NATIONAL PHYSICAL LABORATORY

SYMPOSIUM No. 10

Mechanisation of Thought Processes

VOLUME I

LONDON: HER MAJESTY'S STATIONERY OFFICE

Price £2. 10s. 0d. net for two volumes

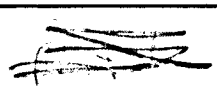
(not to be sold separately)

554

SCIENTIFIC LIBRARY



UNITED STATES PATENT OFFICE



GPO 16—53001-1

MAY 1 - 1962

SCIENTIFIC LIBRARY
APR 22 1960
U. S. PATENT OFFICE

Teddington, Eng.

NATIONAL PHYSICAL LABORATORY

SYMPOSIUM No. 10

Mechanisation of Thought Processes

Proceedings of a Symposium held at the National Physical Laboratory on 24th, 25th, 26th and 27th November 1958

VOLUME I

LONDON: HER MAJESTY'S STATIONERY OFFICE

1959

SL.
QA
76
T4
c.2

P R E F A C E

This Symposium was the tenth in the present series of N.P.L. Symposia. Two are normally held each year - one on a subject of general industrial interest and the other on a theme of more academic research.

This Symposium was held to bring together scientists studying artificial thinking, character and pattern recognition, learning, mechanical language translation, biology, automatic programming, industrial planning and clerical mechanization. It was felt that a common theme in all these fields was "The Mechanization of Thought Processes" and that an interchange of ideas between these specialists would be very valuable.

It is unfortunate that meeting accommodation in the Laboratory is at present very restricted, and a very large number of people had to be turned away. Nearly 200 delegates attended and of these about one third came from overseas.

A total of 32 papers was presented, and on the fourth day there were parallel sessions - one covering implications for Biology and the other for Industry. With one exception these papers are reproduced here substantially as presented i.e. with only very minor revisions. Several authors have added appendices to their papers since the Symposium. The discussion was recorded and all contributors and authors were asked to edit their contributions. The discussion is reproduced in full.

There were also a number of Lecture-Demonstrations, and a list of these is included. Most of the demonstrations are described in the papers or appendices, and the others are described in short papers.

SCIENTIFIC LIBRARY
APR 22 1960
U. S. PATENT OFFICE

CONTENTS

OPENING ADDRESS

DR. G. B. B. M. SUTHERLAND, NPL

PAGE

SESSION 1 GENERAL PRINCIPLES
Chairman: DR. A. M. UTTLEY, NPL

1 Some methods of artificial intelligence and heuristic programming 3
DR. M. L. MINSKY, MIT, USA

Discussion on paper 1 29
Chairman: Dr. W. S. McCULLOCH, MIT, USA

2 Operational aspects of intellect 37
DR. D. M. MACKAY, King's College, London

Discussion on paper 2 67

3 Programs with common sense 75
DR. J. McCARTHY, MIT, USA

Discussion on paper 3 85

4 The mechanism of habituation 93
DR. W. ROSS ASHBY, Barnwood House Hospital, Gloucester

Discussion on paper 4 115

5 Conditional probability computing in a nervous system 119
DR. A. M. UTTLEY, CME Division, NPL

Discussion on paper 5 149

SESSION 2 AUTOMATIC PROGRAMMING
Chairman: DR. E. T. GOODWIN, NPL

1 Automatic programming: present status and future trends 155
DR. GRACE HOPPER, Remington Rand Univac, USA

Discussion on paper 1 195

2 Some technical features of the Manchester Mercury autocode programme 201
MR. R. A. BROOKER, Manchester University

Discussion on paper 2 225
Chairman: PROF. B. S. SOTSKOV,
Academy of Sciences of the USSR

110482

PAGE

3 Automatic programming: properties and performance of FORTRAN systems I and II 231
MR. J. BACKUS, IBM, USA

Discussion on paper 3 249

4 The work of the Computing Center of the Academy of Sciences of the USSR in the field of automatic programming 257
DR. A. P. ERSHOV, Academy of Sciences of the USSR

Discussion on paper 4 275

MECHANICAL LANGUAGE TRANSLATION
Chairman: DR. L. MEHL, Ecole Nationale d'Administration, Paris

5 Tigris and Euphrates - A comparison between human and machine translation 279
MR. R. H. RICHENS, Cambridge Language Research Group

Discussion on paper 5 303

6 Pronoun reference in German 309
MR. L. BRANDWOOD, Department of Classics, Manchester University

Discussion on paper 6 341

7 Automatic translation in the USSR 351
DR. A. P. ERSHOV, Academy of Sciences of the USSR

SESSION 3 SPEECH RECOGNITION
Chairman: PROF. COLIN CHERRY, Imperial College, London

1 Sensory mechanisms and sensation 357
DR. I. C. WHITFIELD, Saint Elizabeth's Hospital Washington, USA

Discussion on paper 1 369

2 An analogue of the speech recognition process 375
PROF. D. B. FRY and MR. P. DENES, University College, London

Discussion on paper 2 385

PAGE

3 The perception of speech 397
MR. P. LADEFOGED, Edinburgh University

Discussion on paper 3 411

LEARNING IN MACHINES
Chairman: MR. J. F. COALES, Cambridge University

4 Two theorems of statistical separability in the Perceptron 419
DR. F. ROSENBLATT, Cornell Aeronautical Lab., USA

Discussion on paper 4 457

5 Learning machines 473
DR. A. M. ANDREW, CME Division, NPL

Discussion on paper 5 507

6 Pandemonium: A paradigm for learning 511
DR. O. G. SELFRIDGE, MIT, USA

Discussion on paper 6 527

SESSION 4A IMPLICATIONS FOR BIOLOGY
Chairman: PROF. J. Z. YOUNG, University College, London

1 Sensory mechanisms, the reduction of redundancy and intelligence 535
DR. H. B. BARLOW, Physiology Laboratory, Cambridge University

Discussion on paper 1 561

2 Stimulus analysing mechanisms 575
DR. N. S. SUTHERLAND, Institute of Experimental Psychology, Oxford

Discussion on paper 2 603

3 Agatha Tyche: Of nervous nets - the lucky reckoners 611
DR. W. S. McCULLOCH, MIT, USA

Discussion on paper 3 627

PAGE

4 Medical diagnosis and cybernetics 635
DR. F. PAYCHA, Paris

Discussion on paper 4 661
Chairman: SIR FREDERICK BARTLETT, Applied Psychology Research Unit, Cambridge

5 Models and the localization of function in the central nervous system 669
MR. R. L. GREGORY, Psychological Laboratory, Cambridge

Discussion on paper 5 683

6 Some questions concerning the explanation of learning in animals 691
MR. A. J. WATSON, Psychological Laboratory, Cambridge

Discussion on paper 6 721

7 Information, redundancy and decay of the memory trace 729
DR. JOHN BROWN, Birkbeck College, London

Discussion on paper 7 747

SESSION 4B IMPLICATIONS FOR INDUSTRY
Chairman: THE RT. HON. THE EARL OF HALSBURY, NRDC, London

1 Automation in the legal world 755
DR. L. MEHL, Ecole Nationale d'Administration, Paris

Discussion on paper 1 781

2 The mechanization of literature searching 789
PROF. Y. BAR-HILLEL, The Hebrew University, Jerusalem

Discussion on paper 2 801

3 To what extent can administration be mechanized? 809
MR. J. H. H. MERRIMAN and MR. D. W. G. WASS, HM Treasury, London

Discussion on paper 3 819

PAGE

4 Possibilities for the practical utilization of learning processes 825
DR. S. GILL, Ferranti Ltd., London

Discussion on paper 4 835
Chairman: MR. J. MERRIMAN, HM Treasury, London

5 Automatic control by visual signals 841
DR. W. K. TAYLOR, University College, London

Discussion on paper 5 857

6 An analysis of non-mathematical data-processing 863
MR. E. A. NEWMAN, CME Division, NPL

7 Physical analogues to the growth of a concept 877
MR. G. PASK, System Research Ltd., London

Discussion on paper 7 923

LECTURE - DEMONSTRATIONS

List of Lecture-Demonstrations (with references to papers describing them) 931

Machina Reproducatrix 933
DR. A. J. ANGYAN, Physiological Institute, Budapest

Conditional Probability Computer 945
DR. A. M. ANDREW, NPL

A Simple Computer for Demonstrating Behaviour 947
DR. W. ROSS ASHBY, Barnwood House Hospital, Gloucester

Automatic Pattern Recognition 951
DR. W. K. TAYLOR, University College, London

Library Retrieval
MR. S. WHELAN, Royal Radar Establishment, Malvern 953

APPENDICES

1. Attendance List 965

2 Index to authors and contributors 979

THE MECHANIZATION OF THOUGHT PROCESSES

OPENING ADDRESS

by

DR. G. B. B. M. SUTHERLAND

THE National Physical Laboratory occupies a position midway between a university and an industrial laboratory. One of the ways of carrying out our function is to organise symposia in fields in which we are actively engaged and in which universities, industry and other Government laboratories are interested. That this symposium is typical is shown by the fact that roughly one third of the delegates are from universities, one third from industry and one third from government institutions. This was not an administrative decision - it just came out that way.

For the National Physical Laboratory to organise a symposium on "The Mechanization of Thought Processes" is very appropriate, because the future program of the Control Mechanisms and Electronics Division is closely concerned with certain aspects of this problem. Our general aim is to develop equipment which will carry out many of the tedious but essential mental tasks which at present are performed by large numbers of human beings, e.g., pattern recognition, retrieval of information, language translation, and the control of complex operations by trial and error learning. How important is it to solve these problems? We are sometimes criticised for spending time on making, or trying to make, machines, the performance of which compares very unfavourably with that of a human being on the same task, e.g., language translation. There are, of course, three answers to this criticism. The first is that the performance of a child bears little relation to that of the adult into which it develops. The second is that it may not be necessary to do the task as well as the human being, provided the machine does it much more rapidly or at much less cost. The third is that in trying to solve problems of this kind one inevitably learns a great deal about the essential principles which are inherent in the biological solution of the problem, but which would otherwise have lain undiscovered. It is becoming clear that the language of the control engineer has a place in the thinking of biologists and, more generally, that the growing interaction between biology, physics and electronics is likely to prove extremely fruitful to science in the quite near future.

(Dr. G. B. B. M. Sutherland, F.R.S., is Director, National Physical Laboratory, Teddington)

The thinking process brings to mind Descartes' famous remark, "Cogito ergo sum" - I think, therefore I exist. We have as yet no glimmering of how a brain thinks about itself. Even after that stage is reached, we still have the problem of simulating a process carried out at the molecular level, by bits of electronic equipment. Our objectives this week are much more limited, but the problems are no less fascinating, and so let me invert Descartes' proposition and say that, since we exist, we can think, and let us now think to some purpose.

SESSION I

GENERAL PRINCIPLES

Chairman: DR. A. M. UTTLEY, NPL

		PAGE
1	Some methods of artificial intelligence and heuristic programming DR. M. L. MINSKY, MIT, USA	3
	Discussion on paper 1 *Chairman:* Dr. W. S. McCULLOCH, MIT, USA	29
2	Operational aspects of intellect DR. D. M. MACKAY, King's College, London	37
	Discussion on paper 2	67
3	Programs with common sense DR. J. McCARTHY, MIT, USA	75
	Discussion on paper 3	85
4	The mechanism of habituation DR. W. ROSS ASHBY, Barnwood House Hospital, Gloucester	93
	Discussion on paper 4	115
5	Conditional probability computing in a nervous system DR. A. M. UTTLEY, CME Division, NPL	119
	Discussion on paper 5	149

SESSION 1

PAPER 1

SOME METHODS OF ARTIFICIAL INTELLIGENCE AND HEURISTIC PROGRAMMING

by

Dr. MARVIN L. MINSKY

BIOGRAPHICAL NOTE

Marvin Lee Minsky was born in New York on 9th August, 1927. He received his B.A from Harvard in 1950 and Ph.D in Mathematics from Princeton in 1954. For the next three years he was a member of the Harvard University Society of Fellows, and in 1957-58 was staff member of the M.I.T. Lincoln Laboratories. At present he is Assistant Professor of Mathematics at M.I.T. where he is giving a course in Automata and Artificial Intelligence and is also staff member of the Research Laboratory of Electronics.

SOME METHODS OF ARTIFICIAL INTELLIGENCE AND HEURISTIC PROGRAMMING

by

Dr. MARVIN L. MINSKY*

SUMMARY

THIS paper is an attempt to discuss and partially organize a number of ideas concerning the design or programming of machines to work on problems for which the designer does not have, in advance, practical methods of solution. Particular attention is given to processes involving pattern recognition, learning, planning ahead, and the use of analogies or "models". Also considered is the question of designing "administrative" procedures to manage the use of these other devices. The paper begins with a discussion of what is meant by "Intelligence" and concludes with a section concerned with some techniques through which a machine might further improve itself by adding to its collection of problem-solving methods.

1. INTELLIGENCE

I feel that it would not be useful to lay down any absolute definition of "intelligence" or of "intelligent behaviour". For our goals in trying to design "thinking machines" are constantly changing in relation to our ever-increasing resources in this area.

Certainly there are many kinds of performances which if exhibited by a man we would all agree, today, require or manifest intelligence. But would we agree tomorrow? For some purposes we might agree with Turing *(ref.24)*, to regard the same performances in a machine as intelligent. In so doing we would be tying the definition of intelligence to some particular concept of human behaviour.

While such a convention might be useful in some kinds of discourse, its use in serious analysis is precluded by two serious faults. First, it

* The work leading to this paper was supported, in too many ways to cite individually, by the joint services of the U.S.A.

directly evades any concise specification of the kinds of behaviour we are looking for. Second, we can often find simple machines which in certain situations do exhibit performances which would be called intelligent if done by a man.

We are, understandably, very reluctant to confer this dignity on an evidently simple machine. Hence the conflict one would suffer in using this definition would threaten any descriptive value it might otherwise have.

In what situations are we less reluctant to attribute intelligence to machines? Occasionally, a machine will seem to be more resourceful and effective than one might expect from casual inspection of its structure. We may be surprised and impressed and we tend to remain so until through analysis or "explanation" the sense of wonder is removed. Whenever a system behaves as though it had more resources than were evident at "first glance" we react in this way, and this reaction is closely related to that involved when we judge a performance to be an exhibition of intelligence. But clearly this reaction depends also on the resources of the individual who is making the observation. The behaviour of any machine (as we use the term) is always explicable in terms of its past states, external contingencies, and the causal or probabilistic relations between them. Hence the significance of the observer's surprise in this; it can be inferred that the observer is not so good a mathematician that his first glance constitutes an adequate analysis of the situation. In the same way, our judgements of intelligence on the part of other humans are often related to our own analytic inadequacies, and these judgements do shift with changes in understanding.

We frequently find that a skill which seemed highly intelligent in others becomes much less impressive when we have learned the trick of doing it for ourselves. Indeed, we refer to many very complicated procedures as matters of "skill" rather than of intelligence apparently just because there happens to be a known method of instruction through which the ability can usually be acquired.

In attempting to design intelligent machines we are, in effect, concerned with the problems of "creativity". Many people are hostile to such an investigation, maintaining that creativity (or intelligence) is some kind of "gift" which simply cannot be understood or mechanized. This view can be maintained only through a constant shifting of definition. As soon as any process or performance has been mechanized or otherwise explained, it must be removed, with qualifications and apologies, from the list of creative performances. This part is perfectly reasonable; once a process has been mechanized one no longer needs terms like "creative" for its description, and we, too, remove it from the list of things to be accomplished. The weakness of the advocate of inexplicable creativity lies in the unsupported conviction that after *all* machines have been examined some items will still remain on the list.

Let us put it clearly then, that in exploring what we call "the artificial intelligence problem" we are not looking for any kind of closed solution to any question like "what is intelligence and how can it be mechanized?". The judgement of intelligence is more a reflection on what we understand than on what we, or machines, can do. Instead, we are searching for new and better ways of achieving performances that command, at the moment, our respect. We are prepared for the experience of understanding and the consequent reshaping of our goals.

2. PROBLEM-SOLVING

How do humans solve problems? To begin with we have to replace our intuitive requirements by reasonably well-defined technical questions. This may require the largest part of the intellectual effort, but we cannot dwell on the subject. A minimal requirement is that one have a method of discerning a satisfactory solution should one appear. If we cannot do this then the problem must be replaced by one which is well-defined in that sense, and we must hope that solution of the substitute problem will turn out to be useful.

In the best case we come equipped with an efficient *algorithm:* a systematic procedure which, given the problem as input, is guaranteed to produce a solution as output; efficient in that the solution will arrive within reasonable bounds on time and effort. But for new and interesting problems we don't usually have algorithms, or at least not efficient ones. At the other extreme we may know nothing about how to get a solution (except for the ability to recognize one). In such a case we have no alternative save to launch into an exhaustive search through the ensemble of potential solutions, e.g., the set of all proper expressions in our language. Random search is no better in general than systematic exhaustion, and may introduce the possibility of failure. It is tempting but irrational to look for a panacea in chaos. But in any case it is well known that for interesting problems exhaustive search is usually out of the question, even with the aid of the most powerful conceivable machines.

Normally, we are not motivated to attempt such problems. "Interesting" problems always have roots in areas which are at least partially understood. We usually have a good deal of partial information about how to get a solution. But this information may occur in fragmentary form: we may have some information about the "form" of a solution, recollections of similar problems solved in the past, general suggestions, hints, and the like.

We need to find ways of writing programs which will be able to use these fragments, or general advice, to reduce the amount of search to reasonable proportions.

"Hints", "suggestions", or "rules of thumb", which only usually work are called *heuristics.* A program which works on such a basis is called a *heuristic program.* It is difficult to give a more precise definition of heuristic program - this is to be expected in the light of Turing's demonstration that there is no systematic procedure which can distinguish between algorithms (programs that always work) and programs that do not always work.

3. METHODS USED IN HEURISTIC PROGRAMS

Below are mentioned some of the ingredients that might go into a complex problem-solving program. Too little is known about heuristics to justify attempting a systematic or exhaustive catalogue.
Some of the methods are later discussed in further detail.

3.1. *Methods which set up searches*

It has often been convenient to imagine problem-solving as taking place over a tree-like structure of trial possibilities. The tree structure is particularly natural for game problems and for theorem-proving; while it may be less appropriate in other areas, it does seem to have general suggestive value. Usually the problem is not so much to find the basic structure (or the domain of things to try) as to find ways of reducing this structure to reasonable size. Thus, in chess, one may use an evaluation function technique to terminate lines of search, and some kind of "plausible move" selection technique to reduce the amount of branching at each node of the game tree. The problem of designing chess-playing programs has provided fertile ground for study of such methods; see, e.g., Shannon *(ref.22)*, Newell *(ref.12)*, Kister *(ref.7)* and Bernstein *(ref.1)*.

3.2. *Methods which set up new problems as subgoals.*

If direct attack fails, it may be possible to replace the problem by one or more (presumably) less difficult problems, in such a way that solution of the subproblems will lead directly to a solution of the original. In complex problems this may extend to many levels. The Logic Theory machine of Newell and Simon *(refs.13,14)* uses a number of such methods.

3.3. *Methods which set up new problems as models.*

Some of the most powerful techniques used by humans replace problems by new ones whose relation to the original is less immediate. These problems are not equivalent, but only heuristically similar to the original, and their solution need not directly lead to solutions of the original. But to the extent that the problems are really related, we may expect that the *methods* which work on the substitutes will be worthwhile trying on the

original. These techniques include the use of models, analogies, and other ways of transforming a problem into one in an area in which we have more experience or more powerful methods.

3.4. *Methods of characterization or description*

The choice of which method to apply to a given problem depends on what kind of problem it is. In order to select appropriate methods, the machine must have facilities for recognizing problems as members of heuristic categories. If the machine is to be able to learn on the basis of past experience, it must have some way of discerning which records are relevant to a given situation, and again the machine must be able to sort things into heuristically useful classes, i.e., to recognize certain "patterns".

3.5. *Administrative methods.*

As the problem-solving process goes on, we accumulate an elaborate structure of related subgoals and analogue problems. New problems are adjoined, and old ones change their status as evidence accumulates. The management of such a structure requires methods which allocate time and effort, deciding where to concentrate attention. It would seem that these methods should include ways of making the following three kinds of estimates:

(a) Difficulty estimates. For each subproblem, one has to be able to make some estimate of its difficulty - an estimated relation between effort expended and the chance of getting a solution. Such estimates would be based on past experience (using methods of type 4) as well as on whatever direct simplicity estimates could be discovered.

(b) Sub-problem utilities. The second estimates the utility of solving each subproblem with respect to solving the main problem. This will provide a basis for deciding which problem next to attack. This method would use the estimates of type (a) as data as well as data describing the interrelations between the problems. While solutions to some problems may contribute directly to solutions of others, frequently solution of one problem will merely provide "evidence" which has a good chance of being helpful on another. The utility estimate has to weigh the different kinds of evidence and take into account the whole sub-problem structure.

(c) Methods for selecting methods. The estimates above help to select the next problem for attack; one must then select what methods to try. This choice may be based on records of past successes and failures, special experiments on models, and perhaps on the basis of "advice" given from the outside. In each case it is clear that the advice obtained

will be most useful if it really does have general applicability, and again the machine must be able to use adequate methods of categorization.

3.6. *Methods of self-improvement*

In the most advanced systems we must include in the framework resources which can improve the operation of the system through adjustment of existing methods and addition of new ones. Discussion of these methods is deferred to section 9.2.

4. PATTERN RECOGNITION

The problem of sorting events and situations into useful categories arises in so many ways that it is tempting to regard it as the central problem of artificial intelligence. The enormity of the usual underlying search process requires that each trial result be used to remove (on the average) a relatively large class of trial possibilities. Each method will be fruitful only when applied to some particular class of problems, and efficient operation requires that these be recognized. The sorting operation involved may be called "pattern-recognition" or "characterization". Each category can be assigned a conventional or a computed name. It is only through such names that we can hope to introduce "general" or "informal" advice.

In a machine designed to recognize, or learn to recognize, visual objects or the like, the characterization problem is itself the centre of attention. The simplest techniques are those in which objects are "matched" more or less directly against standards or "templates". One usually has to introduce some notion of similarity; this will generally involve (1) an appropriate measure of goodness of fit with the template and (2) searching for such a match over a set of transformations of either the object or the template. This type of recognition seems limited, in practice, to patterns defined by equivalence with respect to modest collections of easily performed transformations - otherwise extensive search and too many templates would be involved.

A more flexible scheme of categorization is that of listing "properties" or "characteristics". A characteristic of an object is the value of some function from the space of objects into some smaller, more convenient, range. (The values might be, e.g., numerical, verbal, or even neural firing patterns. Objects having a common characteristic should tend, of course, to be heuristically related, e.g., through equivalence under transformations frequently encountered in the problem area involved. *(refs. 16,21).*

The ordered set of values assigned to a given object by a chosen sequence of characteristic functions may be called the *character* of that object and may be used as a computed name of the whole set of objects which have that "character", i.e., have that sequence of properties.

Recognition by (conventional) name of an object is done by matching its character against a set of standards which now may be very compact, easily stored, expressions. Here the matching criterion and the templates play a very small part in the process, and, unless conventional names are required, they may not be needed at all.

While property lists are quite adequate for many purposes, they do reflect the use of a number of computations performed independently of one another. This is not the most efficient way to use computations, and it may not be adequate for general pattern recognition. Often, the decision concerning which properties are to be examined should depend on the outcome of previous computations. Consider the case in which we want to recognize the (topological) pattern of those figures which contain a *closed loop*. This particular recognition clearly requires some kind of *recursive* process (since topological connectedness is not a local property). The pattern involved is defined by invariance with respect to a class of transformations much larger and more complex than is involved in those for the patterns, e.g., "square" and "circle".

Again, even if we did have a set of properties which could usefully characterize "single" visual objects, how could we handle scenes containing a number of such objects? Clearly we want to be able to characterize them individually, and also characterize their mutual relationships. To do this in terms of independent properties would clearly be very awkward. But it would not be very difficult to do it in terms of a *program* containing appropriate conditional transfers.

We can expect then that the descriptive expressions, or names of categories, will require more than simple lists, at least in advanced systems. They will have to take the form of complex phrases in expressive languages, e.g., of fragments of program. Only in this way can we expect to recognize and express the nature of those most important patterns whose definition contains a *recursive* ingredient.

In any case, the success of a recognition program will depend largely on the adequacy of the elementary characterization operations. In section 9 we discuss some of the problems in generating and evaluating such operations.

5. CHARACTER-METHOD MEMORIES

Each time a subproblem is considered as a candidate for action, it is important to have access to records based on information concerning

similar problems, i.e., problems of the same type (character). These records should be useful in predicting which methods are most likely to succeed, how much effort will probably be involved, and if possible, what kind of new sub-problems are likely to be generated. Such records can be obtained in a number of ways:

5.1. *Learning by experience*

In the very simplest domains the machine's experience may be summarized by processes of the "stimulus-response reinforcement" type. In such a process there is stored, for each *(character, method)* pair, a record which reflects the utility, over the past, of applying that method to problems having that character. After each instance of such an event, the program must evaluate the relative success, and accordingly modify the corresponding utility estimate. (The resulting numbers could be interpreted in various ways, e.g., as giving a priority ordering for the methods, or perhaps determining their probabilities of future application.)

The precise details of the way in which this record is up-dated will constitute a committment, on the part of the programmer, about the machine's basis for *inductive inference.* For they influence the way in which the machine's behaviour depends on the evidence acquired earlier in similar situations.

For that reason we cannot say, once and for all, which "learning" or "reinforcement" operators are the most desirable. Different tasks requires different standards of evidence. These depend on, among other things, (a) the cost of acquiring new evidence, (b) the penalty for being wrong, and (c) the rate at which the environment may be changing - i.e., the way in which old evidence loses value. The choice of the general form of the operators has to be made on such bases. Fortunately, the strongly self-correcting nature of most learning systems makes them insensitive to the finest details of the operators involved.

In particular one has to decide on how much of the history is to be recorded and preserved. One might store for each Character-Method pair a single number representing relative success, or one might store detailed information about what happened in each such instance. Should one allow for a possible retreat to an earlier position? Should one record failure and success information separately? These options always face the programmer. See Gorn *(ref. 25).*

The futility of aspiring to an absolute answer to these heuristic questions can be seen (I think) by examining the evidently equivalent question - "How can we decide when a question has been properly answered?". Clearly this question is itself not well-defined until we have *chosen* (not discovered!) its answer.

In any case simple reinforcement will be nearly useless in those problems, like chess-playing, in which a very large amount of effort is

expended in each trial, e.g., game. One has to obtain much more information than is implied by success or failure in reaching the main goal. This can be done by the independent reinforcement of the mechanisms involved in each of the many subgoals. As Newell *(ref. 12)* notes, we can reinforce for each subgoal achieved, the methods directly responsible, as well as the methods which generated its sub-sub-goals. If such information is to be useful the machine must, of course, be able to evaluate the utility of the subgoals with respect to the main goals.

5.2. *Advice from the outside*

One could also introduce information into the character-method tables directly from the outside. Each entry in these tables entails an injunction to the machine of this form: - "In situations of such and such a type, try such and such a method...". Of course, the ease with which this can be done will depend on the extent to which the characters in use represent abstractions familiar to the operator, or to humans in general. McCarthy *(ref. 9)* discusses this general topic.

Devotees of completely self-organizing systems may feel that this is an irrelevant, if not objectionable, technique. But perhaps we ought not be so ambitious as to be able to design a machine which, from the very start, will have great problem-solving ability. We do not yet know enough about heuristics. Perhaps we have to first learn how to design machines with which we can establish effective communication. At this stage we need, at the least, a channel through which we can introduce suggestions which will help the machine evolve.

5.3. *Guidance through gradation of problem difficulty*

In evolutionary systems, the ability to learn to solve difficult problems usually depends on the opportunities for exposure to sequences of related problems of graded difficulty, e.g., to a slowly changing environment. We do not expect systems like those of Solomonoff *(ref. 23)*, Newell et. al. *(ref. 14)*, or the like, to efficiently solve problems far in advance of the kind they have already worked. Presentation of a graded series of problems provides the opportunity to work up sets of simple heuristics which may later be combined to form more powerful ones. Complex actions are usually formed of previously available elements; blind search through the space of complex methods would be as hopeless as exhaustive search in any other complex problem domain. We note that the presentation of carefully prepared sequences of problems may be regarded as a form of (indirect) *instruction*. Evolution moves rapidly in the presence of a helpful intelligence. Thus, in teaching arithmetic, one presents children (and Solomonoff's machine) first with addition without carries. When that is

mastered, one can introduce carries; to do this at the start would strain the available resources for method-generation or concept-formation.

In the very near future we shall be working with heuristic programs of such complexity that the designers will be unable to keep in mind a clear picture of their manner of operation. This is already the case in regard to programs resulting from collaboration, and in the case of some existing large switching systems. Even so, if we have available a method or channel through which we can tell how a problem is being attacked, we may still be able to guide the system with helpful suggestions. But without such communication, we may quickly lose the ability to diagnose, correct, or improve the operation of the system.

6. PLANNING AHEAD

Before we attempt to solve a problem or sub-problem we want to select promising methods, perhaps arranged in a strategy or plan. People often do this by indirect methods involving simplifications or models of the problem. If the task is to prove a general theorem, we may first try to find a string of methods which will work in some special case, and later extend the proof scheme. In plane geometry we will draw a relatively crude figure and use visual pattern recognition techniques. In a game one might play out continuations confined to a limited part of the board; only if something promising turns up will one examine the rest of the problem in detail.

Heuristic models need not be consistent in order to be useful. A mathematician can often test the plausibility of a high-dimensional geometric assertion through the use of 2- and 3- dimensional figures whose mathematical structure is quite obscure and personal. Much abstract thinking seems to involve the use of such structures, for which some calculations may be especially simple, provided that the operator knows how to steer around the contradictions that may arise. It may be very difficult to find any single model which reflects all the important features of a class of problems. But fortunately this isn't always necessary. For a partial model may serve well on those sub-problems which themselves don't contain all those features. More detailed examples of this will be seen later.

Just what is a model? One cannot give an absolute definition of the relation between an object and a model of it, simply because the adequacy of the model depends on the questions that one is going to ask it. Perhaps one should say that A* is a model of A to the extent that A* and A give similar answers to those questions considered relevant to the current problem (given suitable codings for the questions). This dependency on a third party cannot be avoided by phrases like "preserving the essential

features of the situation"; still, a structure will not be very useful unless it will serve as a source of models for a range of situations.

6.1. *"Algebraic prediction models".*

Suppose that the character-method tables have accumulated a body of reliable predictions concerning (1), which methods are likely to be useful in solving problems of a given character and (2) the expected distribution of the resultant subproblems. It may be possible to use this information in planning ahead, even when the entries do not point directly toward a solution method. This is done by finding, in the algebraic structure of the predictions, chains of methods whose expectation of success is reasonably high. No particular formalism is proposed here, but one can imagine that for each character there is a table indexed by (1) the plausible methods and (2) the plausible resulting subproblem characters. The entries contain data concerning (1) the plausibility, and (2) the estimated difficulty of carrying the method through.

In the most favourable cases the predictions will be relatively definite, asserting that it is good to apply method Mi to problems of type Cj and that the resulting situation will have character Ck. This will tend to occur in machines which have really well-matched characters and methods.

To the extent that these predictions contain reliable information about how problems are transformed, it will be practical to look ahead a number of steps. Naturally, the reliability falls off rapidly with the number of steps. The important point is that this unreliability is out-weighed, for a certain prediction span, by the fact that the look-ahead itself is computationally trivial; at each stage it involves only a few table look-ups. The corresponding application of an actual method to a real problem would involve a substantial effort. Thus we search through the algebraic structure looking for predicted solution chains of reasonable reliability, and we choose those for which the estimated total difficulty is not excessive.

If such a chain of both reasonable estimated reliability and difficulty is discovered, it will be worth-while attempting to perform the corresponding methods. Naturally, there is a good chance that the plan will not work out in detail. But the value of the plan does not then evaporate. It still retains heuristic value. The chain of methods may fail in such a way that it can still be repaired locally; one might eventually invoke, on the basis of experience, a measure of relatedness between methods, and so explore the chains neighbouring those of the plan.

6.2. *Applicability*

In applying the character-method algebra, one proceeds to look far ahead, without paying attention to small details, and taking a rather

optimistic attitude. Perhaps some such technique is required in any system which can hope to solve complex problems. One has to first lay down a plan or plausible sequence of major steps, and only then try to fill in the details. Otherwise one gets involved in hopelessly large search trees. Whenever possible, we do this by using an analogy or model which is either very simple or very familiar. But if no such structure is available, we can try a formal device such as the above algebraic summary of the available information.

Naturally, if the predictions are unreliable, the model will not be very helpful. But if this is the case then the machine would have been in difficulty already, since either (1) the characteristics in use don't describe the situations in heuristically effective ways (so that methods cannot be efficiently selected), or (2) the machine's learning methods are not adequate to discover the relations between problem types and successful attacks.

The plans or chains or predictions could be interpreted as assertions or generalizations of the form: 'Problems of such and such a type can be solved by such and such a strategy'. We might choose for the main goal of a machine, the discovery of reliable statements of this sort. If this were done in a machine working on mathematical problems, such statements might take on the form of assertions of metatheorems or deduction theorems. Clearly the machine would only be asserting such statements, while not necessarily capable of *proving* them. But if they had a good record of validity, this alone might seem impressive. In mathematics, one frequently gives as much or more credit to the man who first conjectures a theorem as one gives to the man who later finds a proof; the former may be considered the more creative achievement.

7. "SEMANTIC" MODELS AND HEURISTIC DIAGRAMS

In order that a model be useful, it is not necessary that it be especially simple; it may even be of value to use a substitute problem which is really more complex than the original, provided that we happen to have more experience and facility in that domain. Consider the behaviour of a student attempting to prove a theorem in plane geometry. He has been given a set of basic propositions and is presumed to be able to apply some (usually unformalized) rules of inference so as to obtain proofs.

Now both the teacher and the student may, and usually do, pretend that they are operating within the formal system. While they make liberal use of "diagrams" they like to believe that the diagrams are "mere heuristic devices" that they could, if necessary, do without. Of course the fact is that most, if not all, geometers do make use of these diagrams, and if paper were not available they would construct them "in their heads".

Most people find that searching within the logic alone seems meaningless - they find that the diagrams are really necessary to give any "meaning" or "interpretation" to the logical expressions. Without such meaning one cannot guide or motivate the search. But with the "meaning" we can use our "geometric intuition"; we can usually tell quite quickly, for example, whether a proposition is true (and thus save a vast amount of work in trying to find proofs for false subgoals). This ability represents, of course, our powerful (informal) resources derived from our spatial experience.

Thus once the mathematical problem is given a geometric interpretation, we can bring to bear heuristic methods acquired in a more familiar domain. A glance at a diagram tells us at once that it is plausible that two lines are parallel - if there is any doubt it is settled by drawing the figure again with some variation of the unconstrained relations. The good geometer may be able to view the figure as a constrained linkage during such a variation, and thus see whether an apparent parallelism is or is not an accident in a particular sketch.

Now such a heuristic diagram certainly represents some kind of (semantic) interpretation of the problem. But is is important to observe that these interpretations need not always be strictly consistent with the logical system as a whole. We may be satisfied with a much more modest kind of "consistency". For our figure needs to supply interpretations only for those expressions in the formal work which are under scrutiny or are likely to arise shortly. If a new expression fails to find an interpretation in the figure, then we draw a new one with, perhaps, a few additional constructed elements. It is far easier to modify figures occasionally than to try to find one that is completely satisfactory from the start.

Thus in working on a problem, our model need not always be fixed from the start. We may modify it whenever an important expression fails to find a meaning. In a geometry theorem-proving program the figure-drawing process could be under the control of an interpretation unit which would make the alterations when necessary. (This is reminiscent of the system proposed by MacKay *(ref. 11)*. A heuristic model may serve well up to the point where trouble is detected, and even then it may turn out that the trouble can be ignored with relative safety. Such is usually the case in practical mathematics; - one seems to stay out of trouble unless one is seeking it, e.g., in connection with the operational calculus or the logical paradoxes.

8. A PROPOSED SEMANTIC MACHINE FOR PLANE GEOMETRY

The experimental programming of machines to prove theorems in Euclidean geometry is of particular interest in connection with the

possibility of substantial use of the interpretations provided by diagrams. (The problem is also particularly attractive because of our highly developed understanding of this interpretation, and because the rather "natural" language used by geometers is not particularly difficult to put into a form suitable for computers.) A program of this sort might consist basically of (1) a logic program, perhaps along the lines of Newell et. al. *(ref. 14)*, (2) a figure-construction program, and (3) a collection of methods which interpret logic expressions, find properties of the figures, and apply the results to control the logical exploration. A program that could do all this would be quite complicated, and our plans are not yet complete. Below are discussed a few of the ingredients that promise to be useful.

8.1. *Construction methods*

Much of the complexity, as well as the richness, of geometry is due to the large domains of constructions available at each stage; one rarely finds a proof that can be easily expressed solely in terms of the geometric elements mentioned explicitly in the statement of the theorem. The new elements are supplied, of course, by "construction", and in the formal system, the construction methods may be regarded as additional rules of inference. If we compare this logic with the one treated by Newell, we see that the available search trees are comparatively overwhelming, and clearly we will need powerful heuristics to minimize the consideration of irrelevant constructions. One might do this by screening proposed constructions for "plausibility", just as one might in a chess-playing program. The program would compute some carefully chosen function of the set of simple relationships between elements of the figures (with special weighting on the kinds of relations important in the current subgoals. Such a program might include some visual pattern recognition techniques, also discussed below.

Introspection seems to reveal that one does not consider a great many constructions in the course of finding a proof, and that many of these are quickly rejected. In addition to the above screening technique, the program would be limited to a small number of construction techniques, e.g., those which produce only a small number of new elements or which have simple symmetry properties.

8.2. *Validity tests*.

We have already noted (section 7) one important way of coupling the model to the formal system, in connexion with deciding whether or not a proposition is true. This could be done by more or less direct measurement with little computer effort. A simple procedure has been found which takes the statement of a theorem as input and constructs an appropriate figure in "general position"; this procedure works for the

great majority of theorems in Euclid. It fails on a few, for which an approximation method can be used. The accuracy required is not large, for if two angles or segments differ by very little there would be no great risk in assuming their equality. It is important to observe that an error in the interpretation programs will *not* induce the machine to present false proofs; an error will only tend to mislead the main program and make it more difficult to find a correct proof. It should also be noted that confining the logical search to the study of propositions which have been shown (by the semantical methods of this section) to be true statements does not in itself make trivial the problem of finding proofs for difficult theorems. It is, however a crucial heuristic step in reducing the otherwise hopelessly large search through the expressions of this big logical system.

8.3. *Pattern recognition methods.*

Another use of the figures suggests itself in connexion with the problem of characterizing situations and choosing methods. For once a subgoal has been chosen, a glance at the diagram will often tell the student how to proceed. He recognizes within the diagram certain patterns which suggest the application of particular methods. Thus, if the subgoal is to prove that a certain pair of angles are equal, the student will become alert to certain configurations of parallels and transversals, congruent or similar triangles, or appropriate patterns involving arcs. (He will recognize these configurations long before he can logically establish their character.) It will not be particularly difficult to program these recognitions. The programmer has the choice of representing the diagrams within the machine either in an abstract algebraic form, or as a real punctate visual pattern. The former is, of course, more convenient for formal manipulations, and the latter would be a valuable aid for the programmer for communicating with, and understanding what is happening within, the machine.

8.4. *Characters and methods.*

The author has hand-programmed the machine to the point that it can handle some of the early theorems of Euclid with reasonably efficient search processes. In the early stages it seems sufficient to characterize problems in terms of their goal relations. Thus theorems are classified simply according to whether they assert that two angles are equal, two lines equal, one line twice another, etc. The methods, too, in the early stages can be simply described. Thus to show that two angles are equal, the character-method tables will list, in order of descending priority, methods whose goals are to (1) show that they are in congruent triangles, (2) in similar triangles, (3) are alternate interior angles, etc. (The proper order is not yet known; they must certainly follow the Euclidean order of

generation, and an important part of the program will contain the methods which decide when the assertions of earlier proven theorems should be incorporated into new methods. This machine will probably be supplied, from the start, with deduction metatheorems which can so exploit earlier results. Later, one can try learning programs for each of these things.)

Each method may itself generate a sequence of subgoals. Thus the first method we mentioned was one which tries to show two triangles congruent. Subgoals for this would be to (1) show equality of two corresponding sides and included angle, (2) three corresponding sides, etc. The very first test of this proposed initial priority structure actually obtained a proof unknown to the author and nearby associates.

The first theorem attempted was the proposition: *The base angles of an isoceles triangle are equal.* The character-method tables generated the following steps; In *ABC*, with *AB* equal to *AC*, we are to show that two angles (*B* and *C*) are equal. Goal: Are they in congruent triangles? (What triangles are they in? *ABC* and *ACB*.) Can we show these congruent by "side-angle-side"? Yes. (*BA,A,AC* matches *CA,A,AB*!) Q.E.D. Most students prove this theorem by constructing the median and proving (easily enough) that the two inner triangles are congruent.

While this doesn't really tell us much about the power of the basic schemes used, it does point out an interesting property of such machines. They need not become so confused as do humans in "degenerate" situations; the student feels uneasy about using the same triangle in two ways at the same time. This does not point to any basic difference between people and machines, however; when more powerful heuristics are programmed, we will not be surprised if the machines too become confused in the special cases in which the internal models fail to give unambiguous representations of things.*

*NOTE. Most of the remarks in this section were based on multilithed notes distributed by the author at the Dartmouth Summer 1956 Artificial Intelligence Project which was supported by the Rockefeller Foundation (with the author attending as a member of the M.I.T. Lincoln Laboratory and also of the Harvard University Society of Fellows). Since then work along these lines has been under way on a program for the IBM 704 by H. Gelernter and others in N. Rochester's group at the IBM Research Laboratories. The work of the latter group will presumably be reported elsewhere. It will be some time before the author's program will actually run on a computer. The actual coding is awaiting construction of a suitable programming language and compiler.

SCIENTIFIC LIBRARY
APR 22 1960
U. S. PATENT OFFICE

9. THE GENERATION OF CHARACTERS AND METHODS

An intelligent system should have the ability to learn to solve more and more difficult problems as it gains in experience. Such a system ought to be able to increase its resources when necessary, e.g., through the creation of new and/or better characteristics and methods. How can such resources be invented?

One could legitimately regard this as just another kind of search problem - a method, or a characteristic, is just a fragment of program. If we had a way of deciding when such an object was useful, then we could, in principle, search through the space of programs. As usual, to attempt such a search without having powerful heuristics, would be folly. How can such a search be made reasonable?

9.1. *Theories or hierarchies of complexity*

An ancient and still attractive idea is that of trying the simplest methods first, and progressing through successively more complex techniques. It would be easy to make this more precise if there were available an acceptable theory or effective method of enumerating methods (or programs, or machines, or logical expressions) in order of simplicity. A number of such rankings have been proposed; these are too numerous to mention here and none of them may be said to give much practical guidance.

In this connection the results of recursive function theory, in establishing hierarchies of relative solvability classes, are certainly suggestive. The trouble is that all of the computable functions get lumped into one base category. Perhaps these "transfinite" techniques contain ideas which could be used to make finer and more useful distinctions within the basic class of computable functions. Time will tell.

9.2. *Invention based on a measure of similarity.*

For want of an absolute ranking of methods, one must, as always, fall back on the basic heuristic principal: Try methods that are similar "in an appropriate sense" to those that have been successful in the past. Each problem domain has, of course, its own structure, and its own relevant kinds of similarity. In each domain it turns out that, *once the available methods are designated by expressions in some chosen language*, then, for better or for worse, particular notions of similarity suggest themselves.

Suppose, for example, that the methods are to be described by programs in the order code of a particular computer. Then we might take two programs to be "similar" to the extent that they have, or share, common instructions. This idea was explored in an interesting program due to R. Friedberg *(ref. 5)*. The method did not seem to work out as a powerful heuristic generator in itself. Presumably, a powerful method generator of

110482

this kind would have to be able to detect those larger fragments of program ("subroutines") which act as more or less separate entities (performing isolable functions). The detection of such entities constitutes in itself a major problem in inductive inference, of course. The author knows of at least one effort under way to work out a program along these lines *(ref. 18)*.

9.3. *Assembly from smaller elements*

It is clear that ordinary elementary machine instructions are too small to serve well as basic terms in languages which can conveniently describe methods. Perhaps such schemes will be much more effective when operating on the expressions of the more powerful programming languages under development. The most effective kinds of schemes are those in which the methods already known to be useful can be "factored" into relatively large common parts (raising again the kind of problem involved in detection of subroutines). Two examples of generators of apparently high heuristic power are described in the works of Selfridge and Dineen *(refs. 4, 15, 16, 17)*, and Solomonoff *(ref. 23)*.

9.4. *The model of Selfridge and Dineen*

The target of this program is to obtain, for the purpose of recognition of visual patterns, a property list description of the type discussed in section 4 above. What is interesting here is the manner in which it is proposed that new characteristics be generated.

The program starts equipped with a basic set of "elementary" transformations A_1, each of which is an operation or transformation which operates on one picture to produce another. These elementary operations might include such potentially useful operations as extraction of boundaries, detection of vertices or line segments, intensification of contrast, filling of hollow figures, and the like. The machine may apply to a given figure any sequence of such elementary operations, each working on the figure resulting from the previous application.

Now each such sequence could be regarded as a particular transformation, so that (starting with a finite base) there is available an infinite variety of picture-transforming operations. Each of these can be converted into a "characteristic" or "property" by following it, e.g., with any function from the space of pictures to the integers. The "blobbing" operation, which replaces a picture by the *number* of separate objects in the picture, was chosen for this purpose.

It would be impractical to generate trial sequences either by enumeration or at random. As Selfridge remarks, we want to try new sequences which are like those that have already proven their worth. Determining the value of a characteristic does not pose any particular problem here; one might compute the correlations between the names of pictures and the values of a given characteristic.

The proposed generation scheme is this: the machine will collect data on the transition properties of the sequences of established utility, and build new sequences for trial on the basis of these transition data. The occurrence of a relatively large coefficient would tend to indicate the presence of something like a subroutine. It would be remarkable if such a generator did not perform much more efficiently than one based on random selection. It may be noted that if one of the elementary operations is itself useless or destructive it will automatically be retired. A worthless but benign operation will be retired, but more slowly. Thus one could occasionally add new elementary operations, if desired.

This promising approach to pattern recognition has never been followed through in full detail, to the author's knowledge, The author has been able to explore some of the properties of these sequences, with the help of a much faster computer than was available to Selfridge and Dineen at the time their work was done, and my results bear out their contention that this way of generating new characteristics is very effective. Further interesting work along this line is reported by Kirsch et. al. *(ref. 6)*.

9.5. *Solomonoff's proposed Inductive Inference Machine*

Each of the method-creation schemes discussed above may be regarded (although the reader is not compelled to do so) as operating on the expressions or strings of a method-describing language. The probably useful methods are obtained by reshuffling in various ways those terms and other grammatical structures that have already been outstanding. As experience accumulates, it might be desirable to revise the language, e.g., by assigning briefer names to the more useful linguistic objects.

The generating schemes of Solomonoff, working on a different class of problems, are much more difficult to describe in brief, and we must refer to his original paper *(ref. 23)*. The difficulty is twofold; his elementary structures are two-dimensional, and the published description is not complete in all details. A revised version is expected to appear in an early issue of the Journal *Information and Control*. The proposed mechanism is quite complex, and can be summarized only incompletely here. The machine is to be given a sample collection of correctly worked examples of (initially simple) mathematical problems. It is then given some uncompleted problems and required to supply the missing parts. To do this, it must be able to discover the meanings of the symbols involved and learn how to use them. This is a prototype problem of the sort in which the task is to make an acceptable generalization about a body of available empirical evidence.

Now as we know, there is no possibility of finding a method of constructing generalizations which will always remain valid. The machine may be able to learn the meaning of the symbol for addition in the context of

problems which do not require "carries". But this interpretation, or generalization, about what the symbol means will certainly turn out to be incorrect as soon as the machine is given addition problems which do require carries. When this happens, the "empirical utility" of the now incorrect generalization must drop to zero. It has no further value in the role of finding solutions. But it would be disastrous if the machine had then to start again from the beginning. The "patterns" involved in addition with carries are very complex (since the most significant digit of the answer is a function of all the digits of the summands). Fortunately, as Solomonoff observes, even when a generalization loses all predictive value, it may still have precious heuristic content, e.g., for the generation of new hypotheses. The next discovery may be accomplished through a relatively small variation of the otherwise discredited structures. To exploit this value, one must of course have some method of building new structures out of old ones, and the body of the paper under discussion is concerned with proposals in this direction.

In connection with the problem of constructing administrative methods to manage these techniques, it turns out that one must deal with at least two distinct notions of utility: (1) the relatively straightforward notion of an empirical or predictive utility, and, (2) some kind of *a priori* estimate of the utility that a structure will have in its role of generating new structures. The forthcoming version of this proposed program will attempt to formulate in mathematical form some of the properties required of these functions.

9.6. *Structural similarity and network machines.*

Most of the systems discussed up to this point may be regarded as dealing with rather clear-cut syntactic processes involving the manipulation of symbolic expressions. But other kinds of adaptive machines have been studied for which there seem to be no such natural linguistic descriptions. These are the "network" machines which are composed of large assemblies of interconnected components, many of which may be operating at one moment. The variations of behaviour, the source of new modes of operations, are obtained from local variations in the properties of and the interconnections between the elements or "cells". In such machines, the cells may be simple and more or less similar to one another, as in the so-called "neural-net" models, or they may be specialized and individually relatively powerful, as in the "pandemonium" schemes of Selfridge (see this volume). In the neural-net models (to which this discussion is confined) the interconnections are often supposed to be organized in a rather loose way on the local level; more precise specifications may be made about how larger sub-assemblies are put together.

The motivation for the study of such machines is at least threefold. First, people feel that the brain must be such a machine, so that eventually we shall be bound to make such studies. Second, we may have here a sort of "natural" source of heuristically useful behavioural variations (useful if we can find ways to reinforce those that we prefer) and the problem is thus of interest from the purely artificial intelligence viewpoint. Finally it is clear that for reasons of speed and economical use of components we shall eventually be forced into the use and study of "parallel" machines.

The variations (and the associated notion of similarity) are derived from local variations in the properties and interconnexions of the individual cells, and it is hoped that these changes on the "microscopic" level will lead to heuristically useful variations on the gross behavioural level. It is well known that this would be a very poor concept to apply to machines like modern computers which are designed so that local variations produce very large behavioural changes. ("Large" has to be interpreted in an appropriate heuristic sense.) We are very nearly forced to the conclusion that this kind of variation can be useful only in *machines whose exact connexions don't much matter from the beginning.* This, I think, is the reason that proposals of this sort have so often been concerned with varieties of "locally random" machines. One should add, of course, that there is general agreement that the genetic control over brain structure is probably not precise enough to permit much more highly organized alternatives. Among the proposals of this sort are, for example, those of Hebb *(ref. 8)*, Farley and Clark *(refs. 2, 3)*, Minsky and Edmonds *(ref. 10)*, Rochester et. el. *(ref. 19)*, and Rosenblatt *(ref. 20)*.

It has been shown, in all these ways, that it is possible to obtain elementary learning and other kinds of adaptive behaviour (including various reinforcement, association, and servomechanical modes) from these locally random or "self-organizing" nets. We cannot escape the feeling that, if really sophisticated behaviour is desired, there must be some way in which the system can acquire highly developed *hierarchies* of organization; that there must develop some kind of relatively substantial physical representation for higher order concepts. Most of the net theorists who were not satisfied with the simplest forms of learning have tended in this direction, usually by trying to show that fragments of the net will tend to become grouped into functionally unitary entities. This was first made plausible by Hebb, who formulated the first reasonably sophisticated, if rather incomplete, theory; he named these functional groups "cell-assemblies". These cell-assemblies are themselves supposed to represent heuristically useful concepts, e.g., individual characteristics of stimuli, and higher order concept-formation is to occur through the development of useful connexions between them. Although at first the development of satisfactory theories about such structures appears nearly hopeless, it would not be too surprising if, once we have mechanisms for

the formation of simple concepts as physical sub-structures, the remaining heuristic theory would not be very different from the kind concerned with the formal or linguistic models. One might soon need have little concern with the underlying nets. This is one reason why the author feels that, even for those whose central interest is in unravelling the mysteries of the brain, it might be well to devote a major share of the effort, at the present time, to the understanding and development of the kind of heuristic considerations that some of us call "artificial intelligence". There seems to be no real danger, at the moment, that we are going to be overwhelmed with so many plausible theories of how people *might* think, that we would not be able to choose from these a few which might help in the understanding of "natural intelligence"!

BIBLIOGRAPHY

1. BERNSTEIN, A. Computer vs. Chess-Player. *Sci. Amer.*, 1958, **198**, No. 6.
2. FARLEY, B. G., & CLARK, W. A., Simulation of self-organizing systems by digital computer. *Trans. IRE Prof. Group on Info. Theory Symposium*, Sept. 1954.
3. CLARK, W. A., & FARLEY, B. G., Generalization of pattern recognition in a self-organizing system. *Proc. Western Joint Computer Conference*, IRE, March 1955.
4. DINEEN, G. P. Programming Pattern Recognition. *Proc. Western Joint Computer Conference*, IRE, March 1955.
5. FRIEDBERG, R. M. A learning machine, Part I. *IBM J. Res. Dev.*, 1958, **2**, No. 1.
6. KIRSCH, R. A., RAY, L. C., CAHN, L., & URBAN, G. H., Experiments in processing pictorial information with a digital computer. *Proc. Eastern Joint Computer Conference*, IRE, 1957.
7. KISTER, J., STEIN, P., ULAM, S., WALDEN, W., & WELLS, M. Experiments in chess. *J. Assoc. for Computing Machinery*, 1957, **2**, 4.
8. HEBB, D. O. The organization of behaviour. *Wiley*, (1949).
9. McCARTHY, J. See this volume.
10. MINSKY, M. L. Neural nets and the brain-model problem. *(Avail. University Microfilms, Ann Arbor, Michigan).*
11. MacKAY, D. M. The epistemological problem for automata. In *Automata Studies*, Princeton, 1956.
12. NEWELL, A. The chess machine. *Proc.* 1955 *Western Joint Computer Conference*, IRE, March 1955.
13. NEWELL, A., SHAW, J. C. Programming the logic theory machine. *Proc* 1957 *Western Joint Computer Conference*, IRE, Feb. 1957.

14. NEWELL, A., SHAW, J. C., & SIMON, H. A. Empirical explorations of the logic theory machine. *Proc.* 1957 *Western Joint Computer Conference*, IRE, Feb. 1957.
15. SELFRIDGE, O. G. Pattern recognition and modern computers. *Proc.* 1955 *Western Joint Computer Conference*, IRE, March 1955.
16. SELFRIDGE, O. G. Pattern recognition and learning. In Third London Symposium on Information Theory, *Academic Press*, (1956).
17. SELFRIDGE, O. G., See this volume.
18. SILVER, R. L. Private communication.
19. ROCHESTER, N., DUDA, HAIBT, L., & HOLLAND, J. Tests on a cell assembly theory of the action of the brain, using a large digital computer. *Transactions on Information Theory*, IRE, IT-2, No.3, Sept. 1956.
20. ROSENBLATT, F. *The Perceptron*. Cornell Aero. Lab. Report VG-1196-G-1.
21. PITTS, W., & McCULLOCH, W. S. How we know universals. *Bull. Math. Biophysics*, 1947, 9.
22. SHANNON, C. E. Programming a digital computer for playing chess. *The World of Mathematics*, 4, Simon and Schuster, 1956.
23. SOLOMONOFF, R. J. An inductive inference machine. 1957 *Convention Record*, IRE Part 2, March 1957.
24. TURING, A. M. Can a machine think? The World of Mathematics, 4, *Simon and Schuster* 1956.
25. GORN, S. Paper to appear in *Information & Control*.

DISCUSSION ON THE PAPER BY DR. M. L. MINSKY

PROF. VAN SOEST: I am very much impressed by so many most interesting papers presented to this symposium, one of which being that of Dr. Minsky. It is of great value for me to come into contact with others working in the same sort of field as we are doing at the Technical University at Delft.

Our aim there is to look for useful and efficient combinations of pure logical machinery with machinery having some degree of intelligence. The pure logic machine, in some cases, can be too complicated, maybe too expensive. The other type, equipped with trial and error and with learning behaviour could be too wasteful in time. It is a search for a reasonable harmony in combining both together. As a very preliminary example indeed, we have built, among others, an apparatus determining the design characteristics of an electromagnetic transformer of which a priori, a few data have been given.

I completely agree with Dr. Minsky on his views on a definition of intelligence and creativity. Doing otherwise, demonaical thoughts of the outer physical world are too easily introduced, confusing the straightforward way in which research in this field seems to be necessary.

A problem related to the topics of Dr. Minsky's paper is of particular interest to me. It is a problem of systematics and taxonomics, say in botany: the identification and recognition of species. Biologists use a complexity of means for this purpose, the most powerful, and the most extensive at the same time, being the use of identification keys, generally of a tree-like structure, as mentioned by Dr. Minsky in para. 3.1. Those keys only can represent in a useful way if the definitions used are sharp, and if the plant material in study is complete enough. If not, the botanist has to take other considerations into account.

An example for this is the use of pictures or figures in addition to the language written or printed keys. They deliver a pattern recognition in a better way than is possible with a rather short text. See Dr. Minsky's paper, para. 3.4.

Very often the identification key is not composed in the most scientific way. Say, for instance, that fruit characteristics are predominant, but that in practice fruits very often are not available; if a first split-up in the tree structure of the identification key depends on it, this will not be very useful at all. So a transformation in key-characteristics,

often to vaguer ones, is necessary; see Dr. Minsky's paper para. 3.3. I could continue to show the resemblance with Dr. Minsky's problem.

There exist three ways to improve, by some mechanisation process, the identifications in question. Firstly; to change the existing key into a, maybe, totally automatic machine that observes colours, patterns and so on. The technology to reach this may be difficult, but it is not impossible. It does not so much belong to the items of this symposium, perhaps. I would go over to the second point, to develop a machine that has the property of pattern recognition in order to compare identification-key results with pictures and other figures. Thirdly, to develop methods with which a machine, exactly in the way as an author-taxonomist has to do, can compose such keys. A priori, the whole knowledge of the species and their variability concerned has to be put at its disposal.

How can all this be done? I am studying very carefully the general philosophy of how the botanist is doing this work himself, and how this analytical process takes place in him. The ideas explained by Dr. Minsky are relevant to all this, and important and stimulating for my study. I wish him all success with his geometrical model.

DR. H. B. BARLOW: I am impressed by the fact that Dr. Minsky does not seem to put forward any ideas as to the general nature of intelligence, or about the end towards which it is normally directed. He is concerned with a particular manifestation of it — that which enables a person to prove simple geometrical thereoms for instance -- and as I understand him he is trying to get this capacity incorporated into a machine. But this is an advanced manifestation of intelligence which is beyond the powers of many human brains. In other words, isn't Dr. Minsky trying to put the thick end of the wedge into his programmes?

People seem agreed that, in their present form, high speed digital computors are fantastically "stupid"; it would therefore seem logical to try inserting the thin end of the wedge first. A task which seems to me to require such elementary "intelligence" would be the processing of data so that they can be displayed or stored more economically. At its simplest this is an easy task, but it links up directly with Ernst Mach (*ref. 1*) and Karl Pearson's (*ref. 2*) contention that "Economy of Thought", or the economic description of observations, is the essential activity of science.

REFERENCES

1. MACH, E. The Analysis of Sensations. Trans. from the fifth German Ed. by S. Waterlow. *Open Court Publishing Co., London & Chicago* (1914).

2. PEARSON, K. The Grammar of Science. *Scott, London* (1892)

The programming of such tasks might throw more light on intelligence than the programming of problem solving routines.

DR. L. C. PAYNE: I should first like to congratulate the author on what I thought was a most stimulating paper. I have about five points to make and five minutes to do it in, and each of them really requires a paper on its own to do it justice. Although there are extremists on both sides I don't think anyone is under any illusions about whether a machine can think in the same sense that a human being can think. We are here this week to discuss the role of data processing and control systems. Nearly all important control systems depend on the flow of formation in closed loop systems - what might be described as data processing circuits - and such loops generally embody the human brain as the processing mechanism of the loop. What we are trying to do is to find how effectively we can replace this by computing techniques. Simple and familiar examples of first attempts in this direction are the thermostat using one bit of information, and analogue circuits with linear feedback. The problem is, in what way, and to what extent, can we complicate the data processing part of the servo-loop so as to achieve more sophisticated forms of control along the lines suggested by Dr. Sutherland; that is doing less than completely emulating a human being, but perhaps doing something faster or for less cost, which is certainly a sufficiently limited first objective.

I believe that starting in this modest way it is realistic to hope that further developments will lead to equipments that simulate thinking more and more. The grounds for this belief will be clearer if I start with a fairly large generalization and show how it can be narrowed and focussed to manageable proportions.

The essence of all scientific activity, what we call the application of scientific method, is talking consistently about what we have observed. This activity distils itself into the laws of science. With these rules - they might be called programmes - the human brain operates. With the rules of, say, Newton's Laws of Motion, a brain can evaluate the motion of a projectile without going through the original labour of Newton. This is analogous to a computer, which once loaded with a set of rules - a programme - can evaluate results without going through the original labour of drawing up the rules - or programme. The human being can therefore be regarded as a digital computer *par excellence;* his inputs are the very diverse sensory inputs, not nearly so simple as punched paper tape and such like, and many orders more subtle. His data processing element - his brain - can draw from the whole "library" of systematized knowledge that we call science. He can draw from Newton's "subroutines", or Darwin's or anybody else's, by which to process the sophisticated input data. Furthermore, he can then output it in a very complex way through elaborate muscular actions, which make punched and magnetic tapes look very primitive indeed.

That is certainly a large generalization, and I do not want to pursue it; but I think it gives hope that we can go a long way in making control loops with sophisticated control elements simulating human control. As A.N. Whitehead said in "Science and the Modern World","it is the large generalization limited by the happy particularity which is the really fruitful conception". Having sketched a rather large generalization I would now like to particularize it, to bring it down to earth and link it with what is being, and could be, achieved.

We may define intelligence, in a way which is perhaps less than adequate to a philosopher but certainly good enough to recognise intelligence, as the ability to discern relationships both analytically and synthetically. The analytical, of course, we associate with the deductive mode of human reasoning, and the synthetic is what we call the inductive mode of human reasoning. The mechanical nature of deductive reasoning has been recognized for a long time; a person trained exclusively in this process, such as many mathematicians, do not make very successful research workers, at least to begin with. The power of deductive logic in tackling new problems is very limited, and indeed mathematics tend to follow in the wake of new developments rather than in the man. As my Cambridge professor aptly put it, "mathematics is no substitute for clear thinking". However, once rules are established, and I have maintained that scientific activity is simply consistent rule making about observed phenomena, deductive logic can elicit many latent ramifications which are by no means immediately obvious. Further, when science has found rules, for example the physics of colour, we do not think it has failed in its object because subjective experiences, such as redness, are no part of the rules. Similarly, I think eventually, rules will be found for human intelligence so that machines built on those lines can simulate intelligence. I suggest that these will be no less complete because they do not treat the subjective experience of intelligence; and that the arguments that a machine can not be intelligent because it can not *feel* or experience intelligence as we feel it are irrelevant. Once such laws are found, doubtless they will become more and more refined as the science of intelligence becomes established. Deductive logic applied to such laws may elicit many ramifications which engineered into programmes will doubtless give the appearance of subtlety to many machine decisions.

However, one feels much less optimistic that a machine will simulate the inductive mode of reasoning, especially of the kind that we call intuitive insight. This constitutes the essence of originality and creativeness which we tend, perhaps rightly, to regard as more essentially characteristic of human intelligence, than the mechanical-like processes of deduction. Inductive inference is controversial, and some people would contend that there are no systems for induction. However there is one

system which Sir Ronald Fisher calls the fiducial argument (*ref. 1*). It only applies to a restricted class phenomena but it is a mode of induction which is logically sound and watertight. It could therefore be programmed into a machine so that one class of inductive inferences could be made by machine. This is a start. I think any intensification of the study of induction which could result in further systemization would greatly extend the ability of computers to emulate inductive mode of human reasoning. There is one thing more I would like to say on this subject of induction, or the synthetic aspect of human reasoning, that is that a good part of research, of solving problems, is formulating them in such a way that a solution is possible. This should make us very sceptical about a machine solving problems in the more general sense.

The history of science, for example the hereditary process, is punctuated with wrong questions, metaphysical propositions which only yielded ground when a scientific approach was adopted. I think a lot of obscure metaphysical thinking surrounds discussions of intelligence today. With a truly scientific approach, a single new concept can clarify the air considerably, making further progress possible, and other questions meaningful. I think it is possible to advance in a similar way with human intelligence and that with the rules or laws established, that is, intelligent behaviour logically systematized, we can look forward to engineering these rules and having machines acting as intelligently as we understand the mechanism of intelligence at any given epoch.

DR. M. L. MINSKY (in reply):

1. (to Prof. van Soest) I wish to thank Prof. van Soest for his kind remarks. The problem of taxonomic retrieval is a challenging one and mechanization of the composition of keys would require a powerful combination of resources. Prof. van Soest's second proposal -- to mechanize the comparison between a real visual object and a set of diagrammatic representations or pictures --is an interesting one and I wonder how this task would compare in complexity with the more usual problem of characterizing objects in terms of symbolic descriptions or names.

2. (to Dr. Barlow) I am surprised that Dr. Barlow feels that it is somehow wrong to attack the more impressive manifestations of intelligence. The apparent "stupidity" of digital computers is, of course, only a reflection of the things that their programmers make them do (or know how to program). Accordingly, most computers spend most of their time either on conventional

REFERENCE

1. FISHER, R. A. Statistical Methods and Scientific Inference. *Oliver & Boyd, London,* (1956).

numerical calculations or indeed on the processing of data so that it can be stored or displayed more efficiently. To be sure, an intelligent system will need a powerful data retrieval facility, *if* it has to work in an area whose problems require large amounts of storage. But a retrieval system alone, i.e., one not concerned with *the meanings* of the data, will not be intelligent: one cannot forever avoid the intellectual structures involved in the strategies and techniques of problem-solving.

I think there is some misunderstanding about the reasons for the choice of problem domains in the earlier "artificial intelligence" programs. It is true that these have tended to be concerned with problems of formal or mathematical character. But this was not done as a deliberate attempt to simulate the most "advanced manifestations" of intelligence. On the contrary, such problems have been chosen for their relative clarity and simplicity! And they bring out the problems of strategy with a minimum of concern for ordinary data-processing and retrieval routines. Indeed, we have found it more difficult to deal with the "ordinary" manifestations of intelligence that the "extraordinary" ones of games and mathematical problems. In the work on the "Advice-Taker", Dr. McCarthy and I hope to make some attempts on this retrieval problem that Dr. Barlow has raised. The system will have to file advice away in a manner which reflects what is known about its relevancies.

3. (to Dr. Payne). I feel that Dr. Payne is over-pessimistic in two respects. The first matter is connected with the desire for a logically sound, watertight, basis for inductive inference. Now everyone should agree that we cannot expect to find an absolute basis for inductive inference. (See my paper, 5.1). For the appropriate inductive inference basis depends on the universe in which we are embedded, and this in turn can be hypothesized only through an already assumed basis of inference. Clearly *for either men or machines* the basis of induction must be empirical and heuristic; suitable for some universes and not others; and only partially suitable for most (including, presumably the one in which the machine will actually find itself). Now I think that everyone will also agree that the problem of mechanization of inductive inference is indeed basic to the artificial intelligence problem. If we are realistic about this we see that we need not and must not concern ourselves with the question of absolutes but can attend instead to the practical problems of 1) formulating heuristically useful (pragmatic) bases and 2) effectively mechanizing them. The same is true of the problem of pattern-recognition or categorization.

I disagree with Dr. Payne in a second matter (we have no real quarrel on the first) concerning our guesses about how to proceed and what we can expect to achieve. It may seem like common-sense to say that scientific

procedure requires that each phenomenon be well-understood before more complex systems can be effectively studied. But one must be cautious about such assertions when one is in an essentially *engineering* domain. Obviously Mechanical Engineering goes far beyond what can be inferred on the basis of present-day *chemical* theories of structural materials. And the latter, as we all know, goes far beyond what can be inferred from the state of atomic physics, etc. Similarly, I believe, the overall flow diagrams or "principles" of some complex intellectual processes can profitably be studied without full attention directed at the fine structure of, e.g., some of the "learning" sub-processes. For one thing the fine details of the sub-processes may not have very much influence on the overall performance of the system. For another, we will be in a position to evaluate the significance of various proposed sub-processes only when they are embedded in larger systems involved in really respectable performances.

We have to be careful not to attack problems that are really too simple. Otherwise our models will perhaps not tell us anything new or, more serious, not tell us whether the methods that we have tried are really extendable to more challenging problems. This extensional inadequacy is a property of most present-day psychological models of elementary learning and association. They can be as elaborate and mathematical as you like, but they simply don't suggest what to do when, e.g., the rewards are contingent in an interesting way on what the creature might do. This is because they summarize, in a "scientific" way the data obtained when experimental animals are challenged with perfectly trivial problems. As logicians, perhaps, we ought to "walk before we can run" but, as engineers, we must run just as fast as we possibly can: Science will stop if we wait to allow the "foundations" catch up with the pragmatic theories!

Now what I am maintaining is that Dr. Payne is over-pessimistic in his statement that "we can look forward to...having machines acting as intelligently as we understand the mechanism of intelligence at any given epoch".

Let me explain my position with a few remarks on the question of *systems design* and the problem of how one is to understand how very complex processes work. These remarks tie in very closely with those of my colleagues John McCarthy and Oliver Selfridge. (See Session 1, paper 3 and Session 3, paper 6). If you will examine the systems proposed in these papers you will see that underlying each of them is a deliberate attempt to show how one might be able to create large systems *without actually understanding them* in full "chemical" or "atomic" detail. (As a visitor I must bow to your great humorist who has so relentlessly shown us how to do so many things "without actually" doing anything much at all. In America there is rising, among those connected with computers, a science called "Research Simulation" which seems distinctly related to this.)

The idea is that we hope to be able to build up some very complex structures with these systems by use of empirical, heuristic, evolutionary

techniques. These techniques depend critically on being able to observe the system in operation and *make small changes in the desired direction.* Both the *Pandemonium* and *Advice-Taker* systems are expected to be capable of tolerating such interference: in the first by alteration of the "sub-demons" and in the second by addition of general advice to handle new classes of situations. The hope is that in the usual case this interference can be tolerated without extensive revision of the system. We are prepared to find that as these modifications accumulate we will lose our detailed understanding of all but the more general principles of the system's operation. But we should learn, so to speak, the art of steering such evolutionary processes into levels of performance well beyond what we could reasonably hope to obtain through any kind of analytic solutions of systems equations (if any exist).

CHAIRMAN: I wonder whether, in your last few minutes you could say a little bit more about the word 'heuristics', as that is something which is not quite clear. It appears that in different papers the word is used in rather different senses. We would be very grateful if you would say a little about its use.

DR. M. L. MINSKY: By "heuristic" we mean to refer to things related to problem solving. In particular we tend to use the term in describing rules or principles which have not been shown to be universally correct but which often seem to be of help, even if they may also often fail. The term "heuristic program" is thus used in reference to the distinction between programs which are guaranteed to work (and are called "algorithms") and programs which are associated with what the programmer feels are good reasons to expect some success. The term "heuristic" has come into use as a noun as well. We ask someone "what heuristics does your program use?" and he answers with a listing of particular tricks, short-cuts, inductive bases, and the like, regarding each heuristic feature as an individual "heuristic".

(The term has been much used by Polya, whose pioneering work on problem-solving and induction is, it seems, being widely neglected at the present time. *(ref. 1)* The reason for this neglect is, I think, the Polya's writing is actually so rich in unsystematic heuristic suggestions that one's own attempt at systematic development is overwhelmed. As the field becomes more organized we will be more able to exploit this great quarry of insight.)

REFERENCE

1. See the bibliography in POLYA, G., Patterns of Plausible Inference, *Princeton,* (1954).

SESSION 1

PAPER 2

OPERATIONAL ASPECTS OF INTELLECT

by

DR. D. M. MACKAY

BIOGRAPHICAL NOTE

Dr. MacKay is a lecturer in Physics at King's College, London. After graduating from St. Andrew's University in 1943 he spent three years on Radar work with the Admiralty. Since 1946, when he joined the staff of King's College, he has been active in the development of information theory, with special interest in its bearing on the study of both natural and artificial information-systems. In 1951 a Rockefeller Fellowship enabled him to spend a year working in this field in U.S.A. His experimental work has been mainly concerned at first with high-speed analogue computation, and latterly with the informational organization of the nervous system.

OPERATIONAL ASPECTS OF INTELLECT

by

DR. D. M. MACKAY

SUMMARY

THIS paper is concerned with some theoretical problems of securing and evaluating 'intelligence' in artificial organisms, - particularly the kind of operational features that distinguish what we call 'intellect' from mere ability to calculate. Among those discussed are

(a) the ability to take cognizance of the 'weight' as well as the structure of information.

(b) the ability to take tentative steps which are not logically forced, but are disciplined by the evidence.

(c) the ability to steer the exploratory process by a sense of the proximity of a solution.

(d) the ability to envisage a range (not too large) of conclusions compatible with given evidence.

These and other considerations are offered to justify earlier suggestions that the mechanization of intellect requires a hybrid information-system, wherein the conditional probabilities of digital decision-processes are determined by a separate (though interacting) computing process which could operate best on 'analogue' principles.

---

1. INTRODUCTION

'INTELLIGENCE' covers an almost hopelessly tangled knot of concepts in present day literature, thanks largely to the variety of efforts to quantify different aspects of mental performance under that name. If ability to perform complex calculations were a sufficient criterion, then even a conventional digital computor could lay claim to more intelligence than any of us - and perhaps we had better let it make away with the word and be done with it.

At any rate, for our present purpose we are concerned with an aspect of mentation so different that it really needs another word.

Every scientific subject has its 'honest hodmen' (as Philip Henry Gosse, F.R.S., was once classified) - men who excel in a technique, and can turn out quantities of valuable results from a persevering and careful application of it. A conventional computor, however capacious and speedy, is the honest hodman *par excellence*. Like its human prototype, it earns respect and gratitude for indispensable service; but for all its dexterity and reliability, one thing is lacking. It is not 'bright'. It is indeed desperately less bright that the dullest human hodman, for even he permits himself an occasional spontaneity when off duty. It is humanly impossible to be as dull as a digital computor.

Now it cannot be claimed that the term 'intellect' which I have used in the title is much less vague than 'intelligence'; but at least it has so far escaped the quantifiers; and when we speak of 'a man of intellect' we have in mind something which quite transcends the man's ability to calculate and make deductions. He must indeed be able to follow and test and carry through trains of reasoning, and to store information and instructions. But when we admire his intellect it is not for these abilities alone; it is rather for such qualities as alertness, ingenuity, initiative, judgment and originality shown in their exercise. These are the qualities that an employer looks for in the man he will leave in charge; they betoken the mental equipment which, in pure or applied science - or indeed in any field - fits a man to lead rather than to follow: to be a source rather than a mere channel or sink of information.

It is this somewhat nebulous complex of qualities that presents the ultimate challenge to the mechanizer of thought - processes. The challenge is twofold, for if we are ever to devise mechanisms to exhibit such qualities, we shall need also - and perhaps first - to devise criteria by which to evaluate their performance. In view of the quagmire that surrounds the testing of 'intelligence' in humans, the prospect of success might seem remote; and perhaps it is. In their raw state concepts such as initiative and originality seem even more intractable than intelligence. An obvious first step however would seem to be to characterise the effects these qualities have on the way the organism operates as an information-system. Anchored to the relatively concrete framework of informational notions, they may shed some of their elusiveness, and it may become easier to see the features that will be essential in artefacts which are to possess them. We may even venture to hope that our understanding of human thinking might benefit in turn from the attempt.

2. REPRESENTATION OF INFORMATION

Some general principles of a self-organizing information-system with these potentialities have already been outlined *(refs. 2-8)*, and the present paper may be regarded as an extension of the same approach. The essential idea put forward in the 1949 paper *(ref. 2)* is to represent not only the structure of information, but also the *weight* (of evidence*) attached to it (in other words, to embody both "structural" and "metrical" information-content) *(ref. 4)* in a spontaneously active statistical mechanism, so that the forms of possible action depend in part on the structure of information received or stored, while the (conditional) probabilities of action depend on the weights attached to the items of information.

In the 1951 and later papers the principle was extended to show how even the concepts used in such an information system could be developed by it *ab initio*, being represented by 'organisers' which are built up through trial and error to cope with (and so to 'match') the recurrent features of demands made† on its adaptive mechanism. The same organizing routine developed to match a recurrent or stable pattern of demand could serve as the symbol for that pattern, so that re-evocation of an organizing routine to match a demand would constitute *recognition* of the pattern concerned, while its internal reactivation in the absence of demand would amount operationally to consideration or contemplation (remembering) of it in one sense or another according to the internal process that reactivated it. *(refs. 4,5)*.

The thesis of the present paper is that the features which distinguish intellect from mere ability to calculate are mostly concerned with the 'metrical' aspect of information-processing - (i.e.) with the weighing of evidence, and the disciplining of spontaneous activity whose form is not uniquely deducible from the form of evidence ø. It is scarcely a coincidence that the self-organizing information-system just mentioned is well-adapted for such activities; but in any case it will appear that analogous facilities seem to be required in any system for the mechanization of intellect in its more interesting aspects.

---

* c.f. "Amount of Information" as defined by Fisher in "Design of Experiments" *(ref.1)*.

† 'Demand' here does not necessarily imply a stimulus to *external* activity. A visual stimulus from a familiar face, for example, may demand only the *internal* setting-up of readiness to react to the familiar person i.e. of *conditional* probabilities of action.

ø It is not that a move is made contrary to the evidence, but that the evidence and instructions available do not suffice to select uniquely the move that is made.

3. PROBLEM-SOLVING

Intellect (or the lack of it) is typically shown in the way a problem is tackled. As Polanyi *(ref.11)* illustrates convincingly in a recent paper on problem-solving, there is a fundamental difference between a solution by a strictly formalised procedure and what is termed a 'heuristic' solution entailing the crossing of a logical gap, in that the first is logically reversible and repeatable, while the second is not. "Established rules of inference offer public paths for drawing intelligent conclusions from existing knowledge. The pioneer mind which reaches its own distinctive conclusions by a leap across a logical gap deviates from the commonly accepted process of reasoning, to achieve surprising results". *(loc.cit. p.93)*. From our standpoint, then, the exercise of intellect entails not just the transformation but the generation of information *(ref.6)*.

I. *Any information-system with 'intellect' must be capable of activity which is logically undetermined (to a greater or lesser extent) by the information supplied to it.*

The degree of logical indeterminacy (the amount of selective information lacking) defines the width of the logical gap crossed in the solution.

A logical gap can of course be crossed in several ways.

(a) Within a given information-system, *even if its mechanism be physically determinate*, any proposition purporting to predict its own future state *in toto* is logically undetermined (strictly unformulable) until after the event. *(refs.12,9,10)*.

(b) The result of consulting a table of 'random' numbers (or any unknown numbers) is logically undetermined by the user's prior information (though not by the printer's)

(c) The course of a 'random' physical process is by definition logically undetermined by its macro-physical specification.

Indeterminacy of type (a) is logically inevitable and need not be specially contrived; but by the same token it does not help us to mechanize the distinction we require. Type (b) is often used to relieve digital computors from the dilemma of Buridan's ass; and where all possibilities are equally worth trying a random programme offers as good a strategy as any. On the other hand this would be just the situation in which 'intellect' is of no avail; and in all others a completely random strategy would be unintelligent. Novel discoveries might indeed be made; but we should rightly attribute them simply to luck (whatever that means). To make full use of relevant information the table of numbers would have to be weighted accordingly, and a vastly cumbersome library of such weighted tables made available to deal with digram, trigram and higher conditional probabilities.

The alternative to a systematic mechanism guided by a library of tables would seem to be a stochastic mechanism of the kind whose activity suffers

the same statistical constraints as the library of tables, - indeed, the kind that could be used to generate such a library in the first place. Obviously a purely random ('white') noise generator of type (c) would be open to the same initial objection as (b) above. The point is however that a suitable physical system, unlike a table of numbers, can economically and flexibly adjust its conditional probabilities of action (or 'propensities', to use a term suggested by Popper *(ref.13)*, so that the same physical mechanism could mediate both the weighing and combining of evidence and also the generation of stochastic activity appropriately disciplined by it. *(refs.2,5,7,8)*. The primary requirement here is that -

II. *The weight attached to evidence should be represented by a separate physical variable*

which can determine the contribution it makes to the pattern or 'matrix' of conditional probabilities of action in given circumstances. The obvious form of variable to represent 'weight' would be the strength of the signal embodying the evidence *(ref.2)* either in terms of amplitude (i.e. S/N ratio) or of the number of signal elements recruited in parallel. Combination of 'weights', either positively or negatively, could then take place on an 'analogue' principle within the elements whose probabilities-of-action they determine. (With sufficient labour a systematic or random digital approximation to the same procedure could of course be devised as an alternative; but its basic logic would still be that of the 'analogue' stochastic process which it must model, and it would in general be much less economical of equipment).

4. THRESHOLD OF EVIDENCE

The ability to attach weights to evidence is not alone sufficient to secure intelligent action. It may suffice to fix the *relative* probabilities of various forms of action at a given moment, but it does not determine the probability that action will take place.* In real life no finite body of evidence, for however long we accumulate it, can prove a generalisation; and any finite body of evidence is compatible with an infinite number of generalisations. We must constantly draw essential conclusions from insufficient data, while recognizing that the correct conclusion may have escaped us among an infinity of alternatives.

A continuous range of attitudes are possible between the precipitate confidence of the over-speculative and the impotent caution of the rigid logician. They differ in what we may call the *threshold of evidence* below

---

* Supposing the rule is to adopt the highest probability, this does not determine the moment at which it shall be adopted.

which no conclusion is reached. Since neither extreme offers a viable policy, the choice of the *optimal threshold* presents a second-order problem in every intelligent decision, the reasonable solution of which is one of the marks of the man of intellect. Too high a threshold means too low a frequency of adaptive changes in organization or action to meet the flux of demands from the environment; too low a threshold means too frequent and ineffective changes.

III. *The mechanization of intellect thus requires facilities not only for the faithful processing of evidence, but also for evaluating the sufficiency of evidence according to the importance of the decision in hand* -

what, with a fashionable, prefix, we might term a 'meta-computation'. This type of computation is familiar in the theory of games, but its role in the mechanization of intellect is perhaps less generally recognized. Without this ability an ingenious but fastidious logician is tormented and delayed in coming to a conclusion, by the undisciplined clamour of logically possible alternatives. To show true intellect he must know when the occasion demands a lowering of thresholds of evidence so as to make his oscillatory progress converge to a decision. Conversely, the unstable scatterbrain, whose decisions are too frequent and ill-considered, shows a lack of intellect by his inability to set thresholds of evidence high enough to match the occasion.

Where 'weight' of evidence is represented by the strength and/or number of signals, we have at first sight a choice of methods of embodying the threshold of evidence. Operationally, a low threshold means a high probability of action on slight evidence. Zero threshold means spontaneous activity in the absence of evidence. Adjustment of threshold could thus be achieved by altering either the amplification of a signal or the sensitivity of the element receiving it. Since evidence-bearing signals may however converge on one element from several channels, the first method (control of channel gain) would require the amplification in all channels to be adjusted together, while the second (control of bias) would leave gain-control free to be used for the adjustment of 'weight' in each channel, and seems clearly preferable. As always in an informational context, the presence of noise (whether natural or artificially augmented) sets the scale of magnitude, so that

IV. *The system has two basic degrees of freedom, which can be taken as (a) the signal-to-noise ratio, S/N, determining the effective weight of evidence, and (b) the bias-to-noise ratio, B/N, determining the effective threshold of evidence,*

and hence the probability or frequency of spontaneous activity.

It is the second which would be affected in 'metacomputations' concerned with the sufficiency of evidence.

5. ORIENTATION TOWARDS A SOLUTION

There is another feature frequently observed in intellectual problem-solving on which Polanyi *(ref.11)* and others lay stress. This is the thinker's sense of the *proximity* of a solution. An intelligent man wrestling with a problem not only inarticulately moulds his exploratory strategy according to his evidence and past experience; when his thinking is 'getting warmer', he often knows it and guides himself accordingly. After a certain stage we have what F. C. Bartlett calls the 'point of no return' beyond which the thinker finds himself inevitably impelled to the solution.

There are of course trivial cases where this faculty needs no explanation. In a jig-saw puzzle the area remaining unfilled clearly indicates our distance from success; and in hunting for an elusive radio signal we can pick up occasional snatches as we rock the dial which tell us we are 'nearly there'. On the other hand when hunting a formula to make sense of a numerical series, we have no such explicit tokens, and have to rely on whatever share we possess of this 'sense of proximity'. Our common experience of feeling a missing word 'on the tip of our tongue' bears witness to this same faculty. We might call it metaphorically an awareness of the *gradient* of the terrain we are exploring.*

Thus:

V. *The mechanization of intellect requires the trial processes to be replicated with variations,*
preferably simultaneously, but otherwise sequentially with storage, in order that the information-system can sense the existence and magnitude of such 'gradients' - and indeed of higher-order differences. It is evident that a good level of ordinary 'intelligence' could be achieved just by a systematic variation of known parameters and the taking of first (and perhaps second) differences in order to guide the process of exploration. Here again however the most fundamental novelty could be achieved only if *unspecified* variations were sometimes possible, their relative probabilities disciplined in turn as a result of the evidence and experience accumulated by the system. What we want in fact is something analogous to an optimal sampling procedure, where the risks of missing a useful clue are balanced against the risks of arriving at a solution systematically but too late.

A human thinker can of course be wrong in thinking himself near the best solution. He may have reached only a small local dip in a plateau, rather

---

* This may indeed be the way we sense the existence of a problem to be solved - by realizing inarticulately that some hitherto untried neighbouring transformation of our data would in some sense be advantageous. I would not go so far as Polanyi, however, (loc.cit.) and say that what we sense is the existence of a *solution*. To sense a slope underfoot is not the same as to sense the existence of a valley-bottom, though the inference may sometimes be justified in the event.

than the valley bottom he seeks. If our mechanized system is to avoid such traps,

VI. *It must try a range of variations, some on a small scale with high discrimination, others on a larger scale with lower discrimination*

to sense more general trends. It may well be that the familiar benefits of practice in problem-solving arise partly from its influence on the range-distribution explored by our own brain-mechanism.

Although in principle either serial or parallel operation could be used in such explorations, both the capacity and the accuracy of storage demanded by serial operation would be prohibitive for more than the simplest of tasks, as anyone who has attempted to solve difference-differential equations serially in even two variables will realise. With parallel trials, first and second differences could be immediately available and could play their part in increasing the efficiency of concurrent as well as successive explorations, while time derivatives of course would also be directly available without storage. Since the simultaneous storage of n partial results of a serial process requires the same order of minimal information-capacity as their simultaneous computation in parallel, while the overall information rate is only $1/n$th that of the parallel process, it is clear that when n is large a serial method could be justified - if at all - only if storage were a correspondingly 'cheaper' process than computation; and it would have to be more than n times 'cheaper' if the overall information-cost were to be comparable.

In many problems, of course, successive trials may differ widely in method, and any idea of a 'gradient' from one to the next may seem absurd. No doubt in some of these cases one could have correspondingly little sense of the nearness of a solution; but in many others the thinker will be found to have classified the component features of different methods (perhaps unconsciously) in such a way that some are logically near-related and others not, and 'proximity' can have a fairly definite meaning in terms of the space of his classification.

To programme an automaton with our own intuitive system of classification (even assuming we could discover it) would once more impose some of our own limitations on its inventive activity. It could doubtless show originality of a second-rate sort in its variations on our themes; but it could never stun us with the novelty of its discoveries: we would have seen to that. The alternative is again to let it develop its own system, from an adequate experience of practice - and successes - in problem-solving. Obviously some general lines may profitably be laid down by way of programme, - as indeed in the training of human beings. But if we are interested in mechanizing a process of genuine discovery, - the possibility of stumbling across something radically new - we shall want to keep programming to a minimum.

Lest it be objected that no discovery is excluded by any universal system of classification, we may remind ourselves that what is in question is not the possibility but the probability - or if you like, the speed - of discovery. Our automaton is intended as an explorer of some field of activity which has an unknown structure of its own. There will be some procedures which our own experience can recommend to it, as having been profitable in the past. I know of no way of estimating the risk that these may in fact be hindrances to further discovery.

6. THE ENVISAGING OF ALTERNATIVES

We have so far unearthed three components of intellect which seem to have fairly definite operational import: (a) the ability to process and combine information faithfully in respect of its weight as well as its structure. (b) the ability, in the presence of deductively insufficient evidence, to make an exploratory 'logical leap' spontaneously in a direction properly disciplined by the evidence (and by implication the ability to evaluate the sufficiency of evidence for the purpose), (c) the ability to steer the exploratory process by a sense of the proximity of a solution.

Implicit in these however is a fourth which may be best illustrated by a familiar example. Asked to state the next integer in the series 1,2,3, both the dull wit unable to think of one, and the ingenious but fastidious logician who can think of too many, would be tongue-tied. The man of moderate intelligence might say '4'. The man of intellect, though quite capable of thinking of alternatives, would be able to grasp the nettle of decision and might also say '4', allowing himself to be scored 'right' or 'wrong' according to the scorer's criterion of the 'obvious'.

As a (not un-typical) test of 'intelligence' the task fails to reveal what matters: the range of possible alternatives recognized by the subject, and of course the grounds of his selection if any. A fourth component of intellect is thus (d) the ability to envisage a range of conclusions compatible with given evidence.

We might at first be tempted to take the number of recognized alternatives as itself a figure of merit; but here as elsewhere the need for a compromise at once asserts itself. Merely to envisage alternatives takes time; and in many cases an infinity of conclusions are compatible with given evidence. *Once formulated* they may of course be ranked as to their probability on given evidence; but by then it is too late: the time has been lost.

VII. *There must be a threshold of probability below which alternatives stand a negligible chance of being envisaged*

- unless and until further evidence changes the picture.

There are thus two statistical filters, or at least two stages in the statistical filtration process by which a conclusion is reached. The first (the one we have just discussed) determines the chances of formulation of various possibilities; the second (discussed in section 4 above) governs the ultimate selection of one of them*.

If we want a figure of merit, then, we must consider not just the number of possibilities envisaged but also their probabilities on the evidence. Some way must be found† to evaluate both the cost in time and equipment and the risk of formulating each, and also their utilities of formulation. This would represent no small programme, and again it is game theory rather than information theory that is likely to indicate the best line of attack. So far as information theory goes we can say simply that for the subject, the amount of selective information generated by drawing one conclusion out of n which were equally likely on his evidence would be $\log_2 n$; but that the proper use of evidence will normally result in an uneven distribution of probabilities, and so reduce this figure. *The object of weighing evidence is to reduce the amount of selective information generated by the subject in taking his decision.*

Paradoxically, the preliminary process of thinking up alternative compatible possibilities would seem to result in an increased selective information content of decision; but this increase would in fact occur only if the evidence were inadequate. The aim should be essentially to *match* richness of evidence with richness of hypotheses; so that a man who has little evidence or is poor at weighing it would do well not to think up too many hypotheses, while one who has more or is good at weighing it can afford to give more rein to his imagination.

The bigger n (and $\log_2 n$) are, the bigger is the amount of information (to the subject) extracted from the evidence if the weighing process succeeds eventually in eliminating all but one hypothesis. The residual information generated by his decision is not from his standpoint a figure of merit but may be the reverse, since it represents the degree of ambiguity left by the weighing process$^{\emptyset}$.

---

* In passing we may note here the informational importance of *questions*. Superficially a question looks informationally neutral; but in fact to raise a question not already present in a man's mind is to push a fresh possibility past the first of his filters, and so in general to alter the whole weight-structure of the ensemble on which his second filter operates. (c.f. the Socratic method).

† By us, the assessors, - not the mechanism!

$\emptyset$ In creative activity however the opposite may be the case. The degree of originality of an invention may perhaps be well characterised (irrespective of its merit) by the selective information content of its specification relative to existing knowledge.

7. THE IMPORTANCE OF ANALOGUES

This brings us to a final question. Desirable as it evidently is to increase the chances of formulation of profitable hypotheses, - how is it to be achieved? Given a set of hypotheses already symbolised, it is a routine matter to attach probabilities to them on given evidence; but how do we control the probabilities of formulation of hypotheses that are not yet formulated? All we have is the evidence which we could use to filter them once they were formulated.

There would seem to be only one possibility. The evidence itself must be embodied in such a physical form that unforeseeable but reasonable hypotheses or bright ideas have a better-than-average chance of spontaneously *growing* out of their embodiment: (i.e.) so that the physical state representing a good idea is 'near' (in the sense of statistical mechanics) to the state representing the data, and that of most useless ideas 'remote'. This would seem to demand the use of *analogue* methods of representation *(ref.2)* which have the further advantage that undiscovered relations implicit in the evidence are less likely to be lost than in a digital approximation.

To take an excessively simple example, two different routes from the railway station to my front door and my back door respectively could be specified by two different lists of straight paths and right-angle bends. In this numerical form the data might be highly unlikely to suggest the hypothesis that the routes terminate within a few yards of one another, even although once suggested this could be verified numerically in a few seconds. If, however, we were to embody the same raw data by setting up a skeleton scale model, our chances of making the same discovery, even by random inspection, would be enormously increased. Of course if we had been explicitly interested in looking for proximities of that sort we could easily have devised a digital routine to track them down systematically from the numerical data, though this would be a slow business. But what we want is a mechanism with a good chance of making such a discovery quickly even although the question it answers may never have occurred to us. There is a sense in which the consequences implicit in data lie *already represented* in an analogue, whereas in numerical or any other arbitrary symbols such consequences are educed only by *ad hoc* operations. This is not to deny that operations (e.g. measurements) are required on an analogue to make the consequences explicit. The point is that these operations are not essentially different from those by which the data themselves are read; whereas the same consequences cannot be read from digital data without additional prior calculations which must in general take a different form for each consequence educed.

It is when *conditional* probabilities have to be represented, however, that an analogue can score most heavily. Consider, for example, the

difference in complexity between the representation of a given area on a map, and an equivalent catalogue of conditional instructions for all possible journeys. Even if data were complete and certain, so that all conditional probabilities reduced to 1's and 0's, the catalogue would be absurdly inefficient. If, as in a typical inventive situation, the map is still growing and its features have marginal uncertainty, the economy of the analogue becomes still more obvious. New information, instead of causing an entry-by-entry recalculation of a whole catalogue of probabilities, would have all its consequences simultaneously and automatically embodied in the act of representing it. Perhaps still more important in practice, the physically-neighbouring states of an analogue represent logically-neighbouring data, so that an analogue system could tolerate a level of perturbation (noise) that would send a digital system into chaotically diverse activity.

But of course we have cheated a little by taking an example where a map is already known to be constructible. The problem in general is to *discover* an analogue that is sufficiently isomorphic with the conditional-probability-matrix derived from data. It seems highly likely that the development of *insight* in a given situation amounts to the discovery of an internal analogue (in this rather sophisticated sense) in which empirical conditional probabilities can thus be correctly embodied - the shift from *enumeration of events* to depiction of a *state of affairs.*

It may be recalled that in the model of intelligent behaviour described earlier (*refs.4,7,8*) the 'state of affairs' (the 'world' of the organism) is represented implicitly in terms of the pattern of demands it has made on the adaptive system of the organism. The adaptive system comprises a hierarchy of 'organizers' developed to match the conditional-probability structure of the demand-pattern. It is clear in terms of our discussion that a major function of the organizing hierarchy should be to serve as a kind of internal workshop, in which continual efforts are made to evolve analogue structures that (a) match and so embody the empirical conditional probabilities of demands and at the same time (b) suggest further tests of their own validity.

While one can separate the present discussion from all questions of human neurophysiology, it is interesting to note that a nervous system with a large and random overlap of fibres from neighbouring discrete elements is not precluded from having many of the representational properties of a continuous physical medium.* Thus although there is no question of postulating

---

* The possibility of representing *continuity* is of course fundamental to a heuristic analogue. A discontinuous approximation to any continuous pattern may be made as accurate as we please; but no degree of refinement can prevent a model made of sand-grains from falling apart when you try to manipulate it. Without some form of glue (continuity) it has no power to represent most of the heuristically important physical properties - rigidity, plasticity, stickiness, compressibility, or what have you.

miniature scale models of the world in the brain (as in some early theories of Gestalt psychology), it is far from unlikely that effectively-continuous analogues are organized by the nervous system to summarise the conditional-probability patterns discovered in its world of activity.

CONCLUSION

I am painfully aware that much of this paper must seem trite and laboured to those who are steeped in the problems of artificial intelligence. Ideas which were fresh and even hotly disputed ten years ago are now common currency, and far more sophisticated developments will be presented by others in this symposium. In reducing a few of the more subtle features of intellect to operational terms, my aim has been chiefly to bring out some of the pressures that have forced me at least to advocate hybrid analogue-digital principles of design. We may well be restricted for some time (if not indefinitely) to less ambitious models in the metal, and my own interest is solely in theoretical problems at the blueprint stage. But a nagging question is sometimes raised whether our mechanical principles may not have an *inherent* limitation that puts the goal of 'true intellect' beyond our power to mechanize. It may perhaps do no harm to morale to do some digging at the question ourselves.

REFERENCES

1. FISHER, R. A. The Design of Experiments. *Oliver and Boyd, London and Edinburgh.* (1937).
2. MACKAY, D. M. On the Combination of Digital and Analogical Techniques in the Design of Analytical Engines. 1949. *Mimeographed.* (Also circulated as an appendix to *ref. 6* at the London Symposium on Communication Theory, 1952).
3. MACKAY, D. M. Mindlike behaviour in artefacts. *Brit. J. Philosophy of Science*, 1951, **2**, 105. Also 1953, **3**, 352.
4. MACKAY, D. M. In Search of Basic Symbols. Proc. 8th Conference on Cybernetics. *Jos. Macey Jr. Foundation, New.York.* (1951).
5. MACKAY, D. M. Mentality in Machines. *Proc. Arist. Soc. Suppt.*, 1952, 61.
6. MACKAY, D. M. Generators of Information. Communication Theory ed. by W. Jackson. *Butterworths, London.* (1953).
7. MACKAY, D. M. Towards an Information-Flow Model of Human Behaviour. *Brit. J. Psychol.*, 1956, **47**, 30.
8. MACKAY, D. M. The Epistemological Problem for Automata. Automata Studies ed. by C. E. Shannon and J. McCarthy. *Princeton.* (1955).
9. MACKAY, D. M. Brain and Will. *The Listener*, May 9 & 16, 1957.

10. MACKAY, D. M. On the Logical Indeterminacy of a Free Choice. *Proc. of XIIth Int. Congress of Philosophy, Venice* (1958).
11. POLANYI, M. Problem Solving. *Brit. J. Philosophy of Science*, 1957, 8, 89.
12. POPPER, K. R. Indeterminism in Classical and Quantum Physics. *Brit. J. Philosophy of Science*, 1950, 1, 173.
13. POPPER, K. R. The propensity interpretation of the calculus of probability. Observation and Interpretation. *Colston Papers*. (1957).

INTRODUCTION TO APPENDIX

I have been kindly urged to publish, in its original form, the 1949 paper referred to on page 41. I do so here with a diffidence that will need no explanation to anyone whose ideas have been on the move for ten years. I am only thankful - and a little surprised - to find as much as I do with which I - and apparently others - now agree.

Section 2 on the Information concept deals only with its descriptive (semantic) aspects. Shannon's recently - published work *(ref. A. 1)* had just come my way, but it was not immediately obvious how it related to the work that Gabor *(ref. A. 2)* and I had been doing on information-measurement *(ref. A. 3)*. One's first impression that it had no bearing on the present discussion turned out to be correct.

The basic question behind the paper is 'what kind of mechanism ought brain-substitutes to be, in order to handle 'intelligently' the kind of thing that information is? The 'autonomous artefact' of sections 6 and 7 is essentially a digital type of self-organising mechanism, but designed to store and take account of *weight* of evidence, computed in associated (preferable analogue) structures.

Where a digital mechanism would compute all-or-none logical conditionals, this mechanism would thus compute and respond according to conditional probabilities - a principle adopted later in Uttley's well-known machine *(ref. A. 4)* (after quite a different train of reasoning), and embodied in most current 'learning machines' in one form or another.

The main change I would now wish to make is in emphasis. Whereas the 1949 artefact could obviously turn up unforeseen and useful theorems, it was limited conceptually by the designer's vocabulary. In 1951 *(refs. 3 and 4 of my preceding paper)* I showed how an artefact on the same probabilistic basis could develop its own concepts, and introduced a much-needed operational distinction between reception and perception. In the later papers cited *(ref. 5-8 above)* I have found it more helpful, especially when thinking of possible brain-models, to talk of the disciplining of potentially random activity, rather than the statistical perturbation of potentially digital precision. Lastly, I must confess to a grievous decline from the youthful optimism of section 9, in respect of the scientific value of actually building such devices. Its last two sentences, with relief, I still heartily endorse.

REFERENCES

A. 1. SHANNON, C. E. A Mathematical Theory of Communication. *Bell. Syst. Tech. J.*, 1948, **27**, 379, 623.

A. 2. GABOR, D. *J. I. E. E. III*, 1946, **93**, 429.

A. 3. MACKAY, D. M. Quantal Aspects of Scientific Information. *Phil. Mag.*, 1950, **41**, 289. Also *Proc. First. London Symposium on Information Theory*, 1950.

A. 4. UTTLEY, A. M. The Classification of Signals in the Nervous System. *E. E. G. Clin. Neurophysiol.*, 1954, **6**, 479.

APPENDIX*

ON THE COMBINATION OF DIGITAL AND ANALOGUE COMPUTING TECHNIQUES IN THE DESIGN OF ANALYTICAL ENGINES

by

DR. D. M. MACKAY

1. TECHNIQUES OF COMPUTATION

1.1. Two district types of mechanism have found application in analytical engines in recent years. In digital mechanisms, a quantal process ensures precision, and order and structure are primary concepts; magnitude, continuity, and trend are derivative, being specified with an accuracy limited only by the complexity of the structure.

In analogue mechanisms on the other hand, magnitude, continuity, and trend are primary concepts; order and structure are abstractions, limited in complexity fundamentally by the amount of information provided by the physical measurement of the primary quantities.

1.2. For many problems in mathematics and mathematical logic, the characteristics of digital mechanisms make them particularly attractive. Where high numerical accuracy is essential, they are much more economical; indeed at present† they appear almost to have excluded analogue processes from the consideration of many of the workers in the field.

On the other hand there are many problems in which the accuracy of known data may be low, and in which the information required is not primarily numerical. One may for example require qualitative information about the *trend* shown by the solutions of a differential equation as certain parameters are varied. This necessitates the rapid presentation of numerous successive solutions, in which great individual accuracy may not be at all necessary.

In such a case a digital machine is not only uneconomical, but cumbersome. (On a general theory of Information, excess accuracy can always in principle be bartered for speed of operation in quite a precise way.) The analogue technique *per contra* offers a method particularly well - adapted to these circumstances.

Any light-hearted rejection of one technique in favour of the other would seem to be unjustified, and it may be useful to consider the kinds of consideration on which a selection of method should be based, and to seek to clarify the principles on which an optimal division of labour might be achieved between the two techniques.

---

* Written for private circulation in May 1949, and mimeographed by the kindness of Dr. Warren McCulloch, then at I.N.I. Chicago. Reprinted here, unaltered, by request. It should not be read apart from the cautionary introduction on p. 53.

† i.e. 1949.

2. THE NATURE OF INFORMATION

2.1. An analytical engine is essentially a device performing transformations on *Information.* This concept in its technical sense has two distinct and complementary aspects *(ref. 1)*, and a brief study of its nature may be found helpful in relation to the present problem.

2.2. Information is defined by the changes which it can cause or validate in a logical pattern. In so far as the scientific method accepts the criteria of two-valued logic, and confines itself to verifiable statements, a scientific statement can be represented by a logical form having a finite number of discrete elementary logical constituents. The author has suggested elsewhere *(ref. 1)* that this leads to a *quantisation* of scientific information in its communicable aspects, so that a unit or *quantum* of information can be defined as that which enables the receiver to make one alteration to the logical pattern describing his awareness of the relevant situation.

2.3. The distinction between the *form* and the *content* of a proposition is fundamental to a proper understanding of information theory. The *form* may be decided *a priori* - that is, before the actual evidence to be subsumed in the proposition has been collected. The *content* is the complex of elementary facts provided by the evidence, and used to depict the form. One might perhaps use the analogy of the construction of a picture from a supply of childrens' blocks, or of the delineation of a diffraction pattern by a series of photons.

Thus if information is to result in the formulation of a proposition, it must have two aspects, (a) *structural*, specifying form, and (b) *metrical* or quantitative, specifying essentially magnitude, and giving content to the proposition with whose form it is associated. Somewhat loosely, we may say that the amount of structural information governs the complexity of a proposition, while the amount of metrical information governs the degree of credibility to be assigned to it.

2.4. A simple formalism enables the relation of these two aspects or kinds of information to be clearly seen. It is discussed at greater length in the paper mentioned above, *(ref. 1)*, but only its essential features are relevant here. A given structure may be defined by a set of "practically independent" data, for example the coefficients in a power series or Fourier expansion in the case of a function of a single variable. In general any proposition based on a finite body of data has a finite number of "degrees of freedom" of this kind, which we may refer to as logons. *(ref. 2)*.

Each logon will be associated with a certain amount of metrical information, analogous to the "quantity of information" defined by R. A. Fisher, *(ref. 3)* and measuring essentially the precision with which the corresponding coefficient or its analogue has been ascertained. The need for

precision in scientific statements effectively quantises the action of this information also *(ref. 1)*, but this need not concern us here. The point is that the total information comprised in a scientific statement can be represented by a set of numbers associated with a finite number of degrees of freedom or logons. These numbers may be regarded as coordinates of a point, in a space of dimensionality equal to the number of logons. The *direction* of the "information vector" from the point to the origin defines the structure of the corresponding proposition. The *dimensionality of* the space defines the amount of structural information contained in the proposition (or required for its formulation). The *length* (or rather the square of the length) of the information vector defines the amount of metrical information in the proposition, and hence the degree of confidence with which it may be received.

2.5. The case is not materially altered when we consider more general non-scientific propositions, except that we need no longer expect the dimensionality to be finite, and the quantal aspects of metrical information are less relevant. In other words in a general transformation of information, as long as it does not issue in a scientifically disciplined statement, there is no restriction on the possible orientations of the information vector, nor on its length. Thus a model of the processes involved in such a transformation must embody both discrete and continuous features. The "excitation" of a logical form is a quantal process; the transformation of information which leads to it may well be a continuous process, or at least may represent a much higher - order approximation to continuity.

3. DIVISION OF FUNCTION

3.1. This representation of information suggests a natural division of function between digital and analogue techniques. The discrete nature of communicable logical structures, and the precise and permanent significance to be associated with each element therein, are factors particularly favouring the digital representation of logical *form*, of structural information. The term "excitation", arising from the element of *decision* which enters into the formulation of a proposition, adumbrates the usual mode of operation of a digital mechanism, via metastable elements.

On the other hand the continuously - variable character of metrical information, and of the concepts associated with it - *trend, probability*, and the like - make it equally natural to represent it in terms of analogue mechanisms employing a continuous variable such as voltage. Within limits the scale (and error-sources) could be chosen so that the metrical information - content of the actual variable was equal to that which it was intended to represent. In any case the provision in the instrument of a source of random fluctuations or "noise" of controllable spectral character, would always make possible the degradation of metrical information where required.

3.2. It is not intended to present an exhaustive analysis of this problem, but it may be useful to examine one or two other functions of such devices. Information storage for example requires normally the repeated regeneration of a given pattern. Evidently a simple re-entrant self-regenerating system must gradually lose precision; so that here the digital method has much to recommend it. If both the length and orientation of an information - vector require to be retained, or even to be communicated via a logical form, this can only be done with permanent precision by way of a preliminary quantisation of the former, to enable it to be stored in digital form. This contingency is recognised as a basis for the quantal view of metrical information put forward by the author in the paper cited above.

On the other hand the short - term retention of information, which is often adequate, would be greatly simplified if the metrical components could remain in association with their respective logons. This could be achieved by representing the requisite information - vector in orientation and length by a signal of appropriate form, the metrical information - content of the signal *qua* signal being greatly in excess of that of the information - vector which it must represent. An analogue retentive mechanism, having a logon - capacity adequately exceeding the logon - content of the signal, could then be used for a limited period to provide a reproduction of calculable fidelity.

3.3. If long-term retentivity is naturally a digital function, the conservation of short-term equilibrium is equally naturally an analogue one. Ultimately only analogue mechanisms can hope to stabilise a system to the limit set by statistical fluctuations, and such mechanisms are well - adapted to perform the functions of control and anticipation, particularly in relation to motor activity.

3.4. Mention of power - control however raises an important point. There are many applications in which a digital, "yes or no" mechanism can usefully take over control when coarse adjustments are required, and can hand over to an analogue mechanism when the point of equilibrium is near. Such arrangements are common for example in large voltage - regulators. They illustrate the fact that as in retentive mechanisms, it may be more satisfactory to hand over longer - term or larger - scale functions to digital mechanisms, retaining the advantages of analogue mechanisms in short - term or small - scale operations.

4. FIELDS OF OPERATION

4.1. The relative importance of analogue and digital techniques in an analytical mechanism will depend on the field from which its information is drawn and with which it interacts. The field of mathematical computation for example, or of mathematical logic, can normally be covered in terms of wholly structural information. Even so there are many processes,

such as division, in which much time could be saved by a combination of the two techniques. An analogue mechanism could provide a first approximation to an answer, enabling the digital mechanism to select the required range of operation and reach the precise conclusion more quickly than if it had to seek it systematically.

This is the converse of the method adopted in power-stabilisation, since now the end - product is discrete structural information, and not a continuous variable.

4.2. Fields of data on physical variables such as temperature, position, velocity, and so on are the province of familiar automatic servo-mechanisms. The information vector is always of finite and (generally) variable length. With pre-quantisation of data the appropriate mechanisms could be made digital in character; but at present at least analogue techniques are normally more convenient, particularly in view of the ease with which differentiation and integration can be performed with respect to time. Previous remarks on the usefulness of digital techniques for coarse control apply to all mechanisms designed to react with a field of physical variables, and the accuracy and range of operation can be greatly increased by a judicious combination of the two techniques.

4.3. The field of immediate relevance to what follows is that of human intercourse. To some extent this combines the two mentioned above, for in principle a given sentence on any subject can be analysed by the discipline of mathematical logic, while its subject matter will normally include data from the time of physical events. This certainly does not mean however that human communication is adequately describable in terms of mathematical logic, nor that its meaningful subject - matter can only be the field of physical events. Philosophical pitfalls abound on the way to be traced out in the next section, and it is necessary to walk warily and to beg as few questions as possible in our choice of words.

5. AN AUTONOMOUS ARTEFACT

5.1. As a basis of discussion, let us now consider an analytical engine designed to manipulate information in terms of a code constructed on the principles of mathematical logic. In such a "perfect logical vocabulary" each code - group is directly dissectible into its definitive constituent atomic facts, and the simple *existence* of an element in a group represents the *assertion* of the corresponding fact. The concept represented by the presence of a given code group is therefore more than a "word" in our passive sense of the term; it is a compresence of *facts*, the assertion of a proposition. Thus, characteristically, the discipline of mathematical logic eliminates the possibility of unrecognised tautology or of statements having purely verbal significance. Any code group generated by the machine

has a definite significance in terms of primitive data - the atomic facts which have been selected by the designer as the basis of its vocabulary.

5.2. Present-day knowledge and technique makes it possible to envisage in practical terms such a machine capable of performing and combining the following functions, among others:

(1) Receiving information from a given field, or from several fields.
(2) Translating that information into terms of its logical vocabulary.
(3) Storing information and reproducing it subsequently.
(4) Performing logical operations on information, whether received or reproduced, so as to reach logical conclusions compatible with all.
(5) Controlling *motor activity*, in accordance with the outcome of logical sequences, so as to *affect the field* and so alter the incoming information.
(6) Continuously *reaching fresh conclusions* and performing fresh operations as a result of (in response to) the changes in incoming data, and hence *inter-alia* the effects of its own activity.
(7) Storing *conclusions*, and subsequently treating them as information.
(8) *Selecting* or rejecting incoming information according to its relevance to the logical process momentarily being carried out, relevance being tested by predetermined criteria.
(9) Selecting *stored* information for reproduction according to similar criteria.
(10) Adjusting or seeking to adjust motor activity according to a given mode of operation until relevant incoming information satisfies given *equilibrium conditions*.
(11) *Changing* from one mode of operation to another if the first does not produce equilibrium, and continuing to do so until success is achieved.
(12) Storing and subsequently treating as information, data on the *relative success* of the various modes of operation, as judged by the rapidity with which they have produced equilibrium (if at all), or by the rapidity per unit of energy dissipated, or some such given criterion.

5.3. A device of this sort can be autonomous, in that it can pursue an active, responsive, and logically - disciplined existence independent of human intervention. In an abstract sense it may be said to show purposeful activity, the purpose being always the attainment of a certain equilibrium configuration, or one of several such configurations, of incoming or present information. The meaning of those purposes in terms of the information field will of course depend on the equilibrium configurations chosen,

and the same choice will therefore determine whether or not they appear "sensible" purposes to a human interlocutor.

This need not mean however that the designer must pre-determine all the purposive characteristics. Provided that an over-riding principle is represented in the design, (e.g. that stability should be sought, and/or that the recurrence - frequency of a particular information - pattern should be maximised) the functions (11) and (12) will enable the machine after a period of active operation to develop subsidiary characteristic patterns of behaviour or "purposes" depending on its history.

It is also important to remember that if the logical vocabulary of the machine relates to an information - field open to human experience, its activity can include the printing or conveying of information from its logical "mill" (4) to a human interlocutor, whilst its input (1) can include rejoinders made by the latter. The development of subsidiary purposes can therefore be constantly under human influence, additional to that exercised by the original designer. The exchange of information can usefully be compared with that which occurs in a discussion by correspondence.

6. THE INTRODUCTION OF INDETERMINACY

6.1. So far the mechanism described has been deterministic. Its potential complexity of behaviour is enormous, but it may be judged to be deterministic by a simple criterion. Two identical devices of the kind described so far, with identical input and output connections to identical fields, would at all times exhibit identical behaviour-patterns. The situation is completely altered however if the operation of one or more of the functions 1-12 is deliberately perturbed by the introduction of a random element, of controllable intensity and spectral distribution.

6.2. There are now two parameters of interest, affecting the operation of a metastable "yes - or - no" logical mechanism. Such a metastable mechanism may be considered to respond to any input in which a particular quantity exceeds a certain threshold - value S_0. The quantity may be a voltage, or the repetition - frequency of a train of stimuli, or some other physical variable, while S_0 can be controlled by an electrical bias, an adjustable time - constant, the concentration of some chemical, or the like. Now perturbation of the operation may be caused by fluctuations in the threshold S_0, or in the appropriate quantity in the input, or by both. Formally, however, all may be attributed to the existence of a random component in the input, having an R.M.S. amplitude N. The probability of an "incorrect" excitation of the metastable mechanism will then be an increasing function of the ratio N/S_0.

6.3. The quantity in the input responsible for excitation may now be regarded as the sum of the random component N, and a component of amplitude S, which in the absence of N would normally exercise full control. The

probability of a "correct" excitation is then an increasing function of the ratio S/S_0. The two probabilities are not of course independent, but the details need not concern us. The essential point is that we now have a mechanism which can be made to show statistical sensitivity to an amplitude, in the same way in which a logical mind shows statistical sensitivity to the metrical information - content of a proposition. If for example the metrical information - content is 1 unit, *(ref. 1)* then by definition a logical mind is as likely to accept as to reject the proposition. Similarly at a certain level of S/S_0, the probability of correct excitation of the mechanism, (of response, that is, to the incoming information represented by the signal S,) will be 1/2. Behaviouristically the two situations are equivalent, and a statistical test of the two would reveal the same frequency of "correct" and "incorrect" decisions. It is easy to envisage a mechanism such that a 1 : 1 correspondence exists between the amplitude of S (or its square) and the metrical information - content of the corresponding proposition.

To the ability to extract logical conclusions from the structure of information, we have now added the ability to act (statistically) according to its *weight* or degree of credibility (para. 2.3.). The possibilities which this opens up in relation to concepts such as judgment, shades of meaning, and the weighing of evidence, seem to be important. It also is suggestive of the possibility of representing, physically, processes of mathematical logic in which *truth-functionals* may be treated as easily as truth - functions are by digital methods. This cannot be considered now, however. It is time to examine briefly some of the effects which controlled "noise" could have at various points in the mechanism of para. 5.2.

7. AUTONOMY AND INDETERMINACY

7.1. In function (4) the effects of controllable perturbation will not be very profound, but will tend if anything to increase the resemblance between the characteristics of the device and those of human beings. There will now be simply a predisposition to commit occasional "vulgar errors." But if the ratios S/S_0 and N/S_0 are controllable, it is easy to see how the analogues of bias, predilection etc. can be introduced, almost literally, in terms of electronic mechanisms. Of such a machine it could be said that *most* of its conclusions are reliable; that it shows a bias against conclusions of particular forms and a *preference* for others; but that if it commits a vulgar error, the latter will normally be followed by a quite logical sequence, and a few cycles of feedback can enable the machine to detect and rectify it unless the bias on the element responsible is unduly large. On the other hand during these cycles, information on the *reaction of the field to the error* has been gathered, and will thereafter count as data in new logical sequences. The future course of the machine cannot be the same as if the error had not occurred, (unless its information - store is deliberately cleared.)

7.2. More interesting effects may be foreseen if selective functions such as (9) are perturbed. If we assume for the moment a perfect (deterministic) logical mill (4), the effect of a limited amount of noise on (9) will be occasionally to allow entry to the mill, of stored information not apparently relevant to the current logical process. Doubtless this information will normally be in fact irrelevant, in the sense that it will lead to no alteration in activity; but particularly if it is allowed, by the design of the machine, to appear even once in its output, its effect on the external field may occasionally be of great interest.

Without begging any questions, it is interesting to compare the process with what we call "getting an idea". Many of the "unbidden" ideas which flash into mind we are accustomed to dismiss immediately and perhaps even subconsciously as "silly". Occasionally, however, such a random thought starts a train of thought which "leads somewhere", and we may then describe our idea as an "inspiration." The analogy in these terms is of course naive, but appears to be highly suggestive of possible concrete and precisely-definable mechanisms simulating some of the functions which tend to be described as "irrational" or "indefinable" in mental activity.

7.3. It is unnecessary to elaborate the picture of the behaviour - pattern which can result from an intelligent extension of these ideas in relation to the other functions postulated in para. 5.2. In some, the presence of noise will decrease the resemblance to normal human activity, or cause characteristics to develop which would be judged undesirable in a human being. In others the resemblance can be still further enhanced. Nor is there space to discuss the question of simulating emotion, though in outline some possibilities are apparent. One characteristic of emotional change is for instance the inhibition of certain functions in favour of others. This corresponds directly to preferential changes, either in bias levels or signal amplitudes, over certain sections of the logical network. It is not difficult to imagine other refinements; indeed the difficulty is to restrain the imagination within the bounds of sobriety!

8. GENERAL CHARACTERISTICS

8.1. Some general conclusions can be drawn about the relation between the characteristics of an autonomous artefact as a whole, and the ratios S/S_0 and N/S_0. It might be of interest to consider the abnormalities resulting from some of the simplest possible contingencies taken in order. The terms used are suggested as some of those applicable to analogous manifestations in human beings, and are of course purely metaphorical. Moreover it is emphasised that in human beings the terms connote much more. The intention is merely to suggest some thoughts of possible catalytic value to those qualified to pursue their implications, if any, in the fields of physiology and psychology.

(a) If S/S_0 and N/S_0 are both low, words such as dull, sluggish, moronic, forgetful, lazy, unreliable, can be given behaviouristic significance in relation to the machine.

(b) If S/S_0 is still low, and N/S_0 increased, characteristics analogous to imbecility appear. Logical sequences are short and disjointed, stored information becomes distorted and hallucinatory impressions arise, motor activity is aimless, and so on. These conditions can of course be present in varying degrees.

(c) If N/S_0 is low, and S/S_0 is increased, behaviour becomes more normal, typical features being reliability, lack of imagination, promptness of response, and so forth.

(d) If S/S_0 is high, and N/S_0 is increased at suitable points, the frequency of "bright ideas" will tend to increase, each being normally followed - up in a logical fashion. If N/S_0 increases further, however, the frequency of new ideas may reach a level at which normal functioning is impaired. In a sense we then have an illustration of the familiar aphorism that "genius is next-door to insanity"; but the "insanity" of the extreme case where N/S_0 and S/S_0 are both high (in the various forms which will differ according to the location of the noise) will be quite different in character from the "imbecility" of group (b).

8.2. No mention has been made of functional disorders caused by organic failure of the functions 1 - 12, as our aim has been only to examine the behaviour - patterns of a perfectly - functioning machine designed to react logically to both metrical and structural information, under varying degrees of degradation of the information. Nor shall we here deal with the many other possibilities which have been mooted, as that sections of such a device might become disordered through the formation of closed loops endlessly repeating non-functional cycles of operation. We might, however, remark that this latter possibility is increased by the presence of noise at any centres which can form such connections between elements.

8.3. In short it is recognised that the foregoing would be a totally inadequate approximation to an account of any nervous disorders. It is felt, however, that in so far as the ideas presented have arisen out of the development of a quite abstract general theory of information as apprehended by the human mind, they might be expected to have at least formal analogues in the mechanism of human thought.

9. CONCLUSION

It might be well to conclude with a question which will certainly be asked: why should anyone want to make such a machine? It is probably not sufficient in these days to suggest that the study of these devices and their potentialities in a worthwhile pursuit of knowledge for its own

sake. In the author's view however the questions raised by these possibilities are of profound importance, and are properly the province of a borderline study, - the science of knowledge itself. As a tool for such work, a device capable of the most general possible transformations of information appears to be of the first importance. Reference has already been made to the possible usefulness of such studies conducted in parallel with the study of cerebral mechanisms, both normal and disordered. In such cooperation between specialists, as Wiener's group has shown,* the benefits to be derived are assuredly not unilateral. The new study indeed would seem to require the full collaboration at least of engineers, logicians, psychologists, physiologists and physicists for its adequate development.

Perhaps of lesser interest is the fact that if what has been said under the heading of autonomy and indeterminacy is true, it is now possible to construct a machine capable of engaging in dialogue, and of making original and unpredictable contributions to it in the terms and categories of human thought. It is not claimed that we have found a necessary and sufficient mechanistic model of human thought - processes, - rather the contrary. But it may well be that one of the most profitable ways of advancing our understanding of the latter is to ask, in no arrogant or rhetorical spirit; what are the differences?

* The reference was to N. WIENER, Cybernetics, John Wiley (1948). Some of the above ideas were put forward by the author in a brief review of Wiener's book in *Electronic Engineering*, July, 1949.

REFERENCE

1. MACKAY, D. M., Quantal Aspects of Scientific Information, (Unpublished).*
2. GABOR D., *J.I.E.E.*, 1946, **93**, III.
3. FISHER, R. A., The Design of Experiments, *Oliver and Boyd* (1935).

* Published later in *Phil. Mag.*, 1950, **41**, 289.

DISCUSSION ON THE PAPER BY DR. D. M. MACKAY

MR. G. M. E. WILLIAMS: All I have to say on this paper by Dr. MacKay is from the point of view of an interested spectator who is not working in this field. I would like to hear more from Dr. MacKay on the following points:-

In Section 4, "Threshold of Evidence" he proceeds to a discussion where he talks about an optimal threshold: He says "Too high a threshold means too low a frequency, too low a threshold means too frequent and ineffective changes". The remarkable thing to me as a layman in this subject, looking through this and other papers on intellect at this symposium is that the word 'criticism' never seems to be used at all. All of us here I imagine have had experience of selecting staff for a post. One of the things we are then interested in, is determining at what level the candidate's intellect lies and one guide is to try to determine how critical he is of what goes on around him and of his work. Criticism is one representation of a feedback process and we have some of the wellknown effects: if one applies too much criticism to considering a set of events, two things can follow, depending on the sign and intensity of the feedback; either vacillation or oscillation arise, or alternatively one is completely fogged and there is no output at all.

This also has a bearing later in Section 4 III. He says: "Since evidence-bearing signals may however converge on one element from several channels, the first method (control of channel gain) would require the amplification in all channels to be adjusted together, while the second (control of bias)...". Another way in which these controls can be effective is by just this critical feedback I have been mentioning.

If I may now step a little outside the terms of this present discussion, I believe it is held that the average information rate in the human brain is about forty bits per second and the storage capacity about ten to the tenth bits. If one does a sum on this basis, assuming a 13-hour day and 40 bits per second, one finds one could occupy the whole of the storage cells in about 14 years. The average span of life is about 70 years - so we are looking for a factor of about five to reconcile a possibly crude theory that the brain stores every bit generated in it. Here storage is clearly different from memory which also involves the recovery of data from storage. I should be interested to hear views on this theory.

DR. GREY WALTER: Dr. MacKay's suggestions are always very provocative and his suggestion that we should consider analogue systems combined with digital ones is welcome to many physiologists who have felt a need for this synthesis during the last ten years. We were all rather easily seduced earlier on by the achievements of digital computers into emphasising the digital aspects of neuronic activity which we were taught in our schooldays as the all-or-none theory of nervous action, but we all know that neurons do not act particularly digitally in the nervous system, and we have been looking for signs of analogue processes in the operation of the intellect.

As Dr. MacKay says, this is a good word to describe what we are interested in and if we take it as meaning something like understanding it can be investigated in laboratories as well as in the lecture theatre. One of the bits of evidence we have been collecting during the last year or so about the intellect of human beings suggests there is indeed an analogue storage process somewhere in the nervous system; not necessarily in the cerebral cortex. This system has one advantage that Dr. MacKay mentioned to topologic mapping as opposed to algebra, that a gross distortion of the co-ordinate system is not fatal, and may be exhilarating. Imagine a map drawn on a rubber sheet; you can find your way round it adequately. Now suppose someone pokes his finger up from below, the easiest route at the beginning may become a difficult one. You may find you have to go over a mountain because the contours seem to have changed, so you have to explore another route, though the Roman road is still available. This seems to be the sort of thing that happens in a human being trying to understand a problem; very often a person makes an assay by a very long way round, assuming some elaborate underlying restrictions, neglecting the fact that the connectivity is quite straightforward. This is an operation of intellect on its lowest level perhaps, but it is important to us in everyday life. Even at this level we find suggestions of exactly what Dr. MacKay speaks of, the advantages of the analogue mapping system which can be perverted or corrupted or embellished by the personality, training or experience of the subject, and yet not lead to total confusion. The person can find his way beyond the mountain poked in his brain by experience or by the experimenter and in doing so, may finally set up a reasonable solution, with or without a complete understanding of the situation; and in the process may make discoveries in realms quite outside the original territory. I think Dr. MacKay's hunch is a good one and that great help could be gained from some such discipline as probabilistic topology, which could tell us the odds on our being able to get from here to there without too much trouble. This is the sort of situation one can study experimentally and model analogue-wise in hardware if called upon to do so.

DR. W. S. McCULLOCH, CHAIRMAN: May I usurp the floor for one moment? One of my old treasured possessions is an article which appeared in the Spring of

1949, never published. It is "On the Combination of Digital and Analogical Techniques in the Design of Analytical Engines" by one MacKay. I would like to see it made available in public. I have often wanted to refer to it.

MR. E. A. NEWMAN: It was interesting to note - on reading the papers of Drs. Minsky, MacKay and Selfridge, that they largely agree with each other as they all imply that all problems can, without real loss of generality, be considered as the recognition or manipulation of patterns. Most of the patterns we have to cope with are very complex - but they are always built up from a host of simple patterns - marks shall we call them? From those marks that interest us it would be possible to build up a well-nigh infinite set of patterns. Because of this - in general - it is impossible to deduce what complex pattern exists given just a few of the marks it contains. One finds, however, from experience that of the enormous number of patterns which could be built from all simple ones, only a few occur which are of practical interest. There is a hierarchy of patterns, those in each layer of complexity being built from patterns in the layer below. At every level only a few of the possibilities occur. It all possible patterns occurred one would expect the numbers existing at each level of complexity to go up almost exponentially with level. In practice one tends to find at each level fewer patterns than at the level below.
Because of this a simple pattern can lead more or less directly to a complicated one, and since many of the very complicated patterns are small varients of each other, one can solve problems by letting a simple clue lead to a complex pattern, and then trying out small random varients of the pattern. Before one can start thinking one must find out the inter-relations that exist between the various primitive patterns. This can only be done by storing the results of experience, but one cannot have a store organization which will allow for every possible inter-relationship since such a store would be much too big. One must in fact start by only allowing for a few crude relationships, and find from this experience what more delicate relationships are worth allowing for.

For example, let us assume we have a computer working at a number of levels. At the lowest level we have basic inputs - n of these say, at the next level we have units which count coincidences between the various possible pairs of inputs - there will be about $n^2/2$ of these. At the next level we have counter units for the various possible triples - about $n^3/6$ of these - and so on. All told there will be 2^n counter units. By comparing these counts we can calculate the probability of having any combination of inputs given any other combination. The number of units required however is very great. If however one restricts the computer, in the first place to the two-at-a-time layer one might well find that only a few of the possible coincidences in fact occur. Indeed there might be enough low count units

available that these could be extracted and used to allow for every combination of useful pair count units and inputs (which would give all likely three units). Similarly there might well prove to be enough three input coincidence counters left little used to allow for all likely four-units, and so on.

In connection with all this, the paper 'Agatha Tyche' by our Chairman suggests a type of neural network which would have the right reorganizing properties and incidentally a system with much redundancy for all the very basic ideas, but very little redundancy for dealing with trivial bits and pieces of ideas.

MR. G. PASK: I would like to ask Dr. MacKay's opinion about one or two issues in his paper and ask him to develop these with reference to certain analogues which have been made in my laboratory.

In the first place Dr. MacKay speaks about mechanisms and the conditions which must be imposed upon an assemblage if -

(i) It is to solve problems in which there is a gap in the logical argument leading to a solution, i.e. if it is to deal with "gap filling" as noted by Prof. Bartlett *(ref. 1)* and

(ii) If it is to specify the 'categories' in terms of which the problem is formulated and solved.

Approaching the field from the direction of problem solving control mechanisms, which in the most general case, are defined as arbitrary assemblages of elements able to interact with their surroundings, at least one of the required conditions is that the elements of the assemblage should be defined with no initial function to serve but should acquire a function in much the same way that a cell in an organism will differentiate in a manner which is partly determined by the cell and its history and partly determined by its surroundings and their history. If the condition is applied it leads to a special but quite realisable analogue and I would like to ask Dr. MacKay if this is the kind of physical entity he envisages when he speaks of a mechanism which builds up a representation of the external world as part of its own structure. If so it is particularly interesting to note the community of various work in progress along these lines. Dr. H. Von Foerster's *(ref. 2)* neurone networks have, for example, a great deal in common with these analogues and they were constructed in pursuit of the same objective.

REFERENCES

1. BARTLETT, F. "Thinking".

2. FOERSTER, HEINZ VON Aspects in the Design of Biological Computors. 2nd. Congress International Association of Cybernetics. Namur, September, 1958.

A characteristic of any such device is that the set of relevant inputs need not be specified at the outset. The assemblage is able to differentiate sensory receptors appropriate for dealing with a problem. I do not wish to examine this in detail for I shall consider it in a later discussion, but mechanisms of this kind seem necessarily sensitive to a variety of physical variables. Suppose that these variables change in some way which disturbs the assemblage and may be regarded as posing a problem. The assemblage will develop sensory receptors able to appreciate these changes so far as doing so will enable it to achieve a stable condition, i.e. will allow it to solve the problem. As my second enquiry I would like to ask whether this is the process that Dr. MacKay refers to when he considers development of an analogue in the substance of an assemblage.

DR. M. L. MINSKY: I would like to make a few remarks relevant to this discussion.

1. There has been much talk about "logical jumps". Programmes can and do make "logical jumps" through the use of heuristics which abandon the certainty of algorithms. There need be nothing mysterious about the appearance of logical jumps in programs whose basis is, to begin with, pragmatic rather than logical.
2. The problem of exponential growth that arises in the model being considered could be regarded as a central problem of artificial intelligence---something important will be contributed when we are shown a model where the structural tree does not grow at such a rate with the input complexity.
3. While I am fully in accord with what Dr. MacKay has said about analogues, I think we should all be reminded that analogues need not be "analogue" in their physical structure! One can, after all, have digital maps and the like. The concept of ability to interpolate is clearly not limited to measurement of physical quantities. In connection with the relation between a problem and an analogue I heard the term "probabilistic topology". I suggest that some new term, e.g., "heuristic connection", be used to describe the relation so that you do not have to worry whether or not there is a better mathematician in the room.
4. In regard to mutations---there is a big logical jump in assuming that the small variations necessarily have to be on a random basis. If you look at your theories carefully you will often find that the probabilistic element has been inserted only for simplicity and convenience of description. Clearly the more 'creative' people choose their ideas from better ensembles (they are too consistent to be merely lucky); it is the establishment of a good trial space and a priori weights that is important—it matters little whether the further search is random or systematic.

DR. D. M. MACKAY (in reply): Mr Williams is right about the importance of 'criticism', and he will find a good deal on the lines he suggests in some

of the earlier references in my paper, where I described mechanisms which seek to have their (internally generated) working hypotheses dis-confirmed and modified by evidence. The evidence-bearing signals, however, would still function by controlling gain or bias, so that this 'critical feedback' does not represent a separate case.

The information-capacity of the brain is hard to define because of the enormous and unspecifiable long-term redundancy in its input. If we bear in mind that the dendritic ramifications of each neuron might well embody several dozen bits of information, it is far from obvious that a lifetime at 40 bits per second would overtax its capacity. Personally however I doubt that even a fraction of this number is in fact reproducibly stored, so that on all counts (*pace* Elsasser, *ref. 1*) there seems no reason as yet to postulate additional and unknown retentive processes.

I have suggested elsewhere *(ref. 8* of my paper*)* that the nervous system may have a simple answer to the problem of exponential growth mentioned by Mr. Newman and Dr. McCarthy. If instead of starting with a vast range of possibilities to be explored, we suppose it to begin with a primitive system having few degrees of freedom, and to use the information gained by experience to guide a gradual differentiation of the exploratory structure, then at no stage need it be faced with an excessive number of useless possibilities to eliminate. The differentiated structure, moreover, will itself be the 'store' of the information that guided its growth.

With regard to analogue representations, I agree of course that one can have sufficiently accurate digital equivalents to answer questions *known in advance;* but one cannot always rely on these to guide heuristic explorations, for which the equivalence cannot have been fully tested. Discontinuous approximations to continuity are notoriously treacherous. We may remember that an incline constructed of finite steps, however small, would require $\sqrt{2}$ times more paint (if thin enough) to cover it than a smooth surface! Again, a model made of digital sand grains may be fine for some heuristic purposes, but it could never suggest hypotheses involving such notions as rigidity or stickiness.

The analogue representations I have suggested, however, need not be *models* in the conventional sense. A 3-dimensional network with adequate connectivity can in fact represent a situation of more than 3 dimensions, for which no model is physically realisable.

I think Mr. Pask's mechanism does belong to the family I have in mind; but although a good many years ago *(ref. 5,* 1952) I described a similar mechanism comprising a network of modifiable statistical links in which, as in his, the minimum of functional detail need be prescribed by the

REFERENCE

1. ELSASSER, W. Physical Foundations of Biology. *Pergaman Press* (1958).

designer, I would recommend in practice a compromise in which some broad lines of development are predetermined, and only the finer details are left to be settled by experience (*ref. 5*, pp.74 & 80). I think that such a mechanism can be informationally more efficient than either a completely unstructured or an over-specified system *(ref. 8)*. It is interesting to hear that others now think this a fruitful principle.

NOTE

The last four references are to those given at the end of Dr. MacKay's paper.

SESSION 1

PAPER 3

PROGRAMS WITH COMMON SENSE

by

DR. J. McCARTHY

BIOGRAPHICAL NOTE

John McCarthy, born at Boston, Mass. in 1927, received his B.S. degree in mathematics at the California Institute of Technology in 1948, and his Ph.D. also in mathematics at Princeton University in 1951. He is at present Assistant Professor of Communication Sciences at the Massachusetts Institute of Technology.

His present interests are in the artificial intelligence problem, automatic programming and mathematical logic. He is co-editor with Dr. C. E. Shannon of "Automatic Studies".

PROGRAMS WITH COMMON SENSE

by

JOHN McCARTHY

SUMMARY

INTERESTING work is being done in programming computers to solve problems which require a high degree of intelligence in humans. However, certain elementary verbal reasoning processes so simple that they can be carried out by any non-feeble-minded human have yet to be simulated by machine programs.

This paper will discuss programs to manipulate in a suitable formal language (most likely a part of the predicate calculus) common instrumental statements. The basic program will draw immediate conclusions from a list of premises. These conclusions will be either declarative or imperative sentences. When an imperative sentence is deduced the program takes a corresponding action. These actions may include printing sentences, moving sentences on lists, and reinitiating the basic deduction process on these lists.

Facilities will be provided for communication with humans in the system via manual intervention and display devices connected to the computer.

THE *advice taker* is a proposed program for solving problems by manipulating sentences in formal languages. The main difference between it and other programs or proposed programs for manipulating formal languages (the *Logic Theory Machine* of Newell, Simon and Shaw and the Geometry Program of Gelernter) is that in the previous programs the formal system was the subject matter but the heuristics were all embodied in the program. In this program the procedures will be described as much as possible in the language itself and, in particular, the heuristics are all so described.

The main advantages we expect the *advice taker* to have is that its behaviour will be improvable merely by making statements to it, telling it about its symbolic environment and what is wanted from it. To make these statements will require little if any knowledge of the program or the

previous knowledge of the *advice taker*. One will be able to assume that the *advice taker* will have available to it a fairly wide class of immediate logical consequences of anything it is told and its previous knowledge. This property is expected to have much in common with what makes us describe certain humans as having *common sense*. We shall therefore say that *A program has common sense if it automatically deduces for itself a sufficiently wide class of immediate consequences of anything it is told and what it already knows*.

The design of this system will be a joint project with Marvin Minsky, but Minsky is not to be held responsible for the views expressed here.

Before describing the *advice taker* in any detail, I would like to describe more fully our motivation for proceeding in this direction. Our ultimate objective is to make programs that learn from their experience as effectively as humans do. It may not be realized how far we are presently from this objective. It is not hard to make machines learn from experience to make simple changes in their behaviour of a kind which has been anticipated by the programmer. For example, Samuel has included in his checker program facilities for improving the weights the machine assigns to various factors in evaluating positions. He has also included a scheme whereby the machine remembers games it has played previously and deviates from its previous play when it finds a position which it previously lost. Suppose, however, that we wanted an improvement in behavior corresponding, say, to the discovery by the machine of the principle of the opposition in checkers. No present or presently proposed schemes are capable of discovering phenomena as abstract as this.

If one wants a machine to be able to discover an abstraction, it seems most likely that the machine must be able to represent this abstraction in some relatively simple way.

There is one known way of making a machine capable of learning arbitrary behaviour; thus to anticipate every kind of behaviour. This is to make it possible for the machine to simulate arbitrary behaviours and try them out. These behaviours may be represented either by nerve nets *(ref. 2)*, by Turing machines *(ref. 3)*, or by calculator programs *(ref. 4)*. The difficulty is twofold. First, in any of these representations the density of interesting behaviours is incredibly low. Second, and even more important, small interesting changes in behaviour expressed at a high level of abstraction do not have simple representations. It is as though the human genetic structure were represented by a set of blue-prints. Then a mutation would usually result in a wart or a failure of parts to meet, or even an ungrammatical blue-print which could not be translated into an animal at all. It is very difficult to see how the genetic representation scheme manages to be general enough to represent the great variety of animals observed and yet be such that so many interesting changes in the organism are represented by small genetic changes. The problem of how such a

representation controls the development of a fertilized egg into a mature animal is even more difficult.

In our opinion, a system which is to evolve intelligence of human order should have at least the following features:

1. All behaviours must be representable in the system. Therefore, the system should either be able to construct arbitrary automata or to program in some general purpose programming language.
2. Interesting changes in behaviour must be expressible in a simple way.
3. All aspects of behaviour except the most routine must be improvable. In particular, the improving mechanism should be improvable.
4. The machine must have or evolve concepts of partial success because on difficult problems decisive successes or failures come too infrequently.
5. The system must be able to create subroutines which can be included in procedures as units. The learning of subroutines is complicated by the fact that the effect of a subroutine is not usually good or bad in itself. Therefore, the mechanism that selects subroutines should have concepts of an interesting or powerful subroutine whose application may be good under suitable conditions.

Of the 5 points mentioned above, our work concentrates mainly on the second. We base ourselves on the idea that: *In order for a program to be capable of learning something it must first be capable of being told it.* In fact, in the early versions we shall concentrate entirely on this point and attempt to achieve a system which can be told to make a specific improvement in its behaviour with no more knowledge of its internal structure or previous knowledge than is required in order to instruct a human. Once this is achieved, we may be able to tell the *advice taker* how to learn from experience.

The main distinction between the way one programs a computer and modifies the program and the way one instructs a human or will instruct the *advice taker* is this: A machine is instructed mainly in the form of a sequence of imperative sentences; while a human is instructed mainly in declarative sentences describing the situation in which action is required together with a few imperatives that say what is wanted. We shall list the advantages of the two methods of instruction.

Advantages of Imperative Sentences

1. A procedure described in imperatives is already laid out and is carried out faster.
2. One starts with a machine in a basic state and does not assume previous knowledge on the part of the machine.

Advantages of Declarative Sentences

1. Advantage can be taken of previous knowledge.
2. Declarative sentences have logical consequences and it can be arranged that the machine will have available sufficiently simple logical consequences of what it is told and what it previously knew.
3. The meaning of declaratives is much less dependent on their order than is the case with imperatives. This makes it easier to have after-thoughts.
4. The effect of a declarative is less dependent on the previous state of the system so that less knowledge of this state is required on the part of the instructor.

The only way we know of expressing abstractions (such as the previous example of the opposition in checkers) is in language. That is why we have decided to program a system which reasons verbally.

THE CONSTRUCTION OF THE ADVICE TAKER

The *advice taker* system has the following main features:

1. There is a method of representing expressions in the computer. These expressions are defined recursively as follows: A class of entities called terms is defined and a term is an expression. A sequence of expressions is an expression. These expressions are represented in the machine by list structures *(ref. 1)*.
2. Certain of these expressions may be regarded as declarative sentences in a certain logical system which will be analogous to a universal Post canonical system. The particular system chosen will depend on programming considerations but will probably have a single rule of inference which will combine substitution for variables with modus ponens. The purpose of the combination is to avoid choking the machine with special cases of general propositions already deduced.
3. There is an *immediate deduction routine* which when given a set of premises will deduce a set of immediate conclusions. Initially, the immediate deduction routine will simply write down all one-step consequences of the premises. Later, this may be elaborated so that the routine will produce some other conclusions which may be of interest. However, this routine will not use semantic heuristics; i.e. heuristics which depend on the subject matter under discussion.

The intelligence, if any, of the advice taker will not be embodied in the immediate deduction routine. This intelligence will be embodied in the procedures which choose the lists of premises to which the immediate deduction routine is to be applied. Of course, the program should never attempt to apply the immediate deduction routine simultaneously to the list of everything it knows. This would make the deduction routine take too long.

4. Not all expressions are interpreted by the system as declarative sentences. Some are the names of entities of various kinds. Certain formulas represent *objects*. For our purposes, an entity is an object if we have something to say about it other than the things which may be deduced from the form of its name. For example, to most people, the number 3812 is not an object: they have nothing to say about it except what can be deduced from its structure. On the other hand, to most Americans the number 1776 is an object because they have filed somewhere the fact that it represents the year when the American Revolution started. In the *advice taker* each object has a *property list* in which are listed the specific things we have to say about it. Some things which can be deduced from the name of the object may be included in the property list anyhow if the deduction was actually carried out and was difficult enough so that the system does not want to carry it out again.
5. Entities other than declarative sentences which can be represented by formulas in the system are individuals, functions, and programs.
6. The program is intended to operate cyclically as follows. The immediate deduction routine is applied to a list of premises and a list of individuals. Some of the conclusions have the form of imperative sentences. These are obeyed. Included in the set of imperatives which may be obeyed is the routine which deduces and obeys.

We shall illustrate the way the *advice taker* is supposed to act by means of an example. Assume that I am seated at my desk at home and I wish to go to the airport. My car is at my home also. The solution of the problem is to walk to the car and drive the car to the airport. First, we shall give a formal statement of the premises the *advice taker* uses to draw the conclusions. Then we shall discuss the heuristics which cause the *advice taker* to assemble these premises from the totality of facts it has available. The premises come in groups, and we shall explain the interpretation of each group.

1. First, we have a predicate *"at"*. *"at(x,y)"* is a formalization of *"x is at y"*. Under this heading we have the premises

1. *at (I, desk)*
2. *at (desk, home)*
3. *at (car, home)*
4. *at (home, county)*
5. *at (airport, county)*

We shall need the fact that the relation *"at"* is transitive which might be written directly as

6. $at(x,y),\ at(y,z) \rightarrow at(x,z)$

or alternatively we might instead use the more abstract premises

6'. *transitive (at)*

and

7'. *transitive (u) → (u(x,y), u(yz,z) → u(x,z))*

from which 6. can be deduced.

2. There are two rules concerning the feasibility of walking and driving.

8. *walkable(x), at(y,x), at(z,x), at(I,y) → can(go(y,z, walking))*
9. *drivable(x), at(y,x), at(z,x), at(car,y), at(I,car) → can(go(y,z, driving))*

There are also two specific facts

10. *walkable (home)*
11. *drivable (county)*

3. Next we have a rule concerned with the properties of going.

12. *did(go(x,y,z)) → at(I,y)*

4. The problem itself is posed by the premise:

13. *want(at(I,airport))*

5. The above are all the premises concerned with the particular problem. The last group of premises are common to almost all problems of this sort. They are:

14. *(x → can(y)), (did(y) → z) → canachult(x,y,z)*

The predicate *"canachult(x,y,z)"* means that in a situation to which x applies, the action y can be performed and brings about a situation to which z applies. A sort of transitivity is described by

15. *canachult(x,y,z), canachult(z,u,v) → canachult(x,prog(y,u),v).*

Here *prog(u,v)* is the program of first carrying out u and then v. (Some kind of identification of a single action u with the one step program *prog(u)* is obviously required, but the details of how this will fit into the formalism have not yet been worked out).

The final premise is the one which causes action to be taken.

16. *x,canachult(x,prog(y,z),w), want(w) → do(y)*

The argument the *advice taker* must produce in order to solve the problem deduces the following propositions in more or less the following order:

1. *at(I,desk) → can(go(desk,car,walking))*
2. *at(I,car) → can(go(home,airport,driving))*
3. *did(go(desk,car,walking)) → at(I,car)*
4. *did(go(home,airport,driving)) → at(I,airport)*
5. *canachult(at(I,desk), go(desk,car,walking), at(I,car))*
6. *canachult(at(I,car), go(home,airport,driving), at(I,airport))*
7. *canachult(at(I,desk), program(go(desk,car,walking), go(home,airport, driving)), → at(I,airport))*
8. *do(go(desk,car,walking))*

The deduction of the last proposition initiates action.

The above proposed reasoning raises two major questions of heuristic. The first is that of how the 16 premises are collected, and the second is that of how the deduction proceeds once they are found. We cannot give complete answers to either question in the present paper; they are obviously not completely separate since some of the deductions might be made before some of the premises are collected. Let us first consider the question of where the 16 premises come from.

First of all, we assert that except for the 13th premise *(want(at(I, airport))* which sets the goal) and the 1st premise *(at(I,desk)* which we shall get from a routine which answers the question "where am I"), *all the premises can reasonably be expected to be specifically present in the memory* of a machine which has competence of human order in finding its way around. That is, none of them are so specific to the problem at hand that assuming their presence in memory constitutes an anticipation of this particular problem or of a class of problems narrower than those which any human can expect to have previously solved. We must impose this requirement if we are to be able to say that the *advice taker* exhibits *common sense.*

On the other hand, while we may reasonably assume that the premises are in memory, we still have to describe how they are assembled into a list by themselves to which the deduction routine may be applied. Tentatively, we expect the *advice taker* to proceed as follows: initially, the sentence *"want(at(I,airport))"* is on a certain list *L*, called the main list, all by itself. The program begins with an observation routine which looks at the main list and puts certain statements about the contents of this list on a list called "observations of the main list". We shall not specify at present what all the possible outputs of this observation routine are but merely say that in this case it will observe that "the only statement on *L* has the form *'want(u(x))'."* (We write this out in English because we have not yet settled on a formalism for representing statements of this kind). The "deduce and obey" routine is then applied to the combination of the "observations of the main list" list, and a list called the "standing orders list". This list is rather small and is never changed, or at least is only changed in major changes of the advice taker. The contents of the "standing orders" list has not been worked out, but what must be deduced is the extraction of certain statements from property lists. Namely, the program first looks at *"want(at(I,airport))"* and attempts to copy the statements on its property list. Let us assume that it fails in this attempt because *"want(at(I,airport))"* does not have the status of an object and hence has no property list. (One might expect that if the problem of going to the airport had arisen before, *"want(at(I, airport))"* would be an object, but this might depend on whether there were routines for generalizing previous experience that would allow something of general use to be filed under that heading). Next in order of

increasing generality the machine would see if anything were filed under *"want(at(I,x))"* which would deal with the general problem of getting somewhere. One would expect that premises 6, (or 6' and 7'), 8, 9, 12, would be so filed. There would also be the formula

$$want(at(I,x)) \rightarrow do(observe(\text{where am } I))$$

which would give us premise 1. There would also be a reference to the next higher level of abstraction in the goal statement which would cause a look at the property list of *"want(x)"*. This would give us 14, 15, and 16.

We shall not try to follow the solution further except to remark that *"want(at(I,x))"* there would be a rule that starts with the premises *"at(I,y)"* and *"want(I,x)"* and has as conclusion a search for the property list of *"go(y,x,z)"*. This would presumably fail, and then there would have to be heuristics that would initiate a search for a y such that *"at(I,y)"* and *"at(airport,y)"*. This would be done by looking on the property lists of the origin and the destination and working up. Then premise 9 would be found which has as one of its premises *at(I,car)*. A repetition of the above would find premise 8, which would complete the set of premises since the other *"at"* premises would have been found as by-products of previous searches.

We hope that the presence of the heuristic rules mentioned on the property lists where we have put them will seem plausible to the reader. It should be noticed that on the higher level of abstraction many of the statements are of the stimulus-response form. One might conjecture that division in man between conscious and unconscious thought occurs at the boundary between stimulus-response heuristics which do not have to be reasoned about but only obeyed, and the others which have to serve as premises in deductions.

We hope to formalize the heuristics in another paper before we start programming the system.

REFERENCES

1. NEWELL, A. and SIMON, H. A. Empirical Explorations of the Logic Theory Machine. A Case Study in Heuristic. *Proceedings of the Western Joint Computer Conference*, p. 218 (February, 1957).
2. MINSKY, M. L. Heuristic Aspects of the Artificial Intelligence Problem. *Lincoln Laboratory Report 34-55*. (December, 1956). (See also his paper for this conference and his Princeton Ph.D. thesis).
3. McCARTHY, J. Inversion of Functions Defined by Turing Machines. In Automata Studies, *Annals of Mathematics Study Number*.
4. FRIEDBERG, R. A Learning Machine, Part I. *IBM Journal of Research and Development*, 1958, 2, No. 1.

DISCUSSION ON THE PAPER BY DR. J. McCARTHY

PROF. Y. BAR-HILLEL: Dr. McCarthy's paper belongs in the Journal of Half-Baked Ideas, the creation of which was recently proposed by Dr. I. J. Good. Dr. McCarthy will probably be the first to admit this. Before he goes on to bake his ideas fully, it might be well to give him some advice and raise some objections. He himself mentions some possible objections, but I do not think that he treats them with the full consideration they deserve; there are others he does not mention.

For lack of time, I shall not go into the first part of his paper, although I think that it contains a lot of highly unclear philosophical, or pseudo-philosophical assumptions. I shall rather spend my time in commenting on the example he works out in his paper at some length. Before I start, let me voice my protest against the general assumption of Dr. McCarthy - slightly caricatured - that a machine, if only its programme is specified with a suficient degree of carelessness, will be able to carry out satisfactorily even rather difficult tasks.

Consider the assumption that the relation he designates by "*at*" is transitive (page 81). However, since he takes both "*at(I, desk)*" and "*at(desk, home)*" as premises, I presume - though this is never made quite clear - that "at" means something like being-a-physical-part-or-in-the-immediate-spatial-neighborhood-of. But then the relation is clearly not transitive. If A is in the immediate spatial neighborhood of B and B in the immediate spatial neighborhood of C, then A need not be in the immediate spatial neighborhood of C. Otherwise, everything would turn out to be in the immediate spatial neighborhood of everything, which is surely not Dr. McCarthy's intention. Of course, starting from false premises, one can still arrive at right conclusions. We do such things quite often, and a machine could do it. But it would probably be bad advice to allow a machine to do such things consistently.

Many of the other 23 steps in Dr. McCarthy's argument are equally or more questionable, but I don't think we should spend our time showing this in detail. My major question is the following: On page 83 McCarthy states that a machine which has a competence of human order in finding its way around will have almost all the premises of the argument stored in its memory. I am at a complete loss to understand the point of this remark. If Dr. McCarthy wants to say no more than that a machine, in order to behave like a human being, must have the knowledge of a human being, then this is

surely not a very important remark to make. But if not, what was the intention of this remark?

The decisive question how a machine, even assuming that it will have somehow countless millions of facts stored in its memory, will be able to pick out those facts which will serve as premises for its deduction is promised to receive its treatment in another paper, which is quite alright for a half-baked idea.

It sounds rather incredible that the machine could have arrived at its conclusion - which, in plain English, is "Walk from your desk to your car!" - by sound deduction. This conclusion surely could not possibly follow from the premises in any serious sense. Might it not be occasionally cheaper to call a taxi and have it take you over to the airport: Couldn't you decide to cancel your flight or to do a hundred other things? I don't think it would be wise to develop a programme language so powerful as to make a machine arrive at the conclusion Dr. McCarthy apparently intends it to make.

Let me also point out that in the example the time factor has never been mentioned, probably for the sake of simplicity. But clearly this factor is here so important that it could not possibly be disregarded without distorting the whole argument. Does not the solution depend, among thousands of other things, also upon the time of my being at my desk, the time at which I have to be at the airport, the distance from the airport, the speed of my car, etc.?

To make the argument deductively sound, its complexity will have to be increased by many orders of magnitude. So long as this is not realized, any discussions of machines able to perform the deductive - and inductive! - operations necessary for treating problems of the kind brought forward by Dr. McCarthy is totally pointless. The gap between Dr. McCarthy's general programme (with which I have little quarrel, after discounting its "philosophical" features) and its execution even in such a simple case as the one discussed seems to me so enormous that much more has to be done to persuade me that even the first step in bridging this gap has already been taken.

DR. O. G. SELFRIDGE: I have a question which I think applies to this. It seems to me in much of that work, the old absolutist Prof. Bar-Hillel has really put his finger on something; he is really worried about the deduction actually made. He seemed really to worry that the system is not consistent, and he made a remark that conclusions should not be drawn from false premises. In my experience those are the only conclusions that have ever been drawn. I have never yet heard of someone drawing correct conclusions from correct premises. I mean this seriously. This, I think, is Dr. Minsky's point this morning. What this leads to is that the notion of deductive logic being something sitting there sacred which you can borrow for particularly sacred uses and producing inviolable results is a lot of nonsense.

Deductive logic is inferred as much as anything else. Most women have never inferred it, but they get on perfectly well, marrying happy husbands, raising happy children, without ever using deductive logic at all. My feeling is that my criticism of Dr. McCarthy is the other way. He assumes deductive logic, whereas in fact that is something to be concocted.

This is another important point which I think Prof. Bar-Hillel ignores in this, the criticism of the programme should not be as to whether it is logically consistent, but only will he be able to wave it around saying "this in fact works the way I want it". Dr. McCarthy would be the first to admit that his programme is not now working, so it has to be changed. Then, can you make the changes in the programme to make it work? That has nothing to do with logic. Can he amend it in such a way that it includes the logic as well as the little details of the programme? Can he manage in such a way that it works the way he does? He said at the beginning of his talk that when he makes an arbitrary change in the programme it will not work usually, and you try to fix that so that it will. He has produced at least some evidence, to me at least, that small changes in his programme will not obviously not make the programme work and might even improve it. His next point is whether he can make small changes that in fact make it work. That is what we do not know yet.

PROF. Y. BAR-HILLEL: May I ask whether you could thrash this out with Dr. McCarthy? It was my impression that Dr. McCarthy's advice taker was meant to be able, among other things, to arrive at a certain conclusion from appropriate premises by faultless deductive reasoning. If this is still his programme, then I think your defence is totally beside the point.

DR. O. G. SELFRIDGE: I am not defending his programme, I am only defending him.

DR. J. McCARTHY: Are you using the word 'programme' in the technical sense of a bunch of cards or in the sense of a project that you get money for?

PROF. Y. BAR-HILLEL: When I uttered my doubts that a machine working under the programme outlined by Dr. McCarthy would be able to do what he expects it to do, I was using "programme" in the technical sense.

DR. O. G. SELFRIDGE: In that case your criticisms are not so much philosophical as technical.

PROF. Y. BAR-HILLEL: They are purely technical. I said that I shall not make any philosophical criticisms, for lack of time.

DR. O. G. SELFRIDGE: A technical objection does not make ideas half-baked.

PROF. Y. BAR-HILLEL: A deductive argument, where you have first to find out what are the relevant premises, is something which many humans are not always able to carry out successfully. I do not see the slightest reason to believe that at present machines should be able to perform things that humans find trouble in doing. I do not think there could possibly exist a programme which would, given any problem, divide all facts in the universe into those which are and those which are not relevant for that problem. Developing such a programme seems to me to be by 10^{10} orders of magnitude more difficult than, say, the Newell-Simon problem of developing a heuristic for deduction in the propositional calculus. This cavalier way of jumping over orders of magnitude only tends to becloud the issue and throw doubt on ways of thinking for which I have a great deal of respect. By developing a powerful programme language you may have paved the way for the first step in solving problems of the kind treated in your example, but the claim of being well on the way towards their solution is a gross exaggeration. This was the major point of my objections.

DR. L. C. PAYNE: First a quick comment on the remark of no woman having ever brought up a child by means of deductive logic, the point surely is obvious. The feedback is very close: if she drops the baby in a disastrous way, she does not get another chance or she gets a great yelp. She learns very quickly by crude techniques of how to achieve precise control. There is direct feedback! If she is trying to win a spouse and tries a move which does not get the right response, she quickly changes her tack. Computer-wise, we have yet to develop an input (sensory system) and data-processing technique that can give even a gesture of such resourcefulness! It is a real-time trial and error process utilizing every bit of every nuance, quickly adapting and re-adapting.

A computer can deal with only a very small amount of information compared with the human brain, and therefore attention has to be concentrated on the efficiency with which this limited amount of information is handled. This is where one may usefully turn to deductive logic, because it will be appreciated that if a person is to benefit from all the studies and knowledge of many people in different places and epochs then synthesis of some sort is essential. Science in general is just this: it's laws subsume with great economy the mechanisms of diverse processes. For example the application of deductive logic to Newton's three laws allows us to treat of a multitude of practical applications. Hence if a computer is to have any range of activity, it must be fed with explicit rules, so that by rapid deductions or transformations of data, it can evolve a host of ramifications from a limited amount of information.

The Countess of Lovelace remarked that "a machine can originate nothing: it can only do what we order it to perform". The essence of my contention is that we can only order to perform by means of transformations of data

having as their basis existing logical systems. Because of this it might be well to summarize, perhaps boldly before an audience like this, what I think is a summary of existing logics. The first is Formal Logic and consists of statements of the form, if all A are B and C is A, then certainly C is B - the syllogistic type of statement by which one can establish direct connections. This sort of logic is reversible; that is, if you start from a given complex of consistent propositions, then it is possible to take some selection of a derivative statements and from them as a starting complex, derive propositions of the original set. If the original set contains a contradiction then all other contradictions are implied latently. The best example I know of this is one recently cited by Sir Ronald Fisher, which is said to stem from the high table at Trinity College, Cambridge. The late Professor G. H. Hardy was asked, "Do you mean to say, Hardy, that you can prove any contradiction whatsoever if you have got one contradiction?". Hardy replied, "Yes, that is so". The questioner went on, "Well, four equals five, you prove that McTaggart is the Pope". Hardy rejoined at once saying, "If four equals five, then by subtracting three from each, one equals two. McTaggart and the Pope are two, therefore McTaggart and the Pope are one"!

The other important logic is Probability Theory. This consists of statements, that if some well defined proportion of A are B, and if C is A, then only an uncertain inference in the form of a probability statement, can be made about C being B. One has to be especially careful in statements of this kind to see that the total reference set is well defined and also the sub-set having some specified attribute. This kind of consideration, treated very carefully by Sir Ronald Fisher in his "Statistical Methods and Scientific Inference", nullifies the casual attitude which, to instance an example, can remark that, "statistics means that if you take enough inaccurate statements and put them together then a more accurate statement can be made". The well defined nature of the statistical mode of reasoning, if it is respected, means you can be as logically precise as with Formal Logic, but that the kind of statement you can be logically precise about is less certain, that is, it is a probability statement.

A more restricted logic can be based on what Sir Ronald Fisher calls "mathematical likelihood". This allows quantitative statements to be made on the fullest information available; it is discussed in the reference already given.

Beyond these systems one is very suspicious of the play with random exercises which purport to produce something out of nothing. It seems to me that computers can do nothing beyond applying the existing logics to effect transformations of data, since these are the limits within which exercises can be prescribed explicitly. These limits in fact are very wide and circumscribe most of the rational procedures used by human beings. They are not limited by pure mathematics, where one is constrained to using a

limited class functions which lend themselves to analysis. Numerical solutions can certainly explore regions which would bog down a more ponderous mathematical attack. In my opinion higher mathematics can be logically precise and very penetrating only about a very small class of entities, ones that are very abstract in content. Statistics allows one to deal with a wider class with less certainty, and so on down the scale until you reach common sense, where one may be rational about fairly concrete entities. Between common sense and high mathematics one has the whole range of human rationality, but for each refinement in logic one must pay the price of dealing with more restricted classes of entities which become progressively more abstract.

DR. J. McCARTHY (in reply): Prof. Bar-Hillel has correctly observed that my paper is based on unstated philosophical assumptions although what he means by "pseudo-philosophical" is unclear. Whenever we program a computer to learn from experience we build into the programme a sort of epistemology. It might be argued that this epistemology should be made explicit before one writes the programme, but epistemology is in a foggier state than computer programming even in the present half-baked state of the latter. I hope that once we have succeeded in making computer programs reason about the world, we will be able to reformulate epistemology as a branch of applied mathematics no more mysterious or controversial than physics.

On re-reading my paper I can't see how Prof. Bar-Hillel could see in it a proposal to specify a computer programme carelessly. Since other people have proposed this as a device for achieving "creativity", I can only conclude that he had some other paper in mind.

In his criticism of my use of the symbol "*at*", Prof. Bar-Hillel seems to have misunderstood the intent of the example. First of all, I was not trying to formalize the sentence form, A is at B as it is used in English. "*at*" merely was intended to serve as a convenient mnemonic for the relation between a place and a sub-place. Second I was not proposing a practical problem for the program to solve but rather an example intended to allow us to think about the kinds of reasoning involved and how a machine may be made to perform them.

Prof. Bar-Hillel's major point concerns my statement that the premises listed could be assumed to be in memory. The intention of this statement is to explain why I have not included formalizations of statements like, "it is possible to drive from my home to the airport" among my premises. If there were n known places in the country there would be $\frac{n\ (n-1)}{2}$ such sentences and, since we are quite sure that we do not have each of them in our memories, it would be cheating to allow the machine to start with them.

The rest of Prof. Bar-Hillel's criticisms concern ways in which the model mentioned does not reflect the real world; I have already explained that this was not my intention. He is certainly right that the complexity

of the model will have to be increased for it to deal with practical problems. What we disagree on is my contention that the conceptual difficulties arise at the present level of complexity and that solving them will allow us to increase the complexity of the model easily.

With regard to the discussion between Prof. Bar-Hillel and Oliver Selfridge - The logic is intended to be faultless although its premises cannot be guaranteed. The intended conclusion is "*do(go(desk,car,walking))*" not, of course, "*at(I,airport)*". The model oversimplifies but is not intended to oversimplify to the extent of allowing one to deduce one's way to the airport.

Dr. Payne's summary of formal logic does not seem to be based on much acquaintance with it and I think he underestimates the possibilities of applying it to making machines behave intelligently.

SESSION 1

PAPER 4

THE MECHANISM OF HABITUATION

by

DR. W. ROSS ASHBY

BIOGRAPHICAL NOTE

Dr. W. Ross Ashby, born in London, studied medicine at Cambridge and London, took M.B., B.Ch. in 1928, M.D. in 1935.

He has since been engaged in research in psychiatry, specialising in the application of generalised dynamic principles (equilibrium, homeostasis, self-repairing systems).

His publications include Design for a brain (*Chapman & Hall, London; John Wiley & Sons, New York;* 1954) and An introduction to cybernetics (1958).

His present interests are: study of complex equilibria, especially in their topological aspects, as applied to the intelligent and adaptive aspects of the brain.

He is now in the Department of Research of Barnwood House Hospital, Gloucester.

THE MECHANISM OF HABITUATION

by

DR. W. ROSS ASHBY

SUMMARY

THE phenomenon of habituation, in which the response to any regularly repeated stimulus decreases, has not so far received any general mechanistic explanation. Its occurrence in systems of widely differing nature (nervous system, protozoa, bacterial cultures) and its occurrence against stimuli that the species has never previously encountered (smells or toxicities of newly synthesised chemicals) show that it cannot always be due to a specialised mechanism, shaped by natural selection.

It is here shown that when any system is subjected to a regularly repeated stimulus or disturbance, the successive responses, if they change in size, do not in general tend to become larger or smaller with equal probability: there is a fundamental bias in favour of the smaller. This bias holds over a great range of systems and disturbances. A theorem is proved, setting out the exact conditions for the decrease.

Some applications of the theorem are given. It is shown that habituation is specially likely to occur in any system made of parts that are rich in states of equilibrium. The theorem, without further assumption, provides an explanation of why any extraneous disturbance typically causes de-habituation.

To demonstrate the generality of the theorem, an example is given showing how habituation and de-habituation appear even out of a table of random numbers when the appropriate operations are applied.

It is suggested that this asymmetry is responsible for much habituation, including much of that shown by the cerebral cortex.

---

1. INTRODUCTION

ONE of the commonest phenomena in the living world occurs when an organism, given an innocuous and unvarying stimulus or disturbance repetitively, reacts briskly at first, then less actively, and finally perhaps not at

all. Known in behaviour as "habituation" and in perception as "adaptation", it has been recognised from time immemorial yet still lacks explanation. Only recently Sharpless and Jasper (1956, *ref. 10*) could say "Habituation ... has yet to be explained by any known neurophysiological principles".

A review of the subject need not be given here as it has been well reviewed by Humphrey (1933, *ref. 6*), Harris (1943, *ref. 5*), and Thorpe (1956, *ref. 11*). On one important matter they are agreed: habituation of typical form occurs in almost every form of life; in particular it appears as readily in forms having no neural apparatus as in the forms having a well developed brain. *Amoeba* shows it as freely as does the cat. *The phenomenon evidently does not depend on specifically neurophysiological details.* Its origin must lie in some property of much wider occurrence.

The possibility of "fatigue" as an explanation must be rejected. A number of workers (e.g. Humphrey, 1933, *ref. 6*; Danisch, 1921, *ref. 4*) have shown that the more violent the stimulus (with correspondingly larger response) the less does habituation occur and the later is its onset. This relation is just the opposite to what would be expected to happen with fatigue, in which the larger responses would lead to a more rapid decay.

The aim of this paper is not, in any case, to relate habituation to concepts of physiological or psychological type such as "fatigue" but to relate it to basic concepts of mechanistic type. I shall examine the phenomenon in the light of the modern logic of mechanism, so as to make use of the full generality of its methods (Ashby, 1956 *ref. 2*). Its use has several advantages. It enables the discussion to be as rigorous as we please. Yet this rigour is coupled with an extreme generality; for while it makes no assumption (as mathematical analysis so often does) of continuity, it includes this possibility as a special case. Further, by being abstract, *general* concepts need not be artificially restricted to those covered by the vocabulary of the neurophysiologist, or the electronic engineer, or the colloidal chemist, or other specialist. As will be shown below, the basic phenomenon of habituation can be identified over a very wide range of systems, and only a language that can range equally widely is appropriate.

Habituation was originally a physiologist's concept. To discuss it abstractly we must re-define it in abstract terms. It can be defined at several degrees of strength. The weakest degree ($\underline{H}_1$) requires only that the response to the later stimulus is *smaller* than that to the earlier. (Such is the form shown by the homeostat; *ref. 1* S.13/8). A stronger degree ($\underline{H}_2$) demands $\underline{H}_1$ and also that the change from the initial large response to the terminal small response shall pass through a *monotonic* sequence of intermediate sizes. An even stronger degree ($\underline{H}_3$) demands $\underline{H}_2$ and also that the fall shall be of approximately *exponential* form, with a rapid fall at first, flattening out. Related to $\underline{H}_1$, though not identical with it, is

degree $\underline{H}_4$: the *maximal* later responses are consistently less than the *maximal* earlier.

At first we shall be concerned with habituation only in the weak degree of $\underline{H}_1$; in Section 4 we will consider the stronger H_2 and $\underline{H}_3$. $\underline{H}_4$ is considered in Section 3.

Before we enter the arguments of Sections 2 to 4 it should be noticed that most of the arguments and implications are not reversible -- what is necessary is not usually sufficient, and *vice versa*. The implications in the two directions will be given in separate Sections.

SECTION 2

In this Section it will be shown that there is a fundamental asymmetry in mechanism, such that over a very broad class, containing most biological systems, there is a tendency for the response to a repeated stimulus to get smaller rather than bigger.

To see how this is so, let us start at the beginning and assume that the experimenter has before him some unknown system, totally unrestricted in nature; he is going to test whether it shows habituation. We will specify in detail the assumptions that he usually makes and the operations that he usually performs. We shall formulate these as abstractly and precisely as possible, taking special care to state explicitly those assumptions that he seldom mentions but that are usually taken for granted. We shall then see that over a wide class of system, general in the sense that it contains no *ad hoc* mechanism for habituation, he is certain to find habituation of degree $\underline{H}_1$.

(1) First to be defined is the operation Ω that the experimenter will apply to the system to test it for habituation.

Postulate I; The operation Ω is the following sequence of sub-operations:

(1) The experimenter allows the system to come to rest.
(2) He applies the given stimulus or disturbance.
(3) He allows the system, without further disturbance, to display its response, which he records.
(4) He repeats the cycle (1)-(2)-(3)-(1)- etc. until the responses become invariant (or approximately so) at the "terminal" response.
(5) He defines a numerical scale on which to measure "size of response".
(6) He compares the size of the terminal response with that of the initial.

Each of these sub-operations, as we shall see below, is significant.

(II) Next we must take account of the fact that the experimenter does not typically apply the test for habituation to a system that is unpredictable or chaotic in its behaviour. He restricts his tests to systems that are known to be law-abiding, or that may reasonably be assumed to be so. Formally:

Postulate II: The system under test is a machine with input (*ref. 2* S.4/1), i.e. such that its present state and its present surroundings (or values at its input) determine uniquely the state it will go to next. Abstractly, there is a set E of its states and a set A of its conditions or input-values, and its behaviour corresponds to a single-valued mapping, T, of the product set $A \times E$ into E. T specifies what the machine does, how it behaves.

If the system is law-abiding in the probabilities of its transitions rather than in the transitions individually (i.e. if it is a Markovian machine; *ref. 2* S.12/8) the deductions below would have to be re-stated in probabilistic form. Only modifications in detail would be needed; as they are almost obvious I need not give them here as they would only obscure the theme.

The sub-operations of Ω can now be converted to algebraic form (a necessity for rigorous discussion) as follows (with the numbering of Postulate I):-

(1) "Let the system come to rest" means allowing U to operate, where $U = \lim_{n \to \infty} T^n$, and where $T^2(x) = T(T(x))$, etc. If T has no cycles, only states of equilibrium, U maps E into Q, where Q is the set of T's states of equilibrium, i.e. such states x as satisfy $T(x) = x$.

(2) A stimulus or disturbance D, acting impulsively, displaces x to a well-defined state $D(x)$. Thus D maps Q into E, if D operates only when the system has come to rest under T.

(3) The return to rest after D implies the operation of U.

(4) The triple (2)-(3)-(1) is the operation $U \circ D$ a composite operator (Bourbaki, 1951, *ref. 3*), such that $(UD) \circ (x) = U(D(x))$; call it Σ. Σ maps Q into Q, if applied only when the system has come to rest under T. For the responses to become invariant (if Σ has no cycles) the system must have reached a state q such that $\Sigma(q) = q$, i.e. a state of equilibrium under Σ.

(III) Next, the experimenter does not usually test for habituation a system that is undergoing some obtrusive cycle of activity. Formally:

Postulate III: The transformation T is assumed to have no cycles, only states of equilibrium (ref.2 S.5/4).

Another reason for stating this postulate explicitly is because Rubin and Sitgreaves (1954, *ref.9*) have shown that if the set of machines obtained by taking all possible mappings of E into E is taken as a sample space (so that its elements are those of the set E^E -- Bourbaki, 1951, *ref.3*) then as the number of states in E increases, so does the probability of any trajectory ending in a *state* of equilibrium tend to zero. (It becomes almost certain that the trajectory will end in a cycle). If therefore we want to talk of machines with some generality yet wish the class discussed to have equilibrial states rather than cycles, we must make clear that we are speaking of some subset of the class of all machines. The assumption is made here simply because the experimenter usually restricts himself to tests on such a subset.

(This Postulate, and those that follow, are not intended to dogmatise about what will be found empirically in the world of real systems, but simply to state precisely what is being discussed in this paper).

(IV) Next, the experimenter would reject as peculiar any system that produced, under the cycle of suboperations (1)-(2)-(3)- of Ω, an invariant *cycle* of responses R_1-R_2-R_3-R_1- etc., rather than a single response repetitively. So, formally:

Postulate IV: Σ is assumed to have no cycles, only states of equilibrium.

Now Σ is a composite operator, for $\Sigma = U \circ D$, where U is $\lim_{n \to \infty} T^n$. For a state q to be one of equilibrium under a composite operator, a *relation* must exist between how D and U affect it. This relation is the point of the paper. What it is will first be sketched picturesquely, and then formally. (The picturesque statement will show its obvious relation to habituation H_1, but leaves obscure exactly what is being assumed; this fault will be corrected in the formal statement).

Let the system's states E be represented by the points within the area of *fig.1*. For clarity, group together all those states that come to the same state of equilibrium (under T); such a set is a "confluent". The whole set E is thus partitioned into confluents, each containing one state of equilibrium, (shown in the *figure* as a heavy dot). The illustration shows nine states of equilibrium each surrounded by its confluent. From each state of equilibrium draw an arrow to show how the representative point will be displaced when the stimulus or disturbance D acts. (An arbitrary set has been drawn in the *figure*).

Now by merely tracing the sequence of events, it is easily verified that the system, if started from any point in the left-hand two-thirds, must end in either the top left or bottom left confluent. Thus, if started

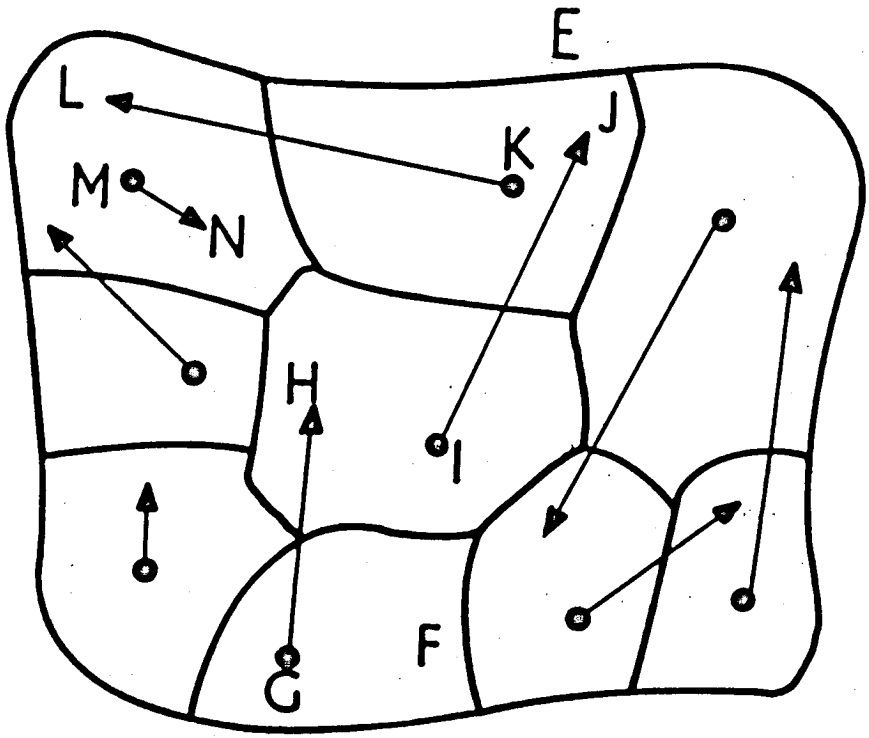


Fig. 1.

at *F*, suboperation (1) of Ω (i.e. *ν*: "let the system come to rest") will take it to *G*; the stimulus *D* then takes it to *H*; then "come to rest" takes it to *I*; and so it goes on, changing from confluent to confluent. When it gets to *K*, however, it is trapped. *N* does not take it outside *M*'s confluent, so it comes back to *M*. Thus the point is trapped in the first confluent such that *D* cannot move it out. Clearly, confluents with short arrows, such as that from *M* to *N*, are more likely to be terminal than those with long ones, such as that from *K* to *L*. So the sequence of arrows, caused by *D*, has a tendency to finish on an arrow *of less than average length*.

Thus the very process of *testing* for habituation, of applying Ω, so acts on the system that by the time the invariant response has been elicited the system's state is no longer an average one but is one such that *D* displaces it less than averagely. Thus, to say "I tested this system for habituation, and I eventually found it to show a diminished response" is to verge on the tautologous.

(Either the representative point ends at such a confluent as M's, or it must enter a cycle made by the disturbance, as on the right of *fig. 1*. The latter possibility has been excluded by Postulate IV).

The necessary theorem will now be stated and proved formally. It will be stated in terms of arbitrary and discrete states, for in this form it has simplicity and the greatest possible generality. It must not be thought, however, that the theorem is true only for systems, such as digital computers, that change by finite jumps; on the contrary, the states may be as close as is wished, and may thus, in the limit, represent the values of continuous variables. (It should also be noticed that the states are not here analysed into components, i.e. the whole system is not regarded as being made of parts; the complication is avoided in this Section as it is not necessary).

Definitions and postulates

(1) A set E of states x is mapped into itself by a single-valued transformation T (Postulate II).
(2) Q is the set of states in E that satisfy $T(q) = q$.
(3) T has no cycles (Postulate III).
(4) D is a single-valued transformation mapping Q into E. (The transition from x to $D(x)$ will be called the "displacement").
(5) U is the transformation $\lim_{n\to\infty} T^n$, mapping E into Q.
(6) Σ is the transformation $U \circ D$, so that $\Sigma(x) = U(D(x))$; it maps Q into Q.
(7) Σ has no cycles (Postulate IV).
(8) Given a state q in Q, the "T-confluent containing q" is the set of all states in E obtainable from q by repeated application of T^{-1}. (It consists of all those states in E that come, under T, finally to q).

Theorem: For a state x to be equilibrial under Σ, it is necessary and sufficient that x and $D(x)$ lie in the same T-confluent.

Proof: (1) Assume x is equilibrial under Σ. Then $\Sigma(x) = x$, and $U(D(x)) = x$, by definition. Now if $U(a) = b$, a must lie in the same T-confluent as b; so $D(x)$ and x must lie in the same T-confluent.

(2) Assume x and $D(x)$ lie in the same T-confluent C. D operates only on states in Q, so x must be a state of equilibrium under T, and it lies in C. $U(D(x))$ is in the same confluent as $D(x)$, so $\Sigma(x)$ lies, with $D(x)$, in C. Σ maps into Q, so $\Sigma(x)$ is also a state of equilibrium under T. But any confluent can contain only one state of equilibrium; so $\Sigma(x) = x$, and x is equilibrial under Σ. (Q.E.D.)

(V) So far, no topology or metric has been assumed over the states, for such an assumption is not necessary within the theorem. In order, however, to introduce the concept of "size", required by (5) of Postulate I, we must

now suppose that some metric, some measure of "distance", holds over the states, so that states have the relation of being "near to" or "far from" each other. (In most practical cases, of course, some metric presents itself as being the obvious one; often it is that in which the system is describable, in three dimensions of space, in units of mass, length, and time).

Given the metric, we must next say, explicitly, how the sizes of the displacements (from x to $D(x)$) are related to the sizes of the responses. We naturally assume that if the displacement is zero the response will be zero (for the system has not undergone any real change), and that an increase in the displacement will show as an increase in the size of the response. Formally:

Postulate V: The size of the response is a positive monotonic function of the size of the displacement, the zeros corresponding.

The point of the theorem (so far as we are concerned) is that whereas the initial response is based on *any* displacement (e.g. on any of the arrows in *fig. 1*), the terminal response can be based only on some displacement that leaves the point inside the confluent. If "in the same confluent as "x" has any implication of "near to x", then the terminal response will tend to be less than the initial, and some habituation (of degree H_1) will tend to occur.

For rigour, the word "tend" requires definition. There are obviously various ways of stating the relation, of various utilities in various applications, and the reader may prefer to formulate his own way. Whichever way is used must be compatible with the fact that the Postulates given so far do not allow the deduction that the terminal response *must* be smaller than the initial (for arbitrary T and D); for nothing prevents the initial displacement from being small yet taking the representative point to another confluent, and then a later large displacement leaving the point in the same (later) confluent. Clearly, our interest is in the fact that the terminal displacements are a subset of the total displacements, biassed in favour of shortness. One way of expressing the relation rigorously is as follows.

To compare the distributions of the initial and terminal displacements, let a "typifying" function f be defined (averaging, taking the mode, taking the maximum, etc.) which maps the set of distributions into the scale of sizes of response, so that each distribution indicates its "typical" size of response. Now if the diameter (topologically) of the largest confluent is l, every displacement in the terminal distribution must be less than l in length. If now, over a sequence of systems (over a sequence of T's) l tends to zero, the terminal displacements must all tend to zero also, and so therefore must the size of the typical response. So whatever the size

of the typical response initially (provided it is non-zero), the typical terminal response will become less than it, and habituation of degree H_1 will be shown.

(This demonstration that there is a fundamental *asymmetry* involved when a composite operation leads to equilibrium is the point of the paper. The theorem, however, leads on to several interesting and easy applications; the remainder of the paper will be concerned with them).

SECTION 3

In the previous Section, we assumed that the experimenter applied the operator Ω to *test* for habituation, and we saw that a very general class of systems, not characterised by possession of any specialised mechanism, would show it. The class was characterised chiefly by having "in the same confluent as x" correlated with "near x". The relation of this property to habituation is further illustrated by consideration of it in the inverse direction. (To make sure that the argument is not circular, and to make sure that we are keeping what is necessary distinct from what is sufficient, we make here a new start).

Starting, then, *ab initio*, let us consider what we can deduce about a system when all we know of it is that it has been tested for habituation and has been found to show it.

This statement must imply that the operator Ω, defined as before in Postulate I, has been applied to it as test. That the test was made implies that the system must have been behaving with some regularity; so Postulate II -- that the system is a machine with input -- is evidently applicable. It implies that the system showed no important cycles when at equilibrium (Postulate III); and it implies that the response to Σ gave a single response rather than a cycle (Postulate IV). (By the preceding Section we could now deduce that *if* T had l less than some value, etc.; but this is not our direction. It is now *given* that the system shows habituation, and we want to know what we can deduce about the particular system given).

Nothing has been said so far about whether the habituation shown by the system is against a particular disturbance D, or against all possible disturbances D_1, D_2, D_3, ... etc. from some class. The two cases will be considered separately.

The first case can be dismissed briefly. It is similar, in the postulates that apply to it, to that discussed in Section 2; and as these Postulates imply that it is likely to show habituation, the information that it does tells us little. So little can be said in this case about T.

The second case, when the system habituates against *any* of a set of D's, is more interesting. This is what is implied when it is said that the

Protozoa show habituation freely, with no specification of the disturbance to be used; for the statement implies that habituation will be shown against any of a wide class of disturbances. An interesting deduction can be made when the habituation is of degree H_4: when all the terminal responses have size less than λ, where λ is much less than the maximal size of the initial responses. (H_1 and H_4 are hardly independent, but will be so treated here, for the sake of rigour). Our assumption is now, formally:

Postulate VI: The system shows habituation (H_4) against each of a set of disturbances that cause, between them, all possible displacements from each state of equilibrium.

We use the fact that if a system is such that all its terminal responses are smaller in size than λ, and if the typifying function f is such that the limitation to λ implies that every displacement is through some distance less than l', then Postulate IV (with the others of this Section) implies that *no confluent in T. can* have a diameter exceeding l'. (For suppose one confluent had a greater diameter; then among the set of displacements would be one (by Postulate VI) longer than l' and lying within the confluent; it would give rise to a response that was both terminal and larger than λ, contrary to hypothesis). So if a system shows habituation H_4 against a wide *class* of disturbances, then, by the theorem, this is evidence that the system's T-confluents are all small. Since, if the number of states does not vary, smallness of confluents implies largeness in their number, the confluents must evidently be numerous. And as each confluent must have a state of equilibrium, it follows that the system's states of equilibrium must be numerous too.

From this fact we can draw an inference about the parts that compose the whole. So far no reference has been made to any parts of which the whole might be made. The "states" referred to so far have been those of the whole system, assumed to be identifiable as such without having to be built up from components. Statements about equilibria in the whole, however, have implications about those in the parts, and *vice versa*. The basic relationship has been described in *ref.2* S.5/12: the whole is at a state of equilibrium if and only if each part is at a state of equilibrium in the conditions provided by the other parts. (The next Section takes up the subject in more detail). Thus if the whole is in equilibrium at a certain state, any particular part must also be in equilibrium at its component-state. So if, over a given sample-space of states, the whole has probability p of being at equilibrium, then the probability π that any given part is at equilibrium cannot be less than p. As we saw above, p is high; so *the π's must be higher.*

This Section thus shows that if a system has been tested for habituation against the set of all possible disturbances and has been found to habituate (H_4) to all, then, with the stated minor qualifications, the

theorem implies that, if made of parts, *the individual parts must have a high proportion of their states equilibrial.*

SECTION 4

As a second application, we can now consider the converse: if a whole is built of parts having a high proportion of their states equilibrial (but not otherwise restricted), will it show habituation when tested by the operation Ω?? (As was said earlier, this converse can by no means be taken for granted).

Let us first make precise what is meant by a part being "rich in states of equilibrium". As we are discussing law-abiding, and not chaotic, parts, each part i will itself be a machine with input (*ref.2*, S.4/1) and will therefore itself behave in accordance with some mapping V_i of the product set $G_i \times B_i$ into B_i, where B_i is its set of possible states and G_i its set of possible input states. When G_i is at a particular state, at g say, there is defined the partial mapping V_{ig} of B_i into B_i generated by V_i and corresponding to g (Bourbaki, 1951, *ref.3*). Then if b is an element in B_i, b is "a state of equilibrium for the input state g" if and only if $V_{ig}(b) = b$. (This definition implies that, if the system is composed of continuous parts or variables y_1, y_2, ..., y_n, behaving in accordance with equations of form

$$dy_1/dt = \phi_1 (y_1, \ldots, y_n)$$

$$\ldots\ldots\ldots\ldots$$

$$dy_n/dt = \phi_n (y_1, \ldots, y_n)$$

then part i is in equilibrium at the state $(y_1, \ldots, y_i, \ldots, y_n)$ if and only if $\phi_i (y_1, \ldots, y_i, \ldots, y_n) = 0$).

For a part i to be "rich" in states of equilibrium, it is implied that the elements in $G_i \times B_i$ satisfying $V_{ig}(b) = b$ are numerous. (Nothing is implied about how they shall be distributed over $G_i \times B_i$). So if the set $G_i \times B_i$, or the domain of the ϕ_i's, is taken as the sample space, a probability can be defined -- that part i should be in equilibrium.

As was said earlier, equilibria in whole and part are related. Thus arises the possibility of relating the richness in equilibria of the parts with that of the whole. Unfortunately, a full treatment of the relation requires attention to how the states are distributed in each part over $G_i \times B_i$, for each i. These combinatorial complexities make the subject hardly worth full treatment here, though the following fact will be required. Suppose each part has a fraction π of its states equilibrial, and that n such parts are coupled (*ref.2*, S.4/6). It can be shown that the whole's fraction p (of states that are equilibrial) can vary from

$1-n$ $(1-\pi)$ (from zero if this function is negative) up to π, according to how the equilibrial states are disposed in the product sets (in the canonical representations) of the parts. A value of p outside this range is combinatorially impossible. If π is near 1, and n is large, this is a broad range, and implies little about the value of p. On the other hand, however large the value of n, a value of π near enough to 1 can force p to be arbitrarily near to 1.

After this preparation, let us start *ab initio* again and assume

Postulate VII: The whole is made of parts each with π so high that the whole has p near to 1, but not otherwise restricted. It is tested for habituation (thus implying the application of Postulates I to V) -- what will be found?

As π tends to 1, so does p. The number of confluents rises, and the average number of states per confluent falls towards one. So the restriction of the terminal displacement to within a confluent (by the theorem) makes its size tend to zero. Whatever the average (or other typical) size of the initial displacement, that of the terminal displacement will become less than it; so the appearance of habituation (H_3 and H_4) certain on the average. Thus, *a whole made of parts rich in equilibria tends to show habituation.* (This result is true even for an arbitrary T-D pair, provided that responses sufficiently small to be terminal exist in D's set of displacements -- i.e. providing D is not everywhere too "strong" in its effects).

Systems made of parts that are rich in equilibria have further interesting relations to habituation. A case of special interest occurs when the parts are not related to one another in some specially arranged way -- when they have been selected, say, as random samples from a distribution of parts, coupled in a way also sampled from a distribution of ways of coupling. Let us then consider the case:

Postulate VIII: In each part the states of equilibrium are distributed independently of one another and of how the equilibria are distributed in the other parts.

As the richness increases (as π tends to 1) the whole tends to be cut into functionally independent subsystems (*ref.1*, S.14/15; *ref.2*, S.4/10); for as the parts on any trajectory, spend more and more time not changing, so do they become unable to transmit variety from part to part within the system.

Suppose now, as is not uncommonly the case, that the disturbance D is actually a vector with components $D = (\delta_1, \delta_2, \ldots, \delta_n)$, so related to the n parts of the system that each component of D acts on some part of the system, so that, e.g.

$$D(y) = (\delta_1(y_1), \ldots, \delta_n(y_n)).$$

This *Postulate IX* is in one sense a simplifying assumption, though it seems more complex algebraically. We are now assuming that how D affects part y_1, say, depends only on the state of y_1, and not also on the states of all the other parts. This means that D affects the whole only so far as each part is affected individually by D. (Often many of the δ's are identity operators, so that D has a non-zero effect only on certain parts; i.e. so that D affects directly only a portion of the whole system).

It is now possible (and the phenomenon will become more evident as π tends to 1) for local regions of the whole to reach equilibrium under Σ while other regions are being changed by it; and they may, for some time, be able to retain their state of equilibrium. Then a final state of equilibrium under Σ (and the invariant terminal response) can be arrived at by degrees, by accumulation of local equilibria (*ref.1*, S.12/5; *ref.2*, S.4/21). If now the system allows

Postulate X: The size of the response is a positive monotonic function of the number of parts changing after each application of Σ, then in the limit (as π tends to 1) the number of parts not in equilibrium can only fall, and the system will thus show the stronger (H_2) degree of habituation.

Finally, if the subsystems go to equilibrium *independently*, so that the number that go is some fraction (approximately constant) of the number of parts still not at equilibrium, then the number not there, and the size of the response, will fall in approximately exponential fashion, thus showing *habituation in the full degree of* H_3.

SECTION 5

The propositions of the previous three Sections have shown that habituation not only can appear but must appear in classes of systems much wider than those with specifically neurophysiological properties. They do not even need to be living, for any system that satisfies the Postulates will show it. In fact, the systems that show it are so general that the process is readily demonstrable on systems formed from random numbers, - a "Monte Carlo" method - all that is necessary being that one must try to adjust the process so that it is neither so simple as to be trivial nor so complex as to be excessively laborious in computation. Here is one that shows the phenomenon readily.

A "part" or "variable" of ten states is given its value by taking a number from a defined place in a table of random numbers. The digit in that place gives the value, and its successive transforms are found by the digits that occur successively below it in the table. Four such variables form a "whole" with 10,000 states. Thus if the table shows 3119 we
8292

interpret it as defining a T-transformation in which state 3119 is changed to state 8292; also that part y_1, at state 3 in conditions $(y_2, y_3, y_4) = (1,1,9)$, changes to 8. (Should 3119 occur again, it must, of course, transform to 8292, not to the number that follows it on the second occasion; but this complication occurs rarely in the process defined below and can be ignored without appreciable error). Thus a T-transformation can readily be defined, devoid of any special structure.

Such a T, however, has on the average only one state of equilibrium, while the T we require must have a fairly high value of p. So the definition is modified. p is made 0.2401 by assuming that all states are stable if they contain no 0,1 or 2. Then the whole progresses to an equilibrium simply by going from state to state (of four digits) until a state occurs having no 0,1 or 2. Thus the start of Kendall and Babington Smith's (1951, *ref. 7*) Table gives

2315
0554
1487
3897

The last state is equilibrial. Thus is defined a T-transformation specialised only in that $p = (0.7)^4 = 0.2401$.

A suitable definition for D was found to be to restrict it to δ_4, and to act on (y_1, y_2, y_3, y_4) so that y_4 was changed by the transformation:

$$D = \delta_4: \quad \begin{matrix} \downarrow 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ 0 & 4 & 0 & 6 & 0 & 8 & 0 \end{matrix}$$

(When at T-equilibrium, values 0,1,2, do not occur). So D, acting on 3897 gives 3890, which is not in equilibrium, and so starts a new trajectory under T;

3890
9731
1174
4336

The last state is an equilibrium for T and for D; it is thus also an equilibrium for Σ.

The response to each application of D is measured by the number of steps taken before equilibrium was restored (corresponding, in *fig. 1*, to the distance from the head of the arrow to the next ensuing equilibrium). Thus the example gave, for its successive responses: 3, 3, 0, 0, ...

The process so far illustrates those of the theorem in Section 2. To show habituation of degree H_3, the processes of Section 4 were followed. A hundred initial states were taken (corresponding to a large system, with many more than 400 parts, being disturbed by D in a hundred places), and to each was applied the operation Ω. The grand "Initial Response" was taken as the average of the hundred individual initial responses; the grand

"Second Response" was taken as the average of the hundred individual second responses; and so on. These average (grand) Responses correspond to what would be observed if a system, much divided into independent local subsystems by a high value of π, were subjected to a D that affected a hundred of them locally, producing a Response that was some function (such as the sum, or average) of the hundred individual responses.

The result is shown on the left half of *fig. 2*. Each column shows the size of the average Response to D, as Σ is applied again and again.

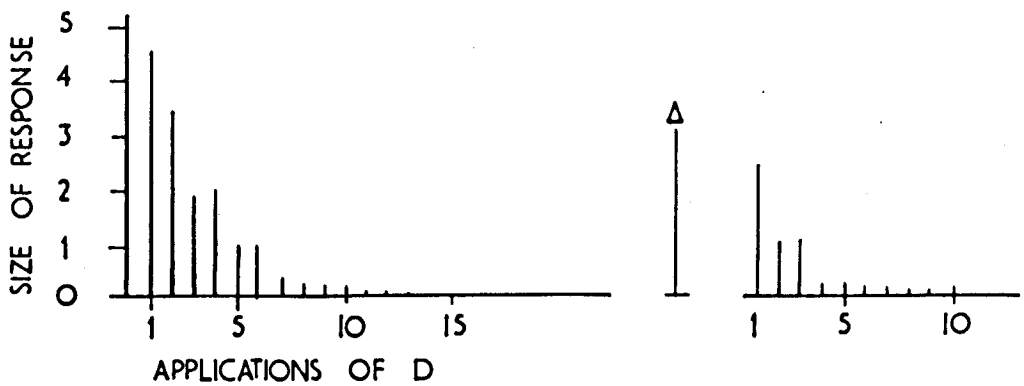


Fig. 2.

It shows typical habituation of degree H_3: the average Response to the first application of D is 4.3; to the second 3.5; and the Responses diminish eventually to zero.

SECTION 6

As ubiquitous as habituation is de-habituation: the fact that a system, habituated to a monotonous stimulus, if given any new stimulus or disturbance, will so change that when the monotonous stimulus is given again it evokes a response greater than those it had shown at the end of the habituation.

This phenomenon occurs widely over the biological world, and is also remarkable in that the interrupting stimulus or disturbance may be of almost any nature. Clearly, in most cases some very general type of process or mechanism is involved. So far, no process or mechanism has been

recognised as being responsible for it. A few explanations have been proposed for particular systems (e.g. an inhibition of an inhibition in the neurophysiological) but these are hardly satisfactory when the same phenomenon appears in systems of quite different type.

If habituation should often be due to the process of Section 2, a reason is readily seen why de-habituation should occur as often. To translate to abstract form, the new stimulus will be a new operator, Δ, mapping the equilibria of Σ into E. That it is "strange" means that Δ (as a set of arrows in *fig. 1*) will be arbitrarily different from D; that it is "strong" means that the arrows will not be so short as to make its effect negligible. Suppose then that the system has become habituated under the repeated action of D (has reached the top left confluent in *fig. 1*, say). An arbitrarily new disturbance Δ is applied -- what will happen?

Δ will displace the system (the representative point) from M to somewhere in the set E of possible states. Should the point fall within M's confluent, the subsequent application of D will evoke only the habituated, small, response; but should the point fall outside the confluent, then the operator Δ has undone the going to equilibrium of Σ, and the new response to D will be, in general, one from the initial, unrestricted, distribution of responses. Thus any new operator Δ does not act symmetrically on the terminal responses: *it tends to free the response rather than to intensify the constraint.*

In biological systems, the processes may often be like those of Sections 3 and 4, the whole having a natural metric and being made of parts rich in equilibria. When this is so, some quantitative relations are likely to appear.

First, Δ can vary on some scale from weak to strong, the weak causing small displacements and the strong large. Obviously, the longer the displacement, the greater is Δ's chance of getting the representative point away to a new confluent and thus of restoring the next response (to D) to its initial value. Thus, the stronger is Δ, the more will the next response return to the initial size.

Secondly, suppose D and Δ have components and a metric of their own, so that the resemblance between them can be measured by the "distance" between them. If Δ is very similar to D it will have similar effects, and will be represented in the phase space by a set of arrows similar to those of D; such an operator will have little tendency to shift the representative point out of a confluent if D cannot. But if Δ is much different from D, Δ's arrow may well take the representative point out of the confluent and destroy the habituation. Thus, the more Δ differs from D, the more effective will it be as a de-habituator.

The theory thus leads, without further modification, to a simple explanation of the phenomenon of de-habituation.

A related property that finds a ready explanation is that the habituated biological system, if left alone, usually recovers its responsiveness "spontaneously". This recovery would occur in systems for two reasons, which are related. First, if being "left alone" means that the system is, in fact, left subject to many little disturbances Δ_1, Δ_2, ... etc., too small to be noticed by the experimenter, then the recovery may be due to the representative point going on a "random walk", under the action of the little disturbances. If it arrives at a new confluent, the response to *D* will be freed from its terminal constraint and will be restored to its initial size. Thus, bombardment by small disturbances will tend to restore responsiveness.

It may also happen that some factors, internal to the system but not noticed by the experimenter, are acting as parameters to the system, are changing slowly in value, and are thus changing its field (*ref. 1*, S.6/3), its *T*-transformation, and the distribution of its confluents. Such, for instance, may happen as an experimental dog grows hungry. When the confluents are changed, a representative point that was previously trapped may be no longer trapped in the new confluent. Thus, secular change in an unobserved parameter may lead to a "spontaneous" recovery of responsiveness.

These facts can be readily illustrated by the "Monte Carlo" method. After the system of *fig. 2* (actually a hundred subsystems) had reached equilibrium under Σ, the operator Δ was applied once:

$\Delta = \Delta_4$:

y_4 =	3	4	5	6	7	8	9
↓ y_4' =	0	0	0	0	9	7	8

(It came after operator *U*, so the state operated on by Δ would be four digits lacking 0, 1 or 2).

As a result of Δ's action, about 4/7 of the habituated subsystems were thrown out of equilibrium under Σ and were forced to follow a trajectory under Σ until stable again. The result is shown on the right of *fig. 2*. The column marked Δ shows the response (under *T*) after it had been disturbed by Δ. The next column to the right shows the response to an application of *D*: previously fallen to zero, it is now restored to the extent of 1.8, and thus demonstrates de-habituation. (Then as Σ is applied again and again, the response falls off to zero).

This Δ displaced only about 4/7 of the hundred subsystems' equilibria. Had it been stronger, e.g. had it been

y_4 =	3	4	5	6	7	8	9
↓ y_4' =	0	0	0	0	0	0	0

it would have disrupted all the hundred equilibria, and would have restored the response to its initial value.

It is not, of course, proposed here that the basic process of the theorem is responsible every time the phenomenon of habituation appears. Showing that Postulates I to V lead to habituation does not exclude the possibility that other processes may also lead to the same phenomenon. Sometimes, especially in sensory adaptation, the phenomenon may be related to survival, either positively or negatively; then natural selection will interfere to give the system what properties are best, departing from those of Postulates I to V. So may develop such specialised, and doubtless more efficient, systems as that of Pringle (1951, *ref. 8*).

On the other hand, when the Postulates hold, habituation follows necessarily. Since the conditions in the cerebral cortex and in protoplasm approximate, from some points of view, to the conditions of the Postulates, it seems likely that much of the unspecialised habituation shown by them is primarily due to this process. With this explanation available, the onus of proof now passes to those who wish to propose that some particular act of habituation is not due to this very generally available process.

The theorem suggests that our attitude to habituation has been basically mistaken. We treated the organism in a particular way, obtained habituation as an outcome, and then asked: what peculiar property in the organism is responsible for this outcome? The theorem suggests that the question itself is wrong, for it assumes that the peculiarity is to be sought in the organism. In fact, the peculiarity, and the reason for habituation, lie at least as much in the particular sequence of operations used by the experimenter. These have sufficient character or pattern to impose some pattern on the response of the system; this pattern shows as habituation, all that is required of the system being that it should be in the class that does not totally destroy such patterns.

REFERENCES

1. ASHBY, W. ROSS. Design for a brain. *London, Chapman & Hall,* (1952).
2. ASHBY, W. ROSS. An introduction to cybernetics. *London, Chapman & Hall,* (1956).
3. BOURBAKI, N. Théorie des ensembles; fascicule de résultats. *A.S.E.I., No. 1141. Paris, Hermann & Cie,* (1951).
4. DANISCH, F. *Zeit. f. allg. Physiol.* 1921, **19**, 133.
5. HARRIS, J. D. *Psychol. Bull.* 1943, **40**, 385.
6. HUMPHREY, G. The nature of learning. *London, Kegan Paul,* (1933).
7. KENDALL, M. G. and BABINGTON SMITH, B. Tables of random sampling numbers. *London, Cambridge University Press,* (1951).
8. PRINGLE, J. W. S. *Behavior,* 1951, **3**, 174.
9. RUBIN, H. and SITGREAVES, R. Tech. Rep. 19 A; Applied Maths. and Stats. Lab.; *Stanford Univ., California,* (1954).
10. SHARPLESS, S. and JASPER, H. *Brain,* 1956, **79**, 655.
11. THORPE, W. H. Learning and instinct in animals. *London, Methuen,* (1956).

DISCUSSION ON THE PAPER BY DR. W. ROSS ASHBY

DR. GREY WALTER: As Dr. Ross Ashby said in his paper, the process rather ineptly called 'habituation' is basic to any consideration of learning. Learning must involve selection, selection implies rejection of more information than is accepted, and this rejection can be called habituation, though this is the reverse of habit-formation. There is a great interest in this field at the present time; at a Colloquium in Moscow some weeks ago this formed the main topic of conversation, though the term used there is "Extinction of the Orienting Reflex".

As you know, there is now ineluctable evidence that habituation is a very active process. This is the question I would like to put to Dr. Ashby. When an animal is subjected to monotonous stimuli there is usually a progressive reduction in the response in the central nervous system. This reduction is due largely to changes at a peripheral level, even in the receptors themselves, yet it seems to depend upon the integrity of rather complex structures in the brain. This should not be confused with Adaptation which is another inherent property of receptors; it is rather a dynamic process of control, of selective blocking from the nervous system outwards. It is susceptible, as Dr. Ashby mentioned, to dishabituation in the presence of another distracting stimulus, and Habituation of this type depends upon integrity of the central nervous system; for example, light anaesthesia may paradoxically augment the brain response to a stimulus which had previously lost its effect by habituation. An anaesthetised animal literally does not perceive the stimulus, yet the electrical response is larger.

This suggests that the structures involved are particularly vulnerable and are therefore probably complex in structure. Does this not imply that habituation in a living organism depends upon a rather elaborate high-level mechanism of selection and control?

Another point is that the degree to which habituation occurs depends on the intensity of the stimulus. If you exhibit a series of auditory stimuli to a human subject, some of which are of moderate intensity and start by being neutral and others are strong enough to evoke an unconditional reflex response, then there may at first be habituation to the moderate neutral stimuli. But if these are associated with the unconditional stimulus so as to become conditioned, the habituation disappears and the response may get larger and larger. The response to the intense unconditional stimulus shows

no habituation however, and in some people may even increase at each repetition.

One can visualise a family of curves plotted on a co-ordinate system in which the ordinate is size or speed of response and the abscissa the number of presentations of a stimulus, the parameter being the intensity of the stimulus. The curves for low intensity stimuli would show a decline with repetition such as Dr. Ashby describes, but those for high intensity stimuli would start off level and then climb away to a limiting value. This sort of relation is seen in the study of nerve fibre excitation where the local potential evoked by subthreshold stimuli declines with time, whereas it rises explosively into an all-or-none propagated impulse in response to supra-liminal stimuli.

Such empirical differences serve in practice to distinguish between neutral stimuli, which show habituation, unconditional ones which show none or even facilitation, and conditioned ones which follow first one rule and then the other. I wonder whether Dr. Ashby has any views about this as a theoretical criterion for identifying the character of stimuli from the point of view of an organism? A classification of this sort would be of great value in experimental studies because it is not at all easy to decide which features in an animal's behaviour are reflex, instinctive, unconditioned, and which are learned, conditioned, exploratory or random.

MR. G. PASK: I have a brief comment to make about this entirely complete paper. It is a request for an extension of the argument to include a more general kind of adaptive behaviour.

Some years ago Eccles *(ref. 1)* suggested that any self oscillatory network of neurone like elements able to interact freely will tend to reject a repeated mode of oscillation in favour of some new mode. If the network also receives stimuli from the external world, and if these sustain a particular oscillatory mode, it is equally the case that this mode will be suppressed, or will be more difficult to sustain, because of the repetition of the stimuli which engender it. The network is thus selective towards novel stimuli and so far as repeated stimuli are concerned it will habituate.

At first sight it seems that all of Dr. Ashby's mathematical arguments are applicable to a structure of this kind and I should be grateful to have a pronouncement on this matter. If such an extension is possible it will provide a calculus for examining not only habituation but the whole process whereby an organism acquires differentiated sensory mechanisms.

REFERENCE

1. ECCLES, J. C. Neurophysiological basis of mind. Waynflete Lecture 1952. *Oxford University Press* (1953).

DR. M. J. BUCKINGHAM: It should be possible to test biological or other systems to see if they behave in a way consistent with that deduced from any particular mechanism for adaptation, such as the interesting one suggested by Dr. Ashby. For example, a system representable by the simple case illustrated in his *figure 1* should show the following property: systems habituated to a stimulus S are subjected to a stimulus of the same type, but with intensity increased just beyond the threshold, until a non-habituated response is elicited. This response should be greater on the average than that shown for the same stimulus by systems which had not been previously habituated. This is a result only for the simple case illustrated, but I would like to ask Dr. Ashby if he has been able to deduce positive predictions of this type, for the more general situation.

DR. W. K. TAYLOR: I believe there is some evidence that the effect of a repeated stimulus can be reduced by an active cancellation process whereby the nervous system learns to cancel the effects of the stimulus. If, for example, the visual scene tends to oscillate, eye movements can compensate for this and the compensation tends to persist after the stimulus is removed. Since many receptors tend to respond to the rate of change of the scene the response will decrease as the compensating mechanism learns to keep the image stationary. This principle also appears to act if we are repeatedly subjected to a force pattern. If, for example, one repeatedly uses the same escalator one gradually builds up an automatic compensation system which reduces the tendency to overbalance on leaving it. Evidence that this is an active cancellation process is afforded by the observation that unbalance tends to occur when one leaves a stationary escalator. It seems to me that there may be a general tendency for living organisms to learn automatically to actively oppose the effects of repeated stimuli.

DR. ROSS ASHBY (in reply): The points that have been raised by Drs. Grey Walter and Taylor are of interest, but hardly concern me, as they refer to mechanisms in which the habituation has been specially developed, usually by natural selection, and I discussed in my paper only a hitherto unnoticed and purely *general* reason for the appearance of habituation. Naturally I do not for a moment suggest that habituation, whenever it appears, is *always* due to this purely general, entropy-like process - I accept freely that sometimes, for reasons of special urgency, natural selection has fostered the development of highly specialised mechanisms that will show habituation with unusual speed and efficacy. The question of what mechanism is at work in any particular case can be answered only by a detailed experimental study of it.

A point I would like to emphasize is that when one uses this type of reasoning about some real "machine" one should realise how free one is to

select *any* meaningful system (as a set of variables) from the infinite number of possibilities that plausibly suggest themselves. A "state", for instance, can legitimately be anything that an observer can recognise with confidence and reliability. It may be one of three typical expressions on an infant's face, or even something like a Fraunhofer line - at which there is actually nothing at all! - no matter, if the observer can recognise it reliably it qualifies as a possible "state". Unless the observer uses freely this freedom of selecting a suitable point of view, he is likely to overlook many possibilities of the theorem's application.

A good example is given by the Eccles' network, referred to by Mr. Pask. We can think of this as changing from state to state so as to *show* the oscillation, or we can ignore these states, think only of the *modes*, and define these as the "states" of a more abstract system, that changes only by going from mode to mode. If the network has plenty of modes at which it can stick, then it will conform to Postulate VII, and will therefore (to answer Mr. Pask's question) certainly tend to show habituation. Thus, though still showing the changes that represent oscillation, it will tend to get to a mode from which the repeated disturbance fails to move it.

Dr. Buckingham's question raises a very interesting extension of my work. Various extensions are being considered at the present time, but I have not yet had time to consider Dr. Buckingham's extension in detail.

It is of interest to notice that these entropy-like processes are all related. Thus, the process of ultrastability that gives coordination and integration is of the same type as this process that gives habituation. In each case the system may go either to a confluent that leaves it immune to disturbance or to one that leaves it still vulnerable. The process is then of the type: Heads, I win; Tails, we toss again - and it tends inevitably to a confluent (or to a set of step-mechanism values) that gives immunity. In general this means habituation; in the particular conditions of the ultrastable system it means coordination and integration.

SESSION 1

PAPER 5

CONDITIONAL PROBABILITY COMPUTING IN A NERVOUS SYSTEM

by

DR. A. M. UTTLEY

BIOGRAPHICAL NOTE

Dr. Uttley took an Honours degree in Mathematics at King's College, London where he also took a degree in Psychology and did post-graduate research in Visual Perception. At the Royal Radar Establishment he designed and built analogue and digital computers. For the last five years Dr. Uttley has been working on theories of computing in the nervous system.

CONDITIONAL PROBABILITY COMPUTING IN A NERVOUS SYSTEM

by

DR. A. M. UTTLEY

ABSTRACT

IN two previous papers it has been suggested that two particular mathematical principles may underlie the organization of nervous systems; the first is that of classification (Uttley, 1954, *ref. 13*) and the second is that of conditional probability (Uttley, 1956, *ref. 14*). The suggestion is based on the similarity of behaviour of these formal systems and of animals. The design of classification computers is discussed in the first paper; the design of conditional probability computers is discussed in a third paper (Uttley, 1958, *ref. 15*); in both papers working models are described. Further reference to these papers will be by date only. It is the aim of the present paper to consider whether the two principles might operate in nervous systems.

There are four requirements for the principle of classification to operate in an area of a nervous system. Firstly, in that area, signalling must be binary; this would be the case if, for example, the impulse frequency were at either a very low rate or at a maximal rate, or if signalling were in terms of standard volleys; in general, if the fibre activity were in one of only two states.

The second requirement is that the fibres which form the input to the area be connected to neurons in as many different ways as possible; there are many areas in which this condition is met.

The third requirement is that more than one synapse of a neuron must become active for it to fire; this appears to be met.

The fourth requirement is that there shall be some way of delaying signals for periods of the order of seconds. A block of isolated cortex does remain active for such periods when stimulated briefly so in this way the requirement might be met.

If these conditions are all met each neuron will indicate, by firing, the occurrence of a particular spatio-temporal pattern of activity in the input to the system.

If the system is to have the additional property of computing conditional probabilities and so of making inferences it must meet four additional requirements three of which have been suggested by Burns (1955, *ref. 4*) to explain the spontaneous firing of a single neuron. The first two demand that some physical quantity in a neuron shall vary (e.g. a chemical concentration). The quantity must change when the neuron fires and recover at a slow rate when the neuron is not being activated.

The third requirement is that the effect of one firing neuron upon another to which it is connected must depend on their relative state. This hypothesis has been made before by Burns, (loc. cit) who has recently found evidence to support it.

The last requirement is met if there are short axon neurons, embedded in the population of indicator neurons already described, whose axonal systems have an inhibitory effect.

From the three papers, two important points emerge. Firstly, a theory designed to explain only conditioning and extinction was found to explain, without extension, a number of other forms of plastic behaviour. Secondly, from the theory, the rules for the design of a model were deduced which can be translated into reasonably biological requirements.

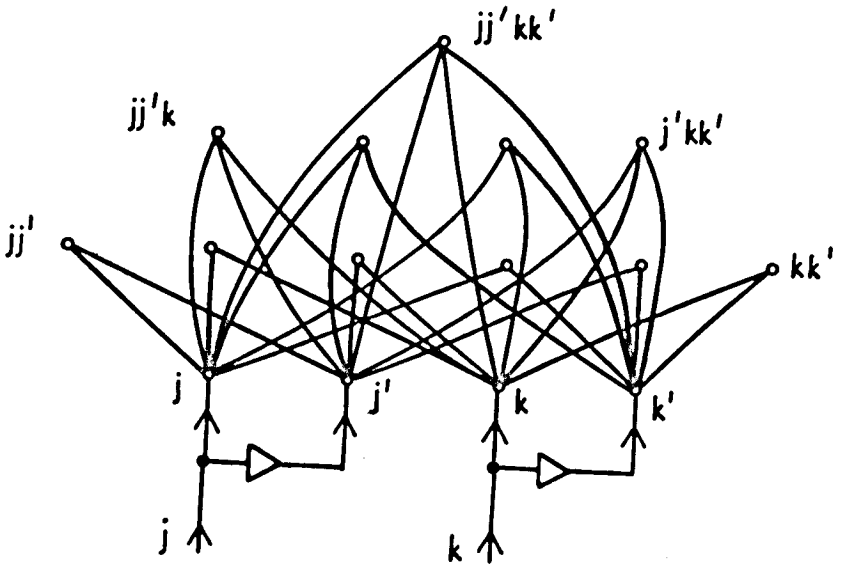
There are at least two lines of work which may lead to tests of the theory. Firstly, a study of the mathematics of random connexions, using anatomical data of cell and fibre density, will lead to laws describing the limited powers of classification of randomly connected neurons. These powers can be compared with known psychological limits to discrimination. Secondly, in terms of the theory, learning laws in experiments on very simple animals may be translatable into laws of reversible chemical reactions; such work may contribute to an understanding of synapse chemistry.

CLASSIFICATION SYSTEMS

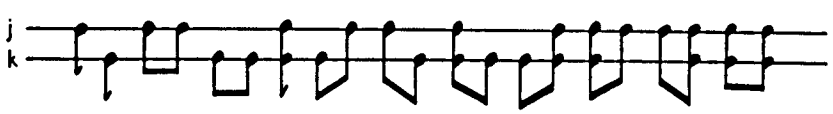
The input to a classification system is a number of separate channels each in a changing state of activity; the system can distinguish complex spatio-temporal patterns of activity in these channels. The output from a classification system consists of a number of channels too and the system can synthesise complex patterns of activity in them.

The conditions for the existence of a classification system have been stated formally (1954, *ref. 13* and 1958, *ref. 15*) and they are repeated here.

1. Each input channel must be always in one of two states active and inactive.
2. The inputs must be combined in as many ways as possible - ideally in all possible ways.



(a)



(b)

Fig.1. The Classification of canalo-temporal patterns of activity in two binary inputs (a). A system of coincidence units and delays, (shown as triangles) which distinguishes only 'before', 'simultaneous' and 'after'. (b) The patterns which can be distinguished by the system.

3. There must be a unit for every combination of inputs, which *indicates* if every input of the combination is active. A combination or set of inputs is said to define a *pattern* of activity. The connexions between inputs and units are called *counting connexions.*
4. If temporal patterns are to be distinguished each input must pass through a series of delays; the output of each delay must provide a separate input to the system of indicating units.

A very simple classification system is shown in *figure 1a*; there are only two inputs *j* and *k* and because each input suffers only one delay the system can distinguish temporally only 'before', 'simultaneous' and 'after'; nevertheless it can distinguish all the patterns of *figure 1b*. Without any extension of principle or of rules of design the number of inputs and of delays can be extended indefinitely so that the system can distinguish very complex patterns; there is always one indicating unit for each pattern.

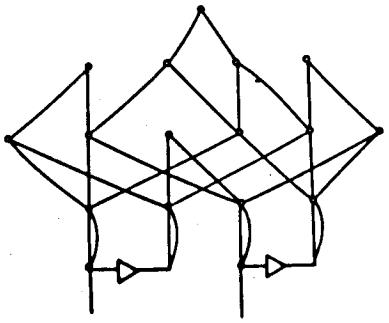
Binary Inputs

Nervous transmission is in terms of standard impulses which meet the requirements of binary classification. However, at low levels in nervous systems, intensity is signalled in terms of impulse frequency. It would not be possible for a train of impulses to signal both an intensity and the number of times that a pattern had occurred. If, at higher levels, patterns are distinguished by classification then intensity must not be signalled in terms of frequency but in terms of 'place'. There are several possible mechanisms by which this coding could occur.

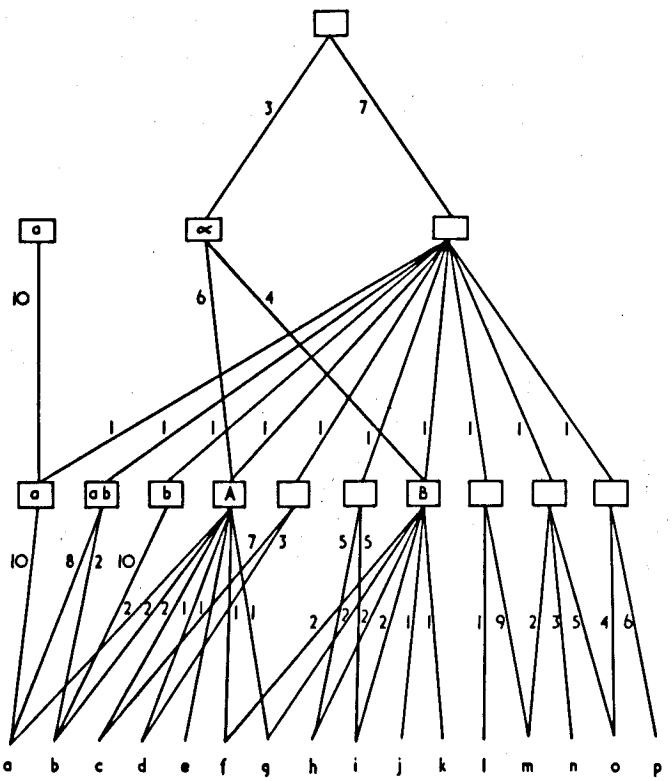
Firstly, by means of a threshold mechanism a system can distinguish signals which are either above or below a certain energy level; all neurons appear to possess such thresholds. The second principle is that of adaptation whereby changes are emphasised; for a signal which consists of pulses of varying frequency if the pulse rate suffers such differentiation many times the signal will be changed into a series of standard volleys - standard because the firing rate of a neuron can lie only between two definite limits. Such a form of activity meets the first requirement.

In a number of areas there is a tendency for firing rates to be standardised; in the auditory nerve, for example, signalling is in terms of pulse frequency (Galambos and Davis 1943, *ref. 8*) but Hilali and Whitfield (1953, *ref. 9*) have shown that, in efferent fibres of the cochlear nucleus and for steady stimulus tones, this frequency is, broadly speaking, either at about 50 pulses per second or it is at the slow background rate. A similar effect appears to occur in ventral horn ganglia so that the force in a muscle is determined more by the number of muscle fibres contracting maximally than by the impulse rate in individual fibres.

SCIENTIFIC LIBRARY
APR 22 1960
U. S. PATENT OFFICE



(a)



(b)

Fig.2. Classification by means of identical units each of which indicates if a fixed number of inputs to it become active, (a) when the number is two, (b) when the number is 10.

110182

The Indicating Units

It has been pointed out (1958, *ref. 15 p. 14*) that a classification system can be constructed from units of a common design if each unit indicates when more than a critical number of the inputs to it are active; the critical number must be two or more. It is easiest to see this if the critical number is taken as two: the system of *figure 1a* then takes the form of *figure 2a*. There are two inputs to each unit; if they are active simultaneously an output of the unit is active and can serve as an input to further units. A classification system using units for which the critical number is ten takes the form of *figure 2b;* it has been discussed in an earlier paper. (Uttley, 1954, *ref. 13*). If the critical number of impulses is different for different neurons classification will still arise but its form will vary in the way shown in *figures 2a and 2b.*

It can be seen that in *figures 2a and 2b* a principle of coincidence is employed so that an *ab* unit, for example, could not fire more often than an *a* or a *b* unit. It follows that the higher the level in such a system the lower the rate at which its units indicate; it would appear inevitable, if more than one synapse must become active for a neuron to fire, that the same statement must be true of a nervous system.

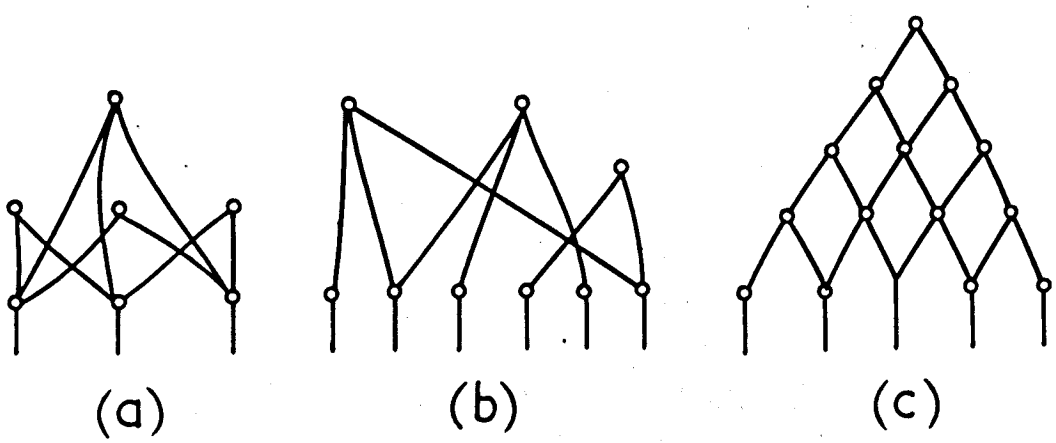


Fig.3. (a) Complete classification of three inputs.
(b) Incomplete classification of specially chosen patterns.
(c) Incomplete classification arising because connexions are limited in length.

The Connexions

A classification system can be complete, as in *figure 3a*, and able to distinguish all possible patterns - this is a mathematical ideal; or it can be incomplete and designed to distinguish special patterns, as in *figure 3b*; or it can be incomplete and yet have no special structure to enable it to distinguish special patterns - such an arrangement is shown in *figure 3c*. It has been suggested (1954, 1956, *refs. 13,14*) that in the second form the system can imitate releaser mechanisms and pointed out (1954, *ref. 13*) that the third form arises if the connexions between input channels and units are entirely random; this possibility will now be considered further.

Consider a set of afferent fibres arborising in a population of identical neurons each possessing a random dendritic system; the axonal systems of such neurons will not be considered at this stage. Suppose that synapses form by chance proximity and that a neuron fires if it receives two simultaneous impulses. Such a system is shown diagrammatically in *figure 4a* and more realistically in *figure 4b*; it possesses three features which do not occur in the formal classification system. Firstly there is more than one neuron with, for example, one connexion to an afferent fibre *a* and another to a fibre *b*; so, corresponding to the *ab* unit of the formal system, there is a diffuse cluster of *ab* neurons, and there is such a cluster for every unit of the system. The different clusters overlap but there is no confusion because each neuron is permanently labelled in terms of its connexions; it does not matter where a neuron is but what it is connected to. Because the units are duplicated such a neural system can sustain much damage before it is unable to distinguish patterns. Secondly, there are neurons with less than two connexions; so they will never fire. In the limit, if fibres *a* and *b* are sufficiently far apart there will be no *ab* neurons, so classification will be incomplete; the actual number of neurons which distinguish the pattern *ab* will therefore depend upon the separation of the afferent fibres. Thirdly, if a neuron has connexions to more than two fibres it will indicate ambiguously, that is, when any two of the connexions are active.

If the neurons possess axonal systems these will make similar connective arrangements with the dendritic systems of further neurons; this is shown in *figures 4a and 4b*. Because these neurons are indirectly connected to the original fibres via one intermediate neuron they will be said to effect *first order indirect classification*. There are two consequences; firstly, there will be more duplication of indicating units; secondly, there will be units which distinguish more complex patterns. The connective arrangements of *figure 2a*. can therefore arise, and the second of the above conditions is met.

The system can be extended to higher orders indefinitely, but it will be stable only under special circumstances. Consider only the *a* neurons of *figure 4*; the number which are indirectly connected to the *a* afferent fibre

through one intermediate neuron might exceed the number of directly connected *a* neurons; similarly the second order neurons might exceed the first order neurons, and so on indefinitely; if the *a* fibre became active there would then be an uncontrollable increase in the number of active neurons. The magnification ratio per order depends on the total length of the axonal system of a neuron (1955, *ref. 16*); if the ratio exceeds unity any input can fire the whole population. In terms of classification each unit is then an *a* unit OR a *c* unit OR ... and there is complete ambiguity of classification.

If, on the other hand, the magnification ratio is kept below unity so that the system is stable, then the actual number of neurons of a given order will decrease with the order. In consequence, as the complexity of a pattern increases the number of neurons which distinguish that pattern will decrease; but the nearer to the instability point that a system can be held, the greater the complexity of the patterns that it can distinguish.

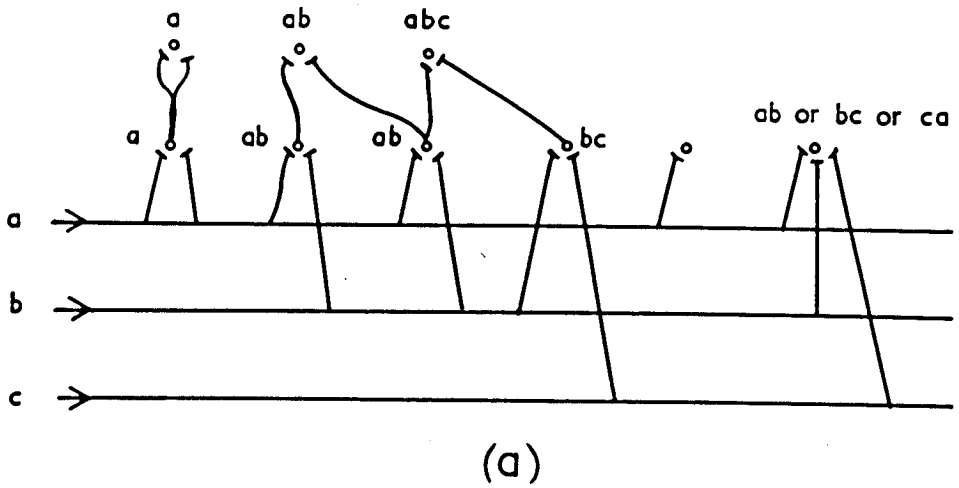
The stability of a population of neurons has been studied by Beurle (1955, *ref. 3*), Allanson (1955, *ref. 1*) and the writer (1955, *ref. 17*); this work must be related to experiments on isolated cortex (e.g. Burns, 1951, *ref. 5*) and to measurements of fibre density (e.g. Sholl, 1953, *ref. 12* and Eayrs, 1955, *ref. 6*).

The Number of Units

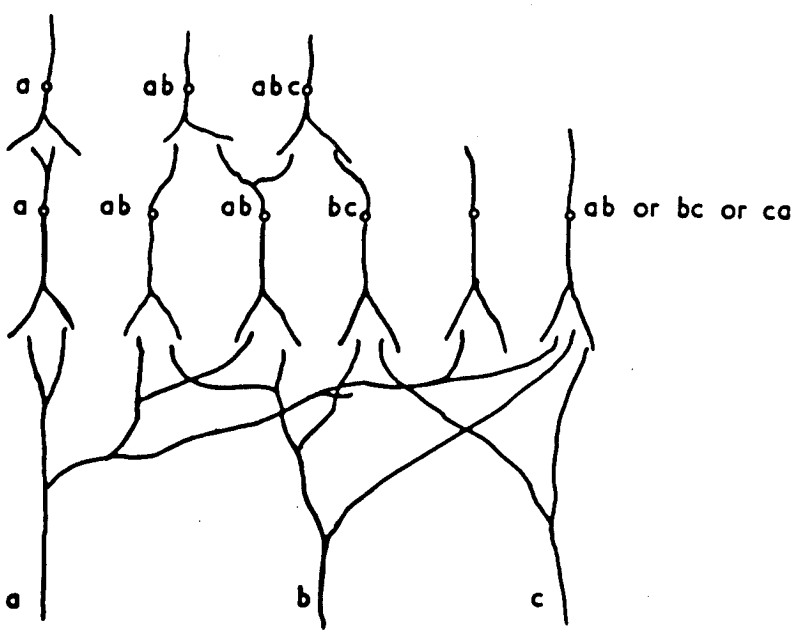
It has been suggested above that each neuron distinguishes a particular pattern of activity by virtue of its connexions. The hypothesis will be made that these connexions do not change, once having formed; although their function may. Consequently it is most important to consider whether the number of neurons in a nervous system are adequate to account for all the patterns of activity that an animal can distinguish.

The first point to be made is that one is here considering the patterns of activity which are initiated within the system - the tunes which are being played upon the receptors - not the limitless patterns of the external environment which can arouse the internal patterns. If there were n fibres which signalled in a binary way to a neural classification system then, for complete classification, that system would have to contain 2^n neurons. From the point of view of this paper the evolution of nervous systems is seen a struggle to solve that intractable problem; the failure to solve it can be seen in a number of known psychological limits to pattern discrimination. Incidently, if each input fibre possessed not two but m discriminable levels of activity, the required number of units would be, not 2^n but m^n; this is the fundamental reason why the input fibres should use a binary system.

At least three methods appear to be used in nervous systems in order to reduce the number of discriminating units. Firstly, the number of fibres entering the brain is reduced to much less than the number of

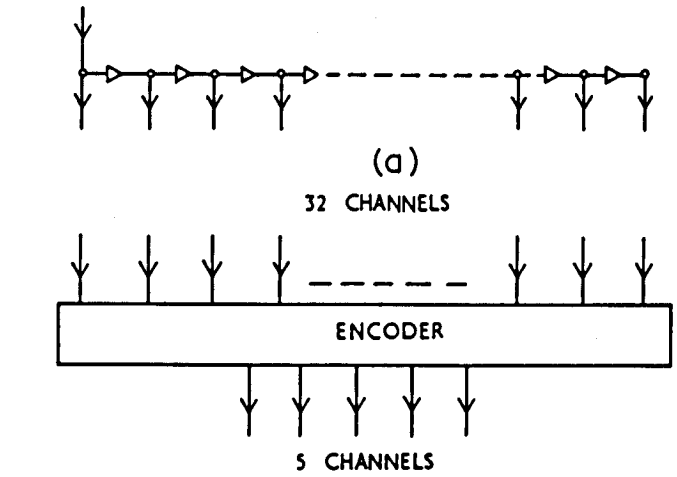


(a)



(b)

Fig.4. Imperfections in classification due to random connexions; some units never indicate, some do so ambiguously. (a) Formal connexions to units which indicate if two inputs are active (b) Neural realization of these connexions.



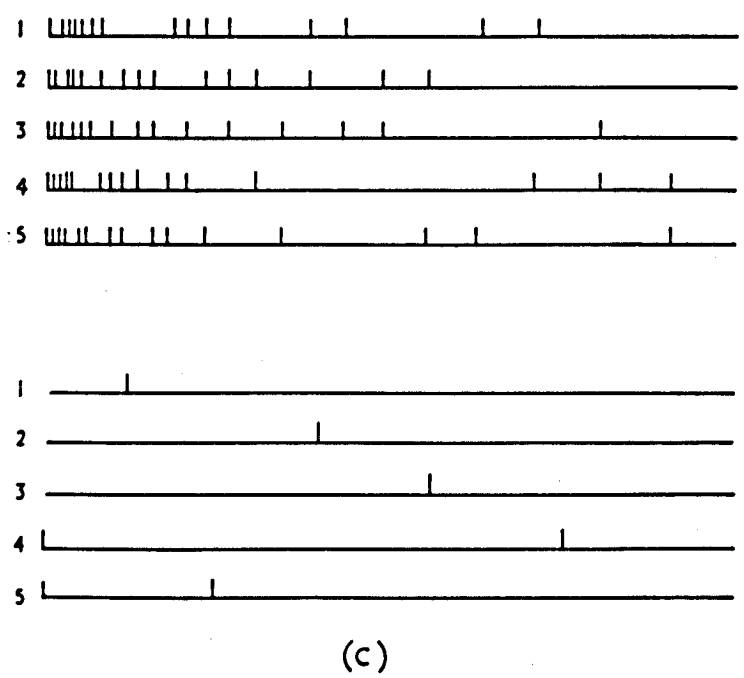


Fig.5. Discrimination of temporal intervals by means of (a) a series of delays followed by (b) an encoder. The output of the combined system, when a single pulse enters, is shown at (c).

receptors. It is suggested that this is one of the primary functions of the neural bodies which are interposed between receptors and brain; in other words they may be said to abstract key features and reduce redundancy. For example, even at the early stage of the retina, there is a hundredfold convergence from receptors to optic nerve. There is a corresponding loss of discrimination in that the information passed on is concerned primarily with contours. In consequence, for man, n is of the order of 10^6, so that the information reaching the brain from all senses is about as much as is contained in black-white television screen.

Secondly, the nervous system does not achieve complete classification of the inputs. If it were so there would be a unit for the Eroica and if it fired one would perceive the whole symphony in that magic instant. But, through the phenomenon of short term memory, one can perceive only the last few seconds' worth. So if r is the number of fibres which fire during the span of short term memory N is of the order of nCr.

Thirdly, there is the phenomenon of span of discrimination whereby a set of unrelated objects can be distinguished, at first glance, only if the number in the set is less than about seven.

Opposing these factors is that of duplication. It has been pointed out earlier that, by random connexions, duplication is likely to occur and that this will enable the system to function even if damaged; but such a gain will be balanced by a loss of discrimination if the total number of neurons is not increased.

To conclude, at the present time there appears to be no evidence to show that the nervous system of any particular animal does not contain enough neurons to match its powers of discriminating patterns of receptor activity.

Delays

It has been shown (1954, *ref. 13*) that for temporal discrimination each binary signal must pass through a series of delays, as in *figure 5a*, and that the output of each delay must provide an input to a classification system.

The human can distinguish temporal intervals between brief stimuli in accordance with Weber's law for intervals up to about three seconds; it will be assumed, for the convenience of binary arithmetic, that one can distinguish 32 such intervals. A complete classification system to distinguish temporal patterns to this extent would require, not 2^n but 2^{32n} units. However, if at any instant only one of the 32 outputs of *figure 5a* is active it is possible to design a system which recodes the information from the 32 channels into 5 channels, each active or inactive: the output from such a coder, shown in *figure 5b*, has 32 possible states. The coding system could be that of binary arithmetic which is shown in column two of the table below.

Number of Delays	Normal Code	Rearranged Code	Total Lapsed Time assuming Weber's Law
0	00000	11111	a
1	00001	01111	ar
2	00010	11101	ar^2
3	00011	10111	ar^3
4	00100	11011	ar^4
5	00101	11110	ar^5
30	11110	00100	ar^{30}
31	11111	00000	ar^{31}

Any fixed arrangement of this code would suffice; let it be according to the number of units in the code number. Such a code is shown in column three; the undelayed signal is represented by five units; the next five discriminable intervals are represented by numbers containing four units and one zero and placed in random order; the last interval produces zero activity in the output of the coder. The fourth column shows the time at which each number should appear at the output of the coder for Weber's law to hold. The data of columns three and four are shown graphically in *figure 5c*; this is the output of the combined delay and coding system when a single pulse occurs at the input. It can be seen that the system behaves exactly like the nearly unstable population of neurons which has been postulated already. Consider all the *a* neurons suggested by *figure 4e*; when the *a* afferent becomes active the directly connected *a* neurons will fire after one synaptic delay; this activity will be passed to *a* neurons of higher and higher order, their number steadily decreasing, until the reverberation finally ceases. It is suggested therefore that a nearly unstable population of neurons, each of which fires if a critical number of impulses falls on it, not only classifies but also incorporates the necessary system of delays. However this system is not exactly of the form originally proposed and in which delays occur only at the input level. Not only will the clusters of *a* neurons behave as in *figure 5c*; to every pattern of input activity there will correspond a cluster of neurons which will behave as a reverberating delay system.

The original formal arrangement of delays is shown in *figure 6a* and the suggested neural realisation of it is shown diagrammatically in *figure 6b*. In the latter there is a redundance of delays but it will have a certain advantage; the number of neurons which indicate the pattern *'A one second before B'* will be independent of the complexity of the patterns *A* and *B*; in other words temporal discrimination of the stimuli will be independent of their complexity.

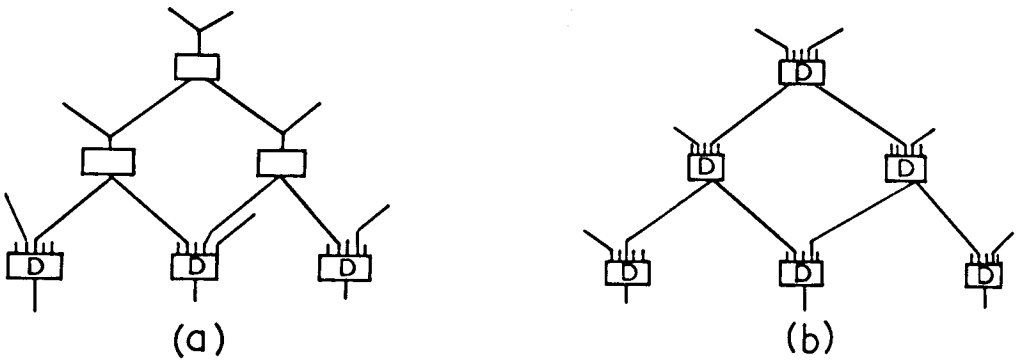


Fig.6. Delays in a classification system may be (a) all at the input or (b) distributed throughout the system.

In making an estimate of the number of neurons required to distinguish spatio-temporal patterns one concludes that it is not necessary to allow one factor for delays and another for duplication; the duplication provides the delays.

CONDITIONAL CERTAINTY SYSTEMS

A conditional probability system is a classification system with an additional function; it has been defined (1956, *ref. 14*) as one in which, when any set of inputs becomes active, the conditional probability is computed of every other possible set of inputs; the conditional probability of each set is contained in the unit which distinguishes that set. The design of such a system has been discussed in a previous paper (1958, *ref. 15*) from which there were the following conclusions.

The system is much easier to construct if each unit indicates only whether the conditional probability of the corresponding set of inputs exceeds some arbitrary threshold. For example, suppose that this threshold is taken as one half and that, in the past, patterns A and B have occurred as follows:-

A 1111111
B 1111111111

If A occurs $p(B/A)$ is 4/7 so B is indicated (inferred); on the other hand if B occurs $p(A/B)$ is 4/10 so A is not inferred. Each unit therefore

indicates in two different ways; if the corresponding set of inputs becomes active it *indicates occurrence;* if this set is only inferred it *indicates conditionally.* Such a system is called a *conditional certainty system.*

The first new requirement of the system *is* that each unit shall count, on a common scale, the number of times that it has indicated. The design of the whole system is eased if counting is on an approximately logarithmic scale, and this can be achieved in the following way.

1. When any unit is not indicating occurrence, some physical quantity associated with it grows - preferably in an approximately exponential manner.
2. When a unit indicates occurrence this physical quantity is decreased, preferably by a fairly constant fraction of its present value.
3. To imitate spontaneous recovery in conditioning there must be storage in depth.

The stored quantity is called the rarity of the set of inputs; it is approximately log K/N, where K is a constant and N is the number of occurrences.

If a set of inputs B includes a set A, then the unit which distinguishes set B is called a superunit of the A unit; conversely the A unit is called a *subunit* of the B unit.

There are two new systems of connexions in a conditional certainty system in addition to the counting connexions. There is a connexion from each unit to all its superunits; its function, which is called *supercontrol* is as follows:-

4. A unit which is indicating occurrence causes any superunit of it to indicate conditionally if the rarity stored in the latter does not exceed that stored in the former by more than some critical amount.

There is also a connexion from each unit to all its subunits; its function, which is called *subcontrol* is as follows:-

5. A unit which has been supercontrolled to indicate conditionally causes all of its subunits to indicate conditionally. If the system has more than two inputs there is one further rule.
6. A unit which is counting inhibits supercontrol by its subunits.

The possibility that these six requirements may be met in nervous tissue will now be considered.

Counting in a Neuron

In a classification system the units must possess two states, active and inactive, in order to indicate occurrence. It has been suggested that in the nervous system there correspond neurons which are either firing or not firing and which possess a rapid recovery process. The additional requirements of counting (1 and 2 above) are met by a reversible process in the neuron whereby its state is changed when it fires. An example of such a process is the transfer of sodium ions across neural membrane; here the

recovery process (the sodium pump mechanism) seems too rapid but it has been measured so far only in fibres of large diameter which are not associated with learning; for fibres of smaller diameter the recovery may be slower. However, because of the long retention of learned inferences, the slow recovery process is more likely to be a chemical than a physical one.

The fourth and fifth requirements imply that a unit can indicate in quite a different way, conditionally. In doing so it must not count; but, by sub-control, it must affect other units to which it is connected. For the present the indication of occurrence and conditional indication will be both identified with the firing of a neuron; the problem of distinguishing them and the consequence of not doing so will be discussed later.

Connexions for Counting Control, Supercontrol and Subcontrol

For a classification system with three inputs the counting connexions are shown in *figure 7a.* For supercontrol (requirement 4) there are connexions from each unit to all its superunits and for subcontrol (requirement 5) there are connexions from each unit to all its subunits; the two systems of connexions are identical and are shown in *figure 7b,* but the two forms of control are in opposite directions and have different functions.

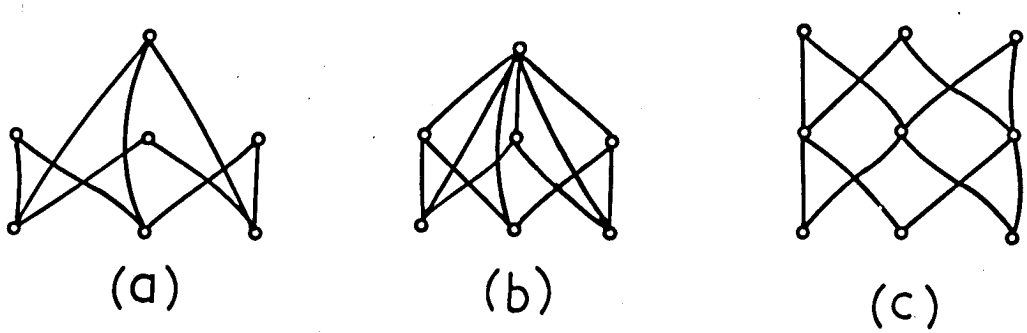


Fig. 7. Connexions for (a) counting control, (b) both super-control and subcontrol, and (c) for all three forms of control.

The connexions for supercontrol and subcontrol include some which are required for counting control and it is an important question whether two or even all three forms of control could be effected by the same physical connexions; this has not been achieved in the electrical computer described in the previous paper and which uses three separate connective systems.

However, at the beginning of this paper, it has been shown that if a unit indicates when more than a fixed number of impulses reach it the

counting connexions take a different form. If the critical number is two and there are only three inputs the system of counting connexions is as in *figure 7c*. It is interesting that this system of counting via subunits also meets exactly the requirements of supercontrol and subcontrol. Not only are no additional connexions needed but incorrect connexions cannot arise; for example, there can be no connexion between a *c* unit and an *ab* unit. This is true by definition; if a unit has been labelled a *ab* unit, for example, this can be only because either it possesses one connexion to an *a* unit and one to a *b* unit, or it possesses two connexions to another *ab* unit. These conclusions are also true if the critical number of impulses is more than two, for example, ten as in *figure 2b*; but then there will be further features of design which have been discussed elsewhere (1954, *ref. 13*). In the rest of this section the critical number of impulses will be taken as two.

The Function of Connexions

An attempt will now be made to discover the conditions under which a neuron must fire in order to meet all the requirements of a conditional certainty system while at the same time using a single connective system for all three forms of control; only the three input system of *figure 7c* will be considered.

For counting control there is the following rule.

A. Whatever the state of a neuron it fires when it receives two or more simultaneous impulses from other neurons whatever their state

For supercontrol a firing neuron must be able to fire a neuron to which it has one connexion but only if the two are in approximately the same state. This difference in state could be effective physically only in an area affected by both neurons but unaffected by others; the synapse meets these conditions.

For a neuron which distinguishes the pattern *a*, suppose that *R(a)*, the rarity of *a*, is stored on the *a* side of all its synapses; similarly for any other neuron *b*. If there is a synapse between neurons *a* and *b*, the independently variable quantities *R(a)* and *R(b)* will be stored on either side of it. A possible mechanism which would meet the fourth requirement is this; suppose that the spike and threshold potentials of a neuron depend on its postulated variable state so that they both decrease after the neuron has fired, and then slowly recover. If neuron *a* fires it is then less able to activate neuron *b*; and if neuron *b* fires, neuron *a* is more able to activate neuron *b*. For supercontrol there is therefore the following rule.

B. A neuron fires when it receives one impulse from a firing neuron if the rarity of the former exceeds that of the latter by less than a threshold amount.

A very simple neural model of a conditional probability system is sketched in *figure 8a.* If neurons *a* and *b* fire then the *ab* neuron will also fire, by rule A; this is counting control. If the *a* neuron fires alone, the *ab* neuron will fire only if *R(ab)* exceeds *R(a)* by less than the critical amount; this is supercontrol. Now the conjunction of *a* and *b* cannot occur more often than that of either *a* or *b* alone so *R(ab)*, the state of the *ab* neuron, cannot be less than *R(a)* or *R(b)*; if the *ab* neuron now does fire conditionally then it must fire the *b* neuron by rule B because *R(b)* is less than *R(ab)*; in consequence, the fifth requirement, of subcontrol, has also been met but with one further physiological hypothesis - that *synapses function in both directions.* However, this hypothesis can be avoided if recurrent collateral axons are introduced in neuron *ab* to mediate subcontrol; in this arrangement, shown in *figure 8b*, there is one system of connexions for counting and supercontrol and another for subcontrol.

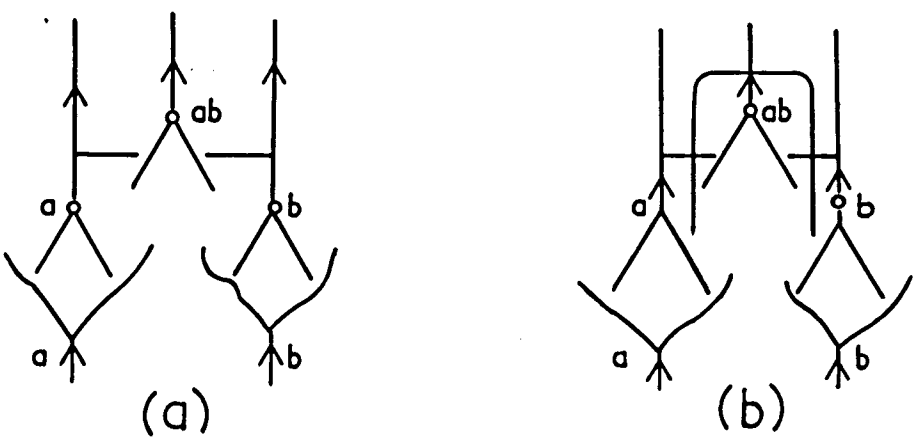


Fig.8. Neural connective system in which (a) each synapse exerts all three forms of control, (b) subcontrol is effected by recurrent collateral axons.

Rule B may be compared with the hypothesis of varying synaptic resistance; according to this theory synaptic resistance is not the same as electrical resistance but it involves the notion of impulses sometimes passing and sometimes not passing over a synapse and of this passage being facilitated by a previous passage of impulses. According to the conditional probability hypothesis the passage of impulses is determined by the difference between two independent variable quantities, the states of the two neurons; but the hypothesis of synaptic resistance refers to only one

variable. In consequence, experimental extinction is more easily explained by the former theory as an active process (1956 *ref. 14* p. 5).

Rules A and B and the connective systems discussed do not seem biologically impossible; indeed they have been suggested by Burns (1955, *ref. 4*) together with the requirements 1, 2 and 3, for an entirely different reason. As will be seen later, the sixth requirement is not unreasonable. The major difficulty must now be discussed; this is that a neuron can give a binary indication in two quite different ways in only one of which its state is changed. In *figure 8b* suppose that some event external to the system causes the *a* afferent fibre to become active; this fibre, by means of two connexions, will cause the *a* neuron to fire by rule A. Suppose that the *a* neuron fires the *ab* neuron by rule B so that pattern *ab* is inferred from pattern *a*. Now supercontrol of further neurons should, by requirement four, be exerted only by the *a* neuron and not by the *ab* neuron. There is no way of making this distinction in the arrangement shown.

This difficulty occurs in an acute form when considering connexions from the system to motor units. Consider a classical alimentary conditioned reflex in which there are the three patterns *Sc*, a conditioned stimulus, and *Su*, an unconditioned stimulus, which always produces a response *R*. Suppose that the three patterns are reported to a conditional certainty system as in *figure 9*; the primitive deterministic control path is shown in full lines; broken lines show the channels by which the activity of the lower level system and of an unrelated receptor system are reported to the higher level system. Whether this system is designed according to *figure 8a or 8b* the input points will also be output points and the broken lines must indicate to two-way signalling. Now suppose that conditioning occurs as in the first sequence of the following table, where a unit in brackets indicates an inferred pattern.

S_c	1111	1 1 1 1
S_u	1111	(1) (1) (1) (1)
R	1111	(1) (1) (1) (1)

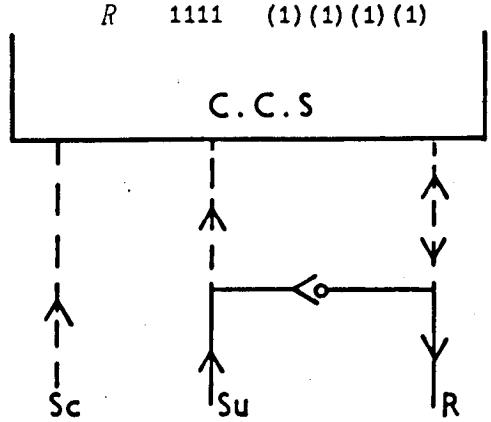


Fig.9. Connexions between receptors and effectors and a Conditional Certainty System which will lead to Classical Conditioning Behaviour.

If S_c occurs alone, as in the second sequence, then S_u and R will be inferred and impulses will pass down the R channel from the conditional probability system to the effector organ in accordance with the facts. But how is the R impulse from the R unit of the conditional probability system to the effector organ to be treated as inferred and not actually counted? If the inferred events, which are shown in brackets in the above table, are counted the probability of R given S_c will remain at unity even though S_c is not reinforced; and this situation will be perpetuated.

The same difficulty occurs in the goal-seeking situation of an instrumental conditioned reflex; this has been discussed in a previous paper (1956, *ref. 14*). In the simplest example there are two situations which arouse the stimulus patterns S_1 and S_2; the animal makes responses which give rise to the proprioceptor pattern P_1 and P_2 in random order. Food which arouses a stimulus pattern G is given for the conjunctions S_1P_1 and S_2P_2 but not for S_1P_2 or S_2P_1. These facts are shown in the following table.

S_1	1	1
S_2		1 1
G	1	1
P_1	1	1
P_2		1 1

Then S_1 will not arouse P_1 rather than P_2 because $p(P_1/S_1)$ and $p(P_2/S_1)$ are both equal to one half. If, however, the pattern G is aroused internally in the system, to represent goal choosing, and S_1 occurs, the P_1 and P_2 units will contain $p(P_1/S_1G)$ and $p(P_2/S_1G)$ which are 1 and 0 respectively; so the pattern will be evoked which has the highest probability. The goal G may now actually occur. How then is the single unit G to be used to indicate at one moment that the pattern G is inferred, imagined, chosen, and at another moment that it is actually occurring? A further important point is that P_1 and P_2 are proprioceptive patterns signals reporting what has happened to joints, tendons, muscles spindles and so on; they are quite different from the efferent patterns of signals sent to muscles. A conditional certainty system which received signals from exteroceptors and proprioceptors but with no input signals reporting the motor patterns emitted could learn what to do but now how to do it.

The problem of connexions from a conditional probability system to effector units cannot be discussed until the problem of the two forms of indication in a unit have been clarified; for this reason the theory developed in the three present papers relates only to the analysis of patterns of afferent signals. The synthesis of efferent signals is the subject of further work.

If the two forms of indication were not distinguished in the neurons of a conditional certainty system there would be important consequences which have been discussed in the previous paper (1958, *ref. 15*). There would, of course, be a failure to distinguish actual and inferred events; but this would occur only in the conditional certainty system; if in other parts of a nervous system neurons were fired only by actual occurrence, then the system as a whole could distinguish actual from inferred events. A further consequence, within the conditional certainty system, would be that a single actual event could set off chains of inferences; if such chains formed closed cycles there would be the possibility of regenerative storage. This would occur if, for example, the past had been as follows:-

A	1	1	1	1	1	0	0	0	0	0
B	0	0	1	1	1	1	1	0	0	0
C	0	0	0	0	1	1	1	1	1	0
D	1	0	0	0	0	0	1	1	1	1
E	1	1	1	0	0	0	0	0	1	1

Then, if *A* occurs, *A* implies *B* implies *C* implies *D* implies *E* implies *A* implies.... An isolated brain does exhibit such continuous activity.

Inhibitory Connexions

The sixth requirement arises in the following way. If a pair of inputs *a* and *c* becomes active then the conditional probability of any input *b* is obtained by subtracting the rarity stored in the *ac* unit, which is counting, from that of its superunit *abc*. But the *(a)* unit is counting too and it must not exert supercontrol, hence the rule. The system of units is shown diagrammatically in *figure 10a*. If *a* and *c* are occurring there will be false supercontrol of the *ab* unit by the *a* unit; this will be prevented if the *ac* unit, in counting, can exert an inhibitory effect on the *ab* unit; similarly for the *bc* unit. In neural terms, the sixth requirement is met, therefore, by the following rule.

C. *A counting neuron inhibits supercontrol by other neurons at the same level.*

The necessary inhibitory links have been indicated in *figure 10a* and shown more realistically as short axon neurons in *figure 10b*. Such neurons occur in large numbers in cerebral cortex and the hypothesis is here made that their axonal systems form inhibitory synapses. This hypothesis has been made to explain inhibitory phenomena in a number of areas, in the cochlear nucleus (Allanson and Whitfield, 1955, *ref. 2*) in the retina and in spinal ganglia (Renshaw, 1946 *ref. 10*).

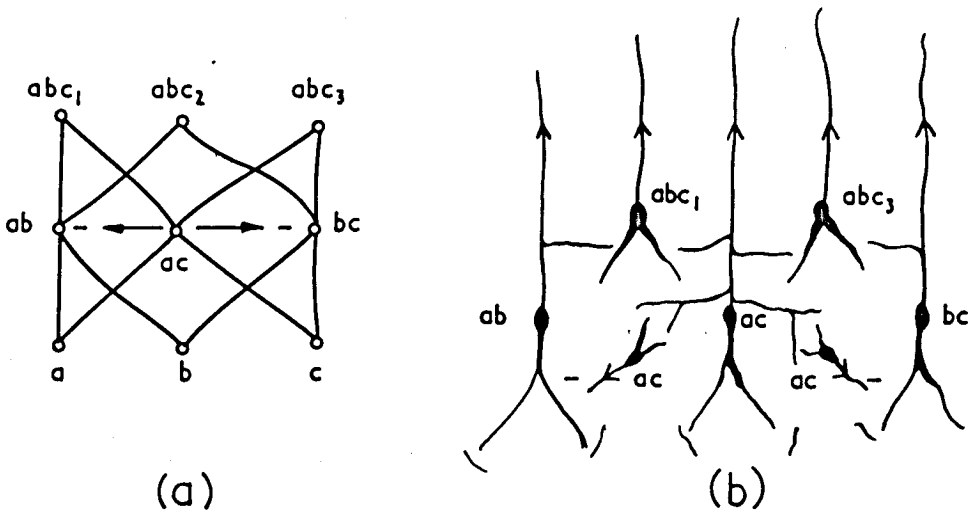


Fig.10. Inhibiting connexions needed in a Conditional Probability Computer.

Storage in Depth

The third requirement has been left until last because it is not essential to the main argument; it is concerned with an additional property of counters which has been discussed in a previous paper (1958, *ref.15)*. In the counting unit of *figure 11a* the first two requirements of this section are met and past events are given an approximately exponential weighting; in consequence the weighting of two temporally separated events is scaled down in exactly the same way as they sink into the past. But in the spontaneous recovery of conditioned reflexes recent events are given enhanced weight in comparison with earlier ones although this effect disappears in time; this property is possessed by the counter of *figure 11b* which gives even greater weight to recent events than does an exponential weighing function. Spontaneous recovery has been demonstrated in a conditional certainty system which uses such counters. The circuit contains two time constants and it is necessary that C_3R_3 exceed C_2R_2; to count an event the switch S is moved to the right and then returned; the voltage R drops immediately and then recovers with the time constant C_2R_2 to the voltage on the condenser C_3. This condenser is hardly affected by a single count - it stores the average rarity over the long period C_3R_3. In a conditional certainty system which uses such a unit inferences (conditional probabilities) can be modified rapidly but there will be a reversion to those inferences which are based on long maintained consistencies in external events.

Spontaneous recovery can also be demonstrated with the circuit of *figure 11c*. A secondary battery B is charged to count and discharged slowly to recover; short time storage occurs in the surface of the electrodes,

long term storage occurs in deeper layers. It is suggested that this storage in depth might also occur in synaptic material lying between the membranes of two neurons; the two surfaces so formed could store the rarities of the patterns distinguished by the corresponding neurons, each backed up by storage in depth.

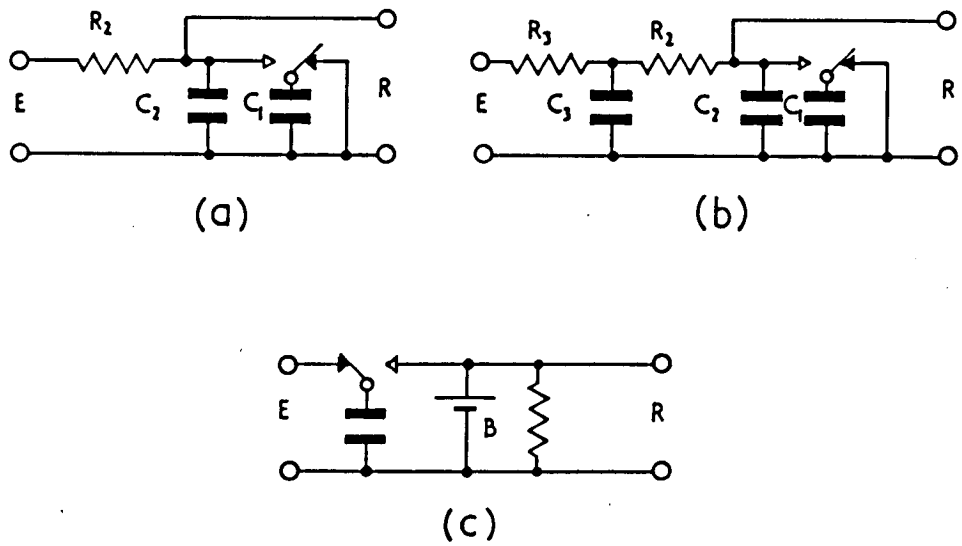


Fig.11. Counters whose recovery processes possess
(a) One time constant
(b) Two time constants
(c) A distributed time constant
The last two imitate the spontaneous recovery of conditioned reflexes.

VARIATIONS OF DESIGN

COUNTING

If there were neurons which met the second requirement of counting but not the first - recovery process, then the system would be capable of learning to make inferences but not of modifying or forgetting them; such a property would be akin to imprinting.

Regenerative storage systems used in digital computers, exhibit no slow recovery processes; however information stored by such methods can be destroyed suddenly and completely if the regenerative loop is broken. If regenerative storage methods were employed in a nervous system, for example in the form of reverberating chains of neurons, then inferences would be made slowly and then either remain fully active, as in imprinting, or be suddenly destroyed.

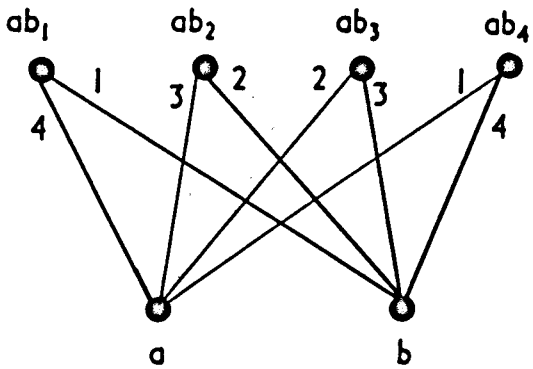


Fig.12. Conditional probability computed in terms of the number of active indicating units.

Conditional Probability Systems

It has been stated at the beginning of this paper that a conditional probability system is more difficult to design than a conditional certainty system; no such computer has yet been built. Yet a salivary reflex, for example, does not suddenly appear; the quantity of saliva secreted increases continuously as the conditional probability of food increases. This suggests that the nervous system does succeed in computing the probability of inferences in a continuous rather than a binary way; two possible mechanisms will be suggested.

First suppose that the critical number of impulses to fire a neuron is more than two, say five; then any neuron, the sum of whose connexions to an *a* and a *b* neuron is five, will be an *ab* neuron; this is shown in *figure 12* in which the different numbers of connexions are indicated. Rule A, suitably modified, implies that, if there have been no conjunctions of *a* and *b*, none of the *ab* neurons will fire if the *a* neuron fires alone; for example, the ab_1 neuron will not fire because it receives only four impulses. But after a sufficient number of conjunctions, *R(ab)* exceeds *R(a)* by less than the threshold value so, by Rule B, the *a* neuron fires the ab_4 neuron to which it has one connexion; even more easily it fires the

other three *ab* neurons since it possesses more than one connexion to each of them. This suggests that at intermediate stages of relative rarity the *a* neuron might fire neurons to which it has two, three or four connexions. If this can occur, the number of impulses (the excitation) arriving at *B* by subcontrol will vary according to the number of *ab* neurons fired by *a*, and hence according to the value of *R(ab) − R(a)*: this is shown in the following table.

R(ab) − R(a) on an arbitrary scale	*ab* neurons fired by *a* neuron	Number of impulses arriving at *b* neuron
5	none	none
4	ab_1 only	1
3	ab_1, ab_2	3
2	ab_1, ab_2, ab_3	6
1	ab_1, ab_2, ab_3 ab_4	10

With the above mechanism the system succeeds, by digital methods, in transferring to the *b* neuron, a *measure* of conditional probability and it solves a problem found difficult by the computer designer. Rules A and B, modified as below, are the criteria for this principle to occur in the nervous system.

(A') *Whatever the state of a neuron, it fires if it receives a critical excitation from firing neurons whatever their state.*

(B') *The excitation contributed by a firing neuron increases as the rarities of the firing and excited neurons approach one another.*

A second possibility has been suggested by Russell (1955). In an electrical conditional probability computer the practical difficulty is that of transferring the difference between two voltages to many parts of the system. It is much easier to transfer a delay. A conditional probability computer has been constructed in which probabilities are computed as delays; the rule for supercontrol (Rule 4 of this section) is as follows:-

4a. *A unit which is indicating occurrence causes any superunit of it to indicate conditionally with a delay proportional to the difference in rarity of the two units.*

For the neural model rule B is modified similarly. Such a delay can be transferred easily from one part of the system to another; it will be approximately proportional to the negative logarithm of conditional probability so that a delay *t* will refer to a probability *exp(−t)*. It is suggestive of this idea that in an alimentary conditioned reflex the delay in response decreases as the probability of food increases. The design of computers using this principle has been described by Russell (*ref.11*).

Equivalence

An instrumental conditioned reflex is extinguished if the correct response is not rewarded. It has been shown (1956, *ref. 14*) that this property arises only if inference is redefined symmetrically as follows:-

A implies B if p(A/B) and p(B/A) exceed a threshold value

It has been shown (1958, *ref. 15*) that for a system to infer in this way the fifth requirement, of subcontrol, must be re-stated as follows:-

A unit which has been supercontrolled to indicate conditionally causes any subunit of it to indicate conditionally only if the rarity stored in the former does not exceed that stored in the latter by more than some critical amount.

In consequence rule B takes the following form.

B A neuron fires when it receives one impulse from a firing neuron if the rarities of the two neurons differ by less than a threshold amount.

CRITICISMS OF THE THEORY

A number of important criticisms have been made of the theory.

Barlow (private communication) has referred to the property demanded by the theory that the threshold pontential of a neuron should decrease after firing (p.16 *l.*22) and so have an increased sensitivity. He has pointed out that this is the opposite of adaptation. Burns (private communication) has shown that it is, in fact, adaptation which occurs in cortical neurons in the isolated brain of cat.

If this reversal in the property of the unit is made the system computes, not probabilities, but rarities; it produces increased outputs as probabilities decrease and should be called a *Conditional Rarity System.* (C.R.S). In trial-and-error learning the system must therefore have an inhibitory function. The all-important Rule B' now becomes

The excitation contributed by firing neuron decreases as the rarities (states) of the firing and excited neurons approach one another.

This rule is met in physiological terms by two neural properties.

1. After a neuron has fired its sensitivity to incoming excitation is reduced. This is adaptation.
2. After a neuron has fired its potential excitation of connected neurons is increased. This is in agreement with the findings of Eccles and others. This is postactivation potentiation (*Eccles, ref. 7*).

Such a change of design is in line with Barlow's demand that there should be a decreased output from units which distinguish patterns of high probability and, conversely, increased outputs for unexpected patterns. In its changed form the system reduces the redundancy in the output channel compared with that of the input channel.

The combination of a C.R.S. with over-connectivity between units appears to resolve the dilemma of the number of units required.

Consider the extreme right-hand unit of *figure 4a* which possesses more input connexions than the two which are necessary for classification.

Suppose that, in the past *a*, *b* and *c* occurred as follows:-

a	111
b	1111
c	1

Then the *a*, *b* and *c* units (which are not shown) will contain 3, 4 and 1 while the ambiguous unit will contain 4. Now, being a C.R.S., the output from the ambiguous unit will be zero when *b* occurs alone. This is reasonable since *b* provides no information as to whether *ab* or *bc* are occurring. The link from *b* to the unit is therefore broken and ambiguity disappears. If now *ab* occurs the unit signals -log(3/4); if the rarer event *bc* occurs it signals the larger quantity -log(1/4).

One is led to the concept of an over-connected system with a reasonable number of units and in which confusion occurs initially. Learning consists in removing the connexions which carry low information; this resolves ambiguities so that the units eventually distinguish uniquely those events which are rare and carry high information.

Watson (private communication) has criticised the demand for symmetrical inference in trial-and-error learning. He considers that there is experimental evidence to show that for a goal *G* an animal chooses a reaction *R* for which the inverse probability p(*G*/*R*) is a maximum regardless of the value of *p*(*R*/*G*). This, in fact, effects a simplification in design.

The neural properties demanded by the system in its changed form will now be summarized. Referring to *figure 8a* and *b*, conditions 1, 2 and 3, (that rarities shall be stored in the units) remain unchanged. Because of the change in Rule B the excitation of the *ab* unit by the active *a* unit is to be proportional to the difference in their states. The output from the *ab* unit and, in general, from all affected superunits thus form a channel of reduced redundancy. Also, the excitation of the *b* unit by the *ab* unit depends only on the difference in their rarity. In consequence, the *b* unit and, in general, all affected subunits form a channel which signals the inverse rarities demanded by Watson.

ACKNOWLEDGEMENTS

Thanks are due to J. A. V. Bates, H. B. Barlow, G. Russell, W. H. A. Rushton and D. A. Sholl for many discussions and suggestions.

This paper is published by permission of the Director of the National Physical Laboratory, Teddington, Middlesex.

REFERENCES

1. ALLANSON, J. T. Some properties of a randomly connected network. Third London Symposium on Information Theory, 1955. London: *Methuen* (1956).
2. ALLANSON, J. T. & WHITFIELD, I. C. The cochlear nucleus and its relation to theories of hearing. Third London Symposium on Information Theory, 1955. London: *Methuen* (1956).
3. BEURLE, R. L. Properties of a mass of cells capable of re-generating pulses. *Roy. Soc. Phil. Trans. B.*; 1956, **240**, No. 699, 55.
4. BURNS, B. D. The mechanism of after-bursts in cerebral cortex. *J. Physiol.*, 1955, **127**, 168.
5. BURNS, B. D. Some properties of isolated cerebral cortex. *J. Physiol.*, 1951, **112**, 156.
6. EAYRS, J. T. The cerebral cortex of normal and hypothyroid rats Acta. *Anat.*, 1955, **25**, 160.
7. ECCLES, J. C. The Physiology of nerve cells. *Oxford University Press*, (1957).
8. GALAMBOS, R. & DAVIS, H. The response of single auditory nerve fibres to acoustic stimulation. *J. Neurophysiol.*; 1943, **6**, 39.
9. HILALI, S. & WHITFIELD, I. C. Responses of the trapezoid body to acoustic stimulation with pure tones. *J. Physiol.*, 1953, **122**, 158.
10. RENSHAW, B. Central effects of centripetal impulses in axons of spinal ventral roots. *J. Neurophysiol.*; 1946, **9**, 191.
11. RUSSELL, G. Learning-machines and adaptive control mechanisms 1957. R.R.E. Memo. No. 1369.
12. SHOLL, D. A. Dendritic organisation in the neurons of the visual and motor cortices of the cat. *J. Anat.*; 1953, **87**, 387.
13. UTTLEY, A. M. The classification of signals in the nervous system. *E.E.G. Clin. Neurophysiol.*; 1954, **6**, 479.
14. UTTLEY, A. M. Conditional probability as a principle of learning. International Congress on Cybernetics Namur, 1956. *Publisher Gauthier-Villars, Paris.*
15. UTTLEY, A. M. The design of conditional probability computers. *Information and Control.* U.S.A., 1958. In the Press.
16. UTTLEY, A. M. The probability of neural connexions. *Proc. Roy. Soc.* (B), 1955, **144**, 229.
17. UTTLEY, A. M. The stability of a population of neurons. 1955. Unpublished work.

DISCUSSION ON THE PAPER BY DR. A. M. UTTLEY

DR. H. B. BARLOW: I am not sure that I have fully absorbed the implications of the modified scheme outlined at the end of Dr. Uttley's paper, but I think this modification brings his ideas and mine much closer together, and it may well answer some of the points I am going to raise.

One thing that worried me in the original conditional certainty scheme indicated in *fig. 8* was this: if it infers that b occurs when a occurs because $P(ab)$ is nearly equal to $P(a)$, then there is no way of distinguishing the rare event a from the common event ab. This does not seem biologically desirable, for rare events are often supremely important. The cat surely keeps a watchful eye for the rare mouse which fails to scuttle to its hole. Does the modified machine detect such rare events?

There are two other points I would like to know more about. Biologists tend to be intimidated by "requirements" proposed by mathematicians, physicists, or engineers, and therefore tend to accept them without argument. I do not, for instance, question the physicist's "requirement" that a light-sensitive receptor cell cannot absorb a fraction of a quantum. I don't feel that you want your "requirements" to be understood as absolute in that sense, but it might help us to understand them correctly if you were to indicate how much they depend upon your particular model, or how far they follow from general properties of machines using conditional probabilities.

The last point is on what Dr. Uttley describes (p.128) as the "intractable problem" of classifying 2^n possible inputs when n is of the order 10^6. There are only about 2^{30} neurones available, so the ratio of cells you have to cells you need is $2^{-999,970}$; in other words, you have virtually none of the cells you need. Of course our capacity to discriminate patterns may be very bad compared with a complete classification system, but I'm sure it's not as bad as that figure suggests.

The conclusion I would like to draw from this absurd ratio is that the nervous system cannot afford to leave to chance the pattern of effective connections a cell makes, and probably cannot allow it to be rigidly predetermined genetically because the environment is not rigidly predetermined. On the contrary, neurones must be allotted to input patterns

according to a scheme which achieves the utmost economy. Does Dr. Uttley's machine, either in its original or in its modified form, achieve this economy?

DR. F. ROSENBLATT: Dr. Barlow has really stolen half of my thunder here with his last comment. However, I would like to congratulate Dr. Uttley for being the first speaker on the programme to propose a theory rigorous and concrete enough in its references to actual physical or biological systems to permit me to take exception to it.

The main point of exception I would like to take is the quantitative one which I believe Dr. Uttley himself has recognised. He has proposed that it might be possible to replace the two to the Nth units which seem to be required for complete coverage of the sensory field by the number of combinations of N things taken M at a time where M is the number of units actually activated or illuminated by a particular stimulus. Or if we use some of the techniques which Dr. Taylor has recommended (*ref. 1*), we might even reduce M to the number of elements necessary to represent a contour which is still smaller. I still think it is perfectly clear that such techniques alone are really not sufficient to adequately reduce the number of units to anything like a biologically plausible system.

On the other hand, this does not mean we are required to reject Dr. Uttley's proposition that the brain is essentially a classifying system in some sense. I think this is a point which does require some clarification, however. There is an error in his assertion that in order to have a classifying system it is necessary to have a single unique element to represent each of the sensory events or patterns that his system may classify. In other words, in Dr. Uttley's system, if we have a hundred billion possible images of a cat on a retina this would require a hundred billion neurons representing these cats. However, a system of only a million neurons could easily represent one hundred billion events by assuming a hundred billion different states; that is the number of possible states is very much greater than the number of elements in the system. It is indeed possible to devise systems capable of classifying many more events than the number of elements in the system. Admittedly the classification is not perfect in these cases, but (as shown by the systems considered in my own paper) it can be made surprisingly reliable on a statistical basis.

I think it might be profitable for Dr. Uttley to reconsider his model

REFERENCE

1. TAYLOR, W. K. Electrical simulation of some Nervous System Functional Activities in Information Theory: Third London Symposium, Ed. by Colin Cherry. *Butterworths Scientific Publications, London* (1956).

in terms of representation of individual events by states of the system rather than by individual elements.

The suggestion that it is necessary to have a unit's delay of each of the impulses coming into the system in order to get recognition of time-patterns, I think is subject to the same sort of error in reasoning. It is not necessary to have a single unit to represent each possible sequence of events. That is, it is not necessary to have a unique unit to represent each entire sequence of configurations of impulses at times 1, 2, 3 and so on. It is really sufficient that the state of the system at time t should be some contingent function of the state of the system at time $t-1$, which is some contingent function of the state at $t-2$, and so on, and we can show that this accomplishes fantastic economies in the number of elements required. I think such systems will indeed prove to be economical.

I think it is out of place in the programme now to go into this in any detail. These comments really apply not only to Dr. Uttley but much more generally to a great many of the people working in this field. Dr. Uttley has fortunately presented us with a rigorous enough theory so that we can be specific. On the other hand, I think this kind of problem (the problem of "neuroeconomy") rears its head over and over again and must really be dealt with by the theorist if the theory is to prove more useful than a mere conjecture. I would like to suggest a methodological postulate (this really comes very close to what Dr. McCarthy would call commonsense) that wherever methods of gaining economy become essential in order to demonstrate that a proposed perceiving mechanism remains practical in a complex environment (not merely the environment of 2 or 3 impulses, but the environment of cats, dogs, and the things it must ultimately recognise) wherever this is the case, I think that a theory of the mechanism can be considered complete only if some specific means of economy are provided and worked out mathematically. Really I am developing a plea for something beyond heuristics here. I think we are all rather enraptured with heuristics at this point. On the other hand I think there is still some hope that rigorous mathematics may have some value. There are still one or two things that can be accomplished with it. I think its main place in this field is the study of the economy problem. I think many of the qualitative behaviours of our models can be discovered by heuristics, but I think these problems of the neuron economy almost always require a very careful mathematical evaluation before a theory can be acceptable.

DR. A. M. UTTLEY (in reply): Regarding Dr. Barlow's first point, the event a cannot be rarer than the event a and b. On the other hand the event a and not b can be rarer than a and b. This distinction cannot be

made in a system based on unitary classification, i.e. which counts only the presence of properties. For this to be done there must be binary classification with the operator *not*. All this is unrelated to the scale upon which probabilities are stored, i.e. upon whether the output at any unit is large for high probabilities, as in my original scheme, or large for low probabilities as in my modified proposal. The latter is very close to Dr. Barlow's proposal and both have the biologically useful property of giving large signals for rare events i.e. when there is a high information content in the incoming message.

On the second point, my "requirements" are necessary deductions from the original specification that conditional probabilities are to be computed.

On Dr. Barlow's last point, which is also Dr. Rosenblatt's first I consider that my modified proposal of a conditional rarity computer together with vast over-connection does achieve the necessary economy of units - with chance connections. Because of the over-connection there is, initially, ambiguity of classification, i.e. failure to distinguish, to a high degree. As information is accumulated those connections which carry little information become less effective until they are disconnected. Ambiguity is thus eliminated and the system learns to discriminate. I suggest that such a system meets the physiological and psychological facts very well.

Dr. Rosenblatt's second criticism is invalid. True, there can be 10^9 different images on a retina of 10^6 receptors; the retina alone can *represent* them all. But it cannot *distinguish* them all; for this we should definitely need 10^9 output channels, the first to indicate Cat. No.1, the second Cat. No.2, and so on. To distinguish all the patterns one would need a complete Boolean lattice with 2^{10^6} units. The vital point is that we know we are not seeking to discover such a system in the brain. To sum up, classification, with computation of conditional rarity, with over-connection, does seem to meet the requirements.

SESSION 2

AUTOMATIC PROGRAMMING

Chairman: DR. E. T. GOODWIN, NPL

		PAGE
1	Automatic programming: present status and future trends DR. GRACE HOPPER, Remington Rand Univac, USA	155
	Discussion on paper 1	195
2	Some technical features of the Manchester Mercury autocode programme MR. R. A. BROOKER, Manchester University	201
	Discussion on paper 2 *Chairman:* PROF. B. S. SOTSKOV, Academy of Sciences of the USSR	225
3	Automatic programming: properties and performance of FORTRAN systems I and II MR. J. BACKUS, IBM, USA	231
	Discussion on paper 3	249
4	The work of the Computing Center of the Academy of Sciences of the USSR in the field of automatic programming DR. A. P. ERSHOV, Academy of Sciences of the USSR	257
	Discussion on paper 4	275

MECHANICAL LANGUAGE TRANSLATION

Chairman: DR. L. MEHL, Ecole Nationale d'Administration, Paris

		PAGE
5	Tigris and Euphrates - A comparison between human and machine translation MR. R. H. RICHENS, Cambridge Language Research Group	279
	Discussion on paper 5	303
6	Pronoun reference in German MR. L. BRANDWOOD, Department of Classics, Manchester University	309
	Discussion on paper 6	341
7	Automatic translation in the USSR DR. A. P. ERSHOV, Academy of Sciences of the USSR	351

SESSION 2

PAPER 1

AUTOMATIC PROGRAMMING
PRESENT STATUS AND FUTURE TRENDS

by

DR. GRACE MURRAY HOPPER

BIOGRAPHICAL NOTE

Dr. Grace Murray Hopper, born New York, obtained her B.A. degree at Vassar College, Poughkeepsie and her M.A. and Ph.D. degrees at Yale University. From 1931-1944 she was first Instructor and later Associate Professor of Mathematics at Vassar College.

From 1944-6 she was a Mathematical officer with the U.S. Navy Computation Project at Harvard University, and, when released from active duty, she became a Research Fellow in the Computation Laboratory.

In 1949 she joined Eckert-Mauchly Computer Corporation, as Senior Mathematician. She went to Remington Rand Univac in 1950 as Senior Programmer and since 1952 has been a Systems Engineer and Director of Automatic Programming Development. Remaining in the Naval Reserve, she now holds the rank of Commander.

She is the author of numerous papers on Computation, Compilers and Automatic Programming.

AUTOMATIC PROGRAMMING
PRESENT STATUS AND FUTURE TRENDS

by

DR. GRACE MURRAY HOPPER

INTRODUCTION

IN engineering and scientific development, it is normal that a break-through relieving an immediate bottleneck, serves to either create or reveal another. The development of electronic computing systems solved the problem of computation and data-processing at high speeds only to create the need for programming and coding the problems. Within the last two years, much has been accomplished toward using the computer itself to replace human coders in the preparation of scientific, data reduction, and data processing problems for computation. The problem of system design or programming is under attack. A few more years should make it also amenable to the computer.

A brief review of the historical development will provide the definitions necessary for a discussion of the present status of automatic coding for data-processing problems. Two lines of development, now merged, can be traced: the assembly system approach and the subroutine approach.

THE ASSEMBLY SYSTEM APPROACH

IT is expected of automatic coding routines that their use shall offer two advantages:

(1) that they shall reduce the cost of coding;

(2) that they shall reduce the elapsed time required to prepare a program.

Obviously, the first is dependent on the second, but, in a given installation, either may be considered of primary importance. Reduction of cost is accomplished by reducing the amount of programmer time, coder time, and computer time required to produce a checked-out program. Reduction of programmer and coder time is accomplished by reducing the quantity of coding to be written and the resulting possibilities of making mistakes. Computer time to prepare a program is reduced by eliminating or materially reducing debugging time.

The mistake frequency of programmers and coders increases with the quantity of coding to be written. Short pieces of program can be written error-free, whereas long routines will inevitably contain mistakes. Typographical errors (reading 13 for a B or vice versa), inversions (writing 839 for 893) can be avoided or at least readily detected in short sections of coding. Following this line of thought, the development of a method by which a large program can be prepared in small sections and later pieced together will reduce the mistake frequency. It also permits division of a large problem among several coders, thus reducing overall elapsed time. Such a system involves a technique for "relative addressing" in order that the sections may be prepared independently.

Consider the problem of adding 100 quantities Y_1 to 100 quantities X_1, adding a constant k to each sum and storing the results as Z_1's. First, let the 100 X's be stored at locations 800-899; the 100 Y's at locations 1200-1299; k at 585; the 100 Z's at 1529-1628; and the instructions starting at location 115. The program then might be (in the instruction code of no known computor):

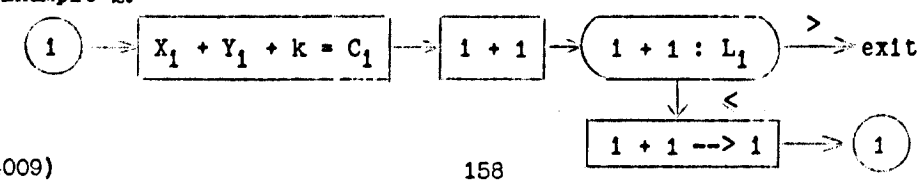
Example 1.

115	Get	800	X_0
116	Add	1200	$+Y_0$
117	Add	585	+k
118	Park at	1529	$=Z_0$ to Z_0
119	Get	115	(Get 800)
120	Add	586	(+1) (a "one" stored at 586)
121	Compare with	587	(Get 899) ("Get" 899 stored at 587)
122	If greater go to	XXX	exit from routine
123	Park at	115	If less, send (Get 801) to 115
124	Get	116	(Add 1200)
125	Add	586	+1
126	Park at	116	Send (Add 1201) to 116
127	Get	118	(Park at 1529)
128	Add	586	+1
129	Park at	118	Send (Park at 1530) to 118
130	Jump to	115	

This is a "fixed subroutine", it is not only written in computer code, but it refers to fixed storage locations.

Suppose, now, the assumption is made that the X's lie from A00 to A99; the Y's from B00 to B99; the Z's from C00 to C99; the k constants start at K00; and the instructions at M00. The subroutine can then be written as:

Example 2.



Example 3.

```
M15   Get              A00   X0
M16   Add              B00   +Y0
M17   Add              K00   +k0
M18   Park at          C00   =Z0
M19   Get              M15
M20   Add              K01
M21   Compare with     K02
M22   if greater go to XXX
M23   Park at          M15
M24   Get              M16
M25   Add              K01
M26   Park at          M16
M27   Get              M18
M28   Add              K01
M29   Park at          M18
M30   Jump to          M15
```

The example has now become a "relatively coded" subroutine. The type of relative addressing employed is one of the X-Series. Such relative addressing systems have been defined for many computers including not only the Remington Rand UNIVAC series, but also for some computers produced by universities or by competitors. For UNIVAC I, a single-address, alpha-decimal computer, the sample problem becomes:

Example 4.

```
M000   B   A000   A   B000
M001   A   K000   C   C000
M002   B   M000   A   K001   here K001 = 000001000001
M003   L   K002   T   XXXX        K002 = B0A099A0B099
M004   C   M000   B   M001        K003 = 000000000001
M005   A   K003   C   M001        called X-1 coding,
M006   0   0000   U   M000
```

and for UNIVAC II:

Example 5.

```
M000   B   A000   A    B000
M001   A   K000   C    C000
M002   B   K001   AH   M000
M003   L   K002   T    XXXX
M004   B   K003   AH   M001
M005   K   0000   U    M000        called X-1 coding
```

While for the UNIVAC 1105, which is a two-address, binary computer,

Example 6.

```
M000   RA   A000   ACC
M001   AT   B000   ACC
M002   AT   K000   C000
M003   RA   M000   K001
M004   RA   M000   ACC
M005   TJ   K002   XXX
M006   RA   M001   K001
M007   RA   M002   K003
M008   MJ   0000   M000          called X-4 coding.
```

For Computer UNIVAC-XYZ, which is one-plus-one address decimal

Example 7.

```
M000   25   A000   M001
M001   70   B000   M002
M002   70   K000   M003
M003   60   C000   M004
M004   25   M000   M005
M005   70   K001   M006
M006   30   K002   M007
M007   87   XXXX   M008
M008   60   M000   M009
M009   25   M001   M010
M010   70   K001   M011
M011   60   M001   M012
M012   25   M003   M013
M013   70   K001   M014
M014   60   M003   M015
M015   00   M000   0000          called X-2 coding.
```

Returning to the original example, the relative coding of Example 3 can be turned back into specific coding by stating that

A = 800 B = 1200 C = 1529
K = 585 M = 115

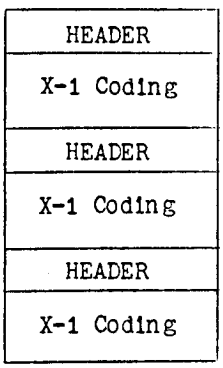
or it may be adapted to any other program by setting values for A, B, C, K and M.

Now a program can be assigned to several programmers, by telling them that

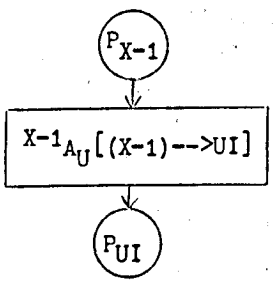
Input: inventory file A
transactions file B

Output: updated inventory file C
shipments file D
orders file E
errors file F

and that they are to use a common pool of constants K, working storage W, temporary storage T, etc. When all of the pieces and sections have been completed, they can be typed on magnetic tape, each section precoded by a header giving the values of M, A, K, C, T, etc. for that section. This X-1 input tape will have the format:



It will serve as the input to the X-1 Assembly routine designed to apply the information in the header to the following section of coding and thus to produce a running program.



Thus a problem stated in X-1 coding is transformed by the X-1 Assembly system (running on the UNIVAC I) into the same problem stated in UNIVAC I computer code. Here, A stands for assembly system

$X\text{-}1_A$ = X-1 Assembly System

$X\text{-}1_{A_U}$ = X-1 Assembly System coded for UNIVAC I or II

$X\text{-}1_{A_U}[(X\text{-}1)\text{-->}UI]$ = X-1 Assembly System coded for UNIVAC I or II to transform X-1 coding to UNIVAC I coding.

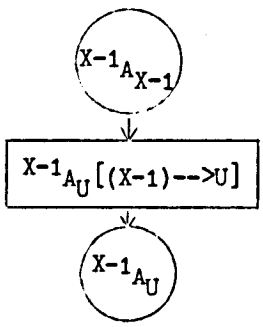
P_{X-1} = problem coded in X-1

P_{UI} = problem coded in UNIVAC I

Similarly $^{X-1}A_U[(X-1)\text{-->}UII]$ indicates the X-1 Assembly system coded for UNIVAC I or II to transform X-1 coding into UNIVAC II coding, and $^{X-2}A_U[(X-2)\text{-->}UXYZ]$ indicates the X-2 Assembly system coded for UNIVAC I or II to transform X-2 coding into UNIVAC XYZ coding. The problem of producing the original X-1 Assembly System was solved by writing by hand

$$^{X-1}A_{X-1}[(X-1)\text{-->}U] \text{ and } ^{X-1}A_U[(X-1)\text{-->}U]$$

then

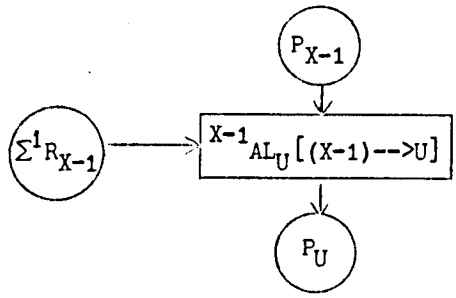


and comparing the computer output with the handwritten coding.

Almost as soon as the X-1 Assembly System was available and in use, the suggestion was made that it be used, not only to process a set of sections of coding but also that it be able to select and copy stored sections from a "library" on tape. Such sections are indicated by $^{1}R_{X-1}$, the 1^{th} subroutine written in X-1 coding; $^{Sin}R_{X-1}$ meaning the sine subroutine coded in X-1. The input to the assembly system now takes the form

HEADER
X-1 Section
HEADER
Library call
HEADER
X-1 Section

and the system may be diagrammed as

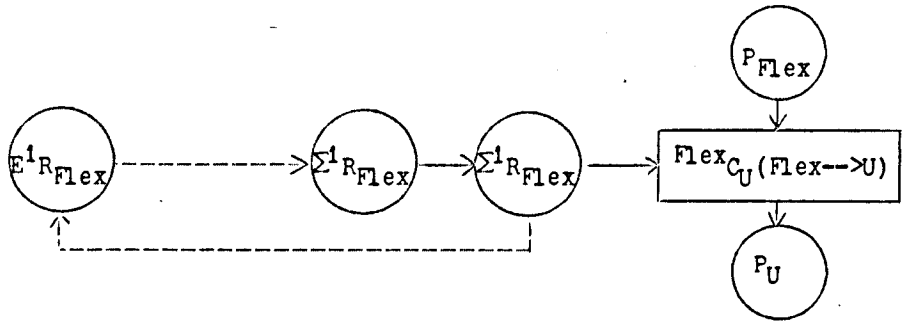


Since programmers and customers are never content, the next request (to save time and effort) is of course: "Why can't the computer write the headers?" Preparing the headers involves storage allocation for input, output, working and temporary storage, constants, and instructions and must precede the operation of the assembly system itself. It leads to the development of a "pseudo-code" to indicate which subroutines are desired from the library and an interpretive or translation routine to decode the pseudo-code. Thus a routine is required which will accept a pseudo-code, interpret the pseudo-code, select subroutines from a library, allocate both data and instruction storage, prepare headers for the required subroutines, and, finally, assemble them into a complete program. Such a master routine has come to be called a "compiler".

History, however, did not follow this logical path of development. The compilers first arose from the subroutine concept rather than from the assembly concept. Presently, the two lines of thought have merged, but before the present can be considered, the subroutine-slanted development must be outlined.

A COMBINED APPROACH

The FlexiMatic system, for UNIVAC I and II, developed under the direction of Dr. John Mauchly is an example of a combined approach. Here the problem is written in a flexible relative pseudo-coding system. The compiler as usual controls a library of subroutines and acts as a final assembly system. However, the library of subroutines is so constructed that one subroutine may call in another at will. Since subroutines may be constructed from subroutines, the creation of an effective library is greatly facilitated. The FlexiMatic system may be diagrammed as:



This automatic system is of particular advantage in the preparation of new and unorthodox programs requiring new subroutines. It appeals strongly to the sophisticated and mature programmer and offers him a wealth of services.

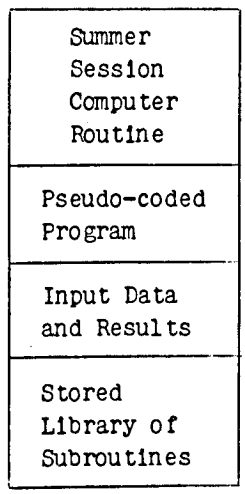
THE SUBROUTINE APPROACH

Subroutines were recognized in the very first large digital computer. MARK I at Harvard, as designed by Professor Howard H. Aiken, contained wired in subroutines for cos x, log x, and exp x. Since these functions were wired into the machine, they, of necessity, offered complete generality: cos x accepted any value of x in radians between minus infinity and plus infinity: it delivered the function, cos x, accurate to 21 decimal places. Almost never was such a degree of accuracy required. If x represents the angle of roll of a ship, it is surely known that $|x|$ is less than $\pi/2$. If the problem data is submitted with five decimal places, 21 places are not required in the cosine function. The wired-in subroutines proved to be too general. Very soon, subroutines were programmed, such as one to compute sin x to 8 decimal places for $|x| < \pi/4$.

Sets of such subroutines developed in the note books of the programmers and it became common practice to say "Joe, can I borrow your sine routine?" But it was wartime, and there was never enough time to organize and systematize this information. Soon, however, the EDSAC crew was gathered together, *before* the computer was finished. They had time to develop and systematize a set of subroutines and provide the first step in automatic coding. They gave us also the first publication in this field, the book by Wheeler, Wilkes, and Gill *(ref. 1)*, "Preparation of Programs for an Electronic Digital Computer". This provided a strong thrust in the direction of building routines to assist the coder.

At the Massachusetts Institute of Technology, it was desired to use WHIRLWIND I, a binary computer with a 16 bit word, for the benefit of students in the summer sessions. The outgrowth of this need was the "Summer Session Computer" (1952-53) a routine which accepted programs written in an alpha-decimal pseudo-code, interpreted the pseudo-code, and executed the

required operation by means of a stored library of double-precision subroutines, before proceeding to the next pseudo-code word. Thus the layout of the storage became, schematically;



It was then possible to write a program in the form

CA 107 0 $\longrightarrow$acc ; (107) $\longrightarrow$ acc

AD 109 (acc) + (109) $\longrightarrow$ acc

and expect the computer to carry out, under control of the interpretive routine, the following operations:

1. examine the word CA 107,
2. convert the decimal address 107 to binary,
3. locate the quantity stored at this address and place it in the augend location,
4. examine the word AD 109,
5. convert the decimal address 109 to binary,
6. locate the quantity stored at this address - and place it in the addend location,
7. arrange for a return jump to the double precision - add-routine stored in the library,
8. execute the stored add-routine,
9. return control to the interpretive routine to pick up the next pseudo-code instruction.

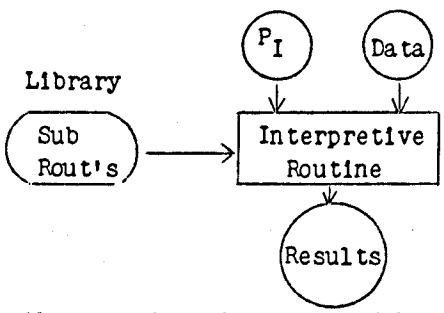
Pseudo-codes and interpretive routines of this type multiplied. Among the best known were the "Short Code" for UNIVAC I (1951) and Speed Coding for the IBM 701 (1953).

A disadvantage of all of the interpretive routines was the necessity of reinterpreting each instruction each time it was executed in the course of

program loops. After the instruction had been interpreted once, the recurring re-interpretations represented non-profitable time - time during which the computer did not "make any numbers". The seeming advantage of the interpretive routines lay in the fact that, at that time, the input and output of most of the computers were manually operated. This placed a premium on packing as much data and as many instructions as possible into the storage each time a problem was loaded. The use of the library of subroutines assisted this packing: the presence of the interpretive routine itself operated against it.

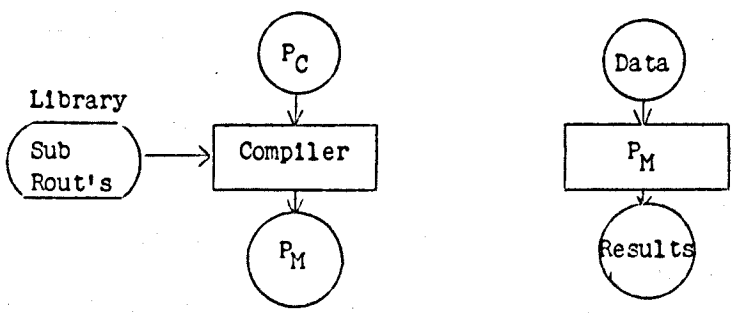
As new computing systems appeared in which the input-output equipment was under control of the central computer, and in which there were multiple input-output units so that they could be used as secondary storage, a new situation was presented. The first of these data-processing systems was the UNIVAC I and it was a natural development that the first "compiler", A-ϕ (1952), was designed for the UNIVAC I. Where the interpretive routine completed a problem in a single pass, the compiler broke the process into two passes and supplied as a by-product the computer-coded problem.

I. Problem in interpretive pseudo-code plus problem data processed by the interpretive routine to produce results:



II. a. Problem in compiler pseudo-code processed by the compiler to produce a machine-coded program for the problem.

b. This program is used to process the data and produce results.



Thus the compiler

1. examines the word of pseudo-code,
2. looks up the required subroutine stored in the library in relative coding,
3. adjusts it to receive the indicated argument and result addresses,
4. positions the subroutine and processes it into fixed coding,
5. writes it as part of the final running program,
6. returns to examine the next word of pseudo-code.

A typical pseudo-code, that of the A-2 Compiler for the UNIVAC I (1953) took the form:

```
AM0 008 000 018     calculate cx
TS0 018 000 020     calculate sin cx
AM0 016 020 024     calculate e^(-x^2) . sin cx

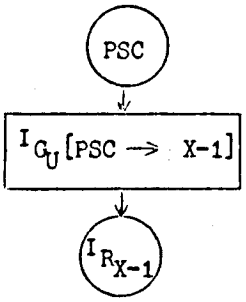
where c is stored in working storage location  008
      x  "   "  "   "       "       "        000
     cx  "   "  "   "       "       "        018
  sin cx "   "  "   "       "       "        020
 e^(-x^2)"   "  "   "       "       "        016
and e^(-x^2) sin cx "   "  "   "       "      024
```

Thus the compiler (a routine on tape) plus its library of subroutines (on tape in relative coding) effectively converted the UNIVAC from a single-address, fixed-decimal computer into a three-address, floating-decimal computer. Use of the pseudo-code reduced the amount of coding to be written by a factor of twenty and with this the number of mistakes by an even larger factor. Further, it was possible to teach this pseudo-code to mathematicians, engineers, statisticians, and other scientists in a few days time in contrast to the weeks required to learn computer code.

This represented a major step forward. However, after a few months in use, it was discovered that the library included about 28 mathematical subroutines, 37 input routines, and 48 output routines, and that each new problem was tending to require a new input routine and a new output routine. The input routines were examined and it was discovered that they were all of essentially the same nature and based on four parameters;

1. the tape (or servo number) to be read,
2. the input storage area through which the data was to be transferred (A or B, or C ...),
3. the number of words to be transferred in each item or record, (S),
4. the position in working storage to which the quantities were to be transferred (W).

A new routine, called GMI, "generator-move-in", was prepared and placed in the library which was capable, upon receiving the pseudo-coded word as input, of creating or *generating* as its input the required subroutine:



Pseudo-code to input generator running on UNIVAC produces as output a specific input routine in X-1 relative coding.

Pseudo-code to input generator running on UNIVAC produces as output a specific input routine in X-1 relative coding.

The pseudo-code word was written in the form:

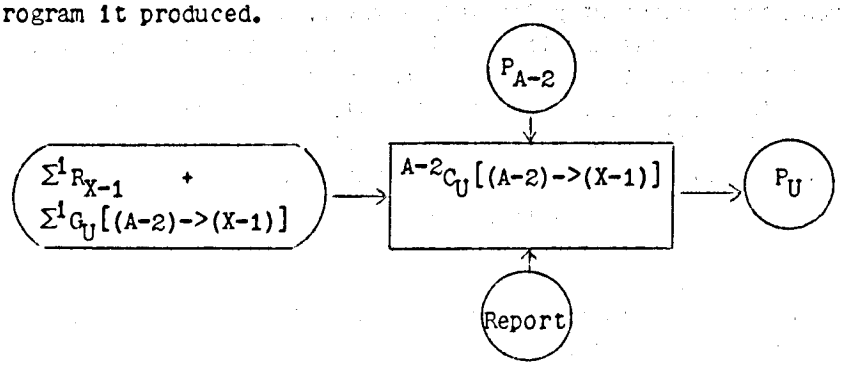
GMI OTA OSS WWW
B

or

GMI 02B 010 122 to generate a subroutine in X-1

relative coding to read from servo 2 through input area B, in 10-word items, to working-storage locations 122-131. All of the input subroutines in the library were replaced by this generator. A similar generated GMO (generator-move-out) replaced the output subroutines.

These two were the first operating generators. Since that time many have been developed and it is their contribution that has made possible the development of data-processing compilers. The A-2 Compiler performed other services in the course of constructing a program besides calling for static or generated subroutines and processing them. It also allocated the storage for data and instructions, arranged for segmenting if more than a single storage load of instructions was required, and prepared a report on the program it produced.



A problem written in A-2 pseudo-code is processed by the A-2 Compiler (running on UNIVAC) from A-2 code to UNIVAC code, making use of static subroutines coded in X-1 and generators (running on UNIVAC) to derive X-1 subroutines from A-2 code.

A DIFFERENTIATOR

Armed with the A-2 Compiler to program mathematical problems, it seemed natural to proceed to the calculus. A few changes were advisable in the A-2 code before the computer was to be inducted into differential calculus; i.e.,

```
ADD 00X 00Y 00Z     x + y = z
MUL 00U 00V 00W     uv = w
```

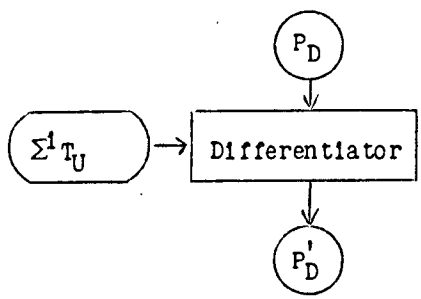
It was agreed that 1x indicated dx

12y indicated $d^{12}y$.

Next a routine, called the Differentiator was prepared which examined each pseudo-code word of a function presented to it and determined the operation to be performed (the first three letters of the word). This "call-word" served to select a "task" routine from a tape library which produced the derivative of the function defined; e.g.,

```
                ADD 00X 00Y 00Z     x + y = z
      yields    ADD 01X 01Y 01Z     dx + dy = dz
+     and       MUL 00U 00V 00W     uv = w
      yields    MUL 00U 01V A19     udv = a19
                MUL 01U 00V A20     vdu = a20
                ADD a19 a20 01W     udv + vdu = dw
```

Higher derivatives can be produced by reapplication of the task routines.



ALGEBRAIC TRANSLATION ROUTINES

In 1953, Laning and Zierler, at MIT set to work to produce a pseudo-code more nearly resembling normal mathematical statements than did the pseudo-code employed by either of the major MIT routines. They produced "A Program for Translation of Mathematical Equations for Whirlwind I" *(ref. 2)*. This routine was followed in 1956 by the Math Matic compiler for

UNIVAC I and II, Fortran in 1957 for the IBM 704, and UNICODE for the UNIVAC 1103A will be completed by the end of the year. All three of these routines accept mathematical symbology and English verbs.

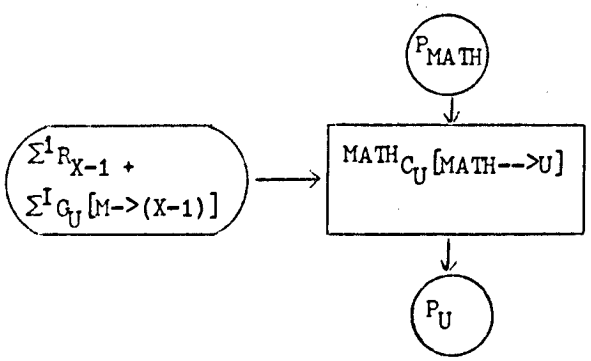
Consider a very small example:

Compute $$Y = \frac{X^3(2 + X)}{3 \cos A} - \sqrt{P}$$

for $0.2 \leq P \leq 0.8$, $P = 0.2$
$0.35 \leq A \leq 1.05$, $A = 0.175$
$1.8 \leq X \leq 3.8$, $X = 0.5$

Coded in MathMathic, the problem becomes

(1) VARY P 0.2(0.2) 0.8 SENTENCES 2 THRU 5.
(2) VARY A 0.35(0.175) 1.05 SENTENCES 3 THRU 5.
(3) VARY X 1.8(0.5) 3.8 SENTENCES 4 THRU 5.
(4) Y = X^3 * (2+X)/(3* COS A) - $P^{1/2}$.
(5) WRITE AND EDIT Y, X, A, P.
(6) STOP.



A problem written in MathMatic pseudo-code is processed by the MathMatic compiler using a library of subroutines coded in X-1 and generators yielding X-1 subroutines to produce a program written in UNIVAC code.

A meeting was held in Zurich last summer between representatives of the Association for Computing Machinery and GAMM, the German Association for Applied Mathematics and Mechanics, to start to define an international mathematical language. A report of their efforts will soon be issued in the "Communications of the ACM".

DATA PROCESSING COMPILER

Meanwhile, with the data-processing systems, UNIVAC I, coming off the production line, a need became evident for a new type of compiling routine -

one to prepare data-processing programs as contrasted to the mathematical type of programs. Work was begun in 1954 and several serious problems immediately became apparent.

1. There existed in commercial and business systems no common symbology, and, in fact, worse, no common terminology. Not only did no two installations use the same terminology but even divisions within a company differed.
2. No longer was the computer to deal with quantities - decimal or binary numbers - fixed in length and type of decimal notation and location. It must deal with alpha, alpha-numeric, and numeric quantities, with and without decimal points in various positions. Several quantities were often packed together in a computer word, or a single quantity or field might spread out over several words,
3. The operations to be performed often required many operands, yielded many results, and operands could also become results by manipulation.

These problems were reduced to

1. A language or symbology must be developed, acceptable to all installations, by means of which operations could be stated.
2. A method must be devised for describing data.
3. A multi-address pseudo-code must be developed.

While the solutions to these problems to be here presented will seem simple and almost obvious, it must be stated that these solutions were not arrived at directly but via many pitfalls, detours, and over many obstacles. The FlowMatic compiling system first operated in 1956 and is still being improved.

FLOWMATIC - PSEUDO-CODE

It was determined that the only common symbology used by the various installations was the English language and it was therefore decided to use a selected English vocabulary as a pseudo-code. It was possible to isolate and define 30 verbs

ADD	INSERT	SELECT
CLOSE-OUT	JUMP	SELECT-LEAST
COMPARE	MOVE	SET
COUNT	MULTIPLY	STOP
DIVIDE	NUMERIC-TEST	SUBTRACT
EXECUTE	OVERLAY	SWITCH
FILL	PRINT-OUT	TEST
HALT	READ-ITEM	TRANSFER
IGNORE	REPLACE	TYPE-IN
INPUT	REWIND	WRITE-ITEM

which as a start would make possible definition of data-processing problems in a common manner. Thus, a sentence could be written

(19) COMPARE INV-ON-HAND WITH INV-LIMIT; IF GREATER GO TO OPERATION 29; IF EQUAL GO TO OPERATION 23; IF LESS GO TO OPERATION 23.

Even such a simple statement turned out to be fraught with traps! The first user dropped the last two clauses and wrote "OTHERWISE GO TO OPERATION 23". The computer stopped, indicating that it did not recognize the word "otherwise". It was necessary to provide for a large number of ways of ending the sentence including "IF UNEQUAL GO TO OPERATION N". The second user did not choose to compare one field WITH another. He elected to compare one field AGAINST another. It was necessary to inform the compiler that any one of a set of words, WITH, TO, FOR, AGAINST, or even a colon, was acceptable between the two noun-names of the fields. Coping with variations in English statements did not end the problems: next came misspelling.

Suppose someone wrote COMPAIR instead of COMPARE. The first time this occurs the computer stops, states that it cannot locate this verb, and requests a correction. The second time someone makes this same error, the computer stops and types out,

"YOU WROTE COMPAIR, DID YOU MEAN COMPARE, IF SO, HIT THE START BAR"

Thus, the symbology selected for the data-processing compiler was restricted to simple English sentences and a preliminary set of verbs was defined. This provided an answer to the first and last problems and left the nouns to be treated.

FLOWMATIC - DATA DESCRIPTIONS

While 30 verbs common to all data-processing problems could be distinguished, no such list of nouns could be prepared. It became clear that each installation must be permitted to define its own nouns, thus describing its own data. It was determined that these definitions must include descriptions of

1. the data as a file or set of reels of tape,
2. the data as items or records of sets of words, and
3. the data as fields or sets of characters.

In order to describe the data as a file, the noun name by which the file will be referred to must be given, the label carried on each tape of the file and its position on each reel, and the sentinels used to indicate end of reel and end of file and their relative positions. This information provides the necessary parameters for the input-output generator to supply all of the tape-handling coding including label and sentinel checks, reading tape, and servo swapping.

To describe an item in a file, in order to position the items for processing, it is necessary that the size be stated (if fixed length) or a description of headers and trailers or an ending convention (if variable length). The names of the fields, or "keys", dictating the sequencing of the file are required for matching, sorting, and merging.

Finally, each field is given a noun name and descriptor stating the type of field, its relative position in the item, its length, and the position of its decimal point and sign, if any. These definitions and descriptors are stored in the library of the particular installation and are thus available for use in the pseudo-code.

A file might be named "inventory", its label could be INV063000001. It could consist of 10-word items corresponding to each "stock-number" and, if arranged in ascending order of "stock-number", that field would become the key. A field such as the stock-number could be described as positioned in the first word of the item, alpha-numeric, starting in the first digit of the word, consisting of ten digits and having no decimal point or sign. Similarly, INV-LIMIT (inventory limit) might lie in the sixth word, be numeric, start in the seventh digit, consist of six digits and have no decimal point or sign. On the other hand, UNIT-PRICE could lie in the fifth word, be numeric, start in the eighth digit, consist of four digits, the decimal point positioned two digits to the right of the beginning of the field and no sign. Simple forms were designed to facilitate the writing of these data descriptions. Only a few weeks later a request was received for "a routine to fill out the forms". This routine now exists.

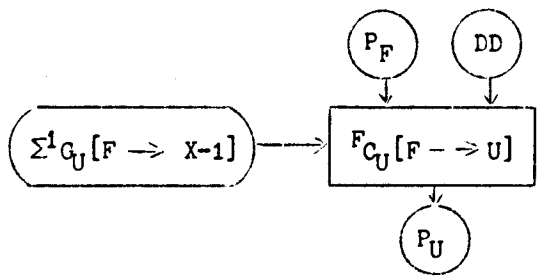
FLOWMATIC GENERATORS

As may be seen from the data descriptions the operation "compare inv-on-hand with inv-limit" might indicate a simple comparison of two computer words. More likely, the phrase would require that the two quantities be extracted from surrounding fields and shifted to line up decimal points before the comparison could be made. Corresponding to each permissable verb, there is a generator in the library which, upon receiving as input data descriptions and parameters, is capable of generating a relatively coded subroutine to perform the particular operation upon the specific fields named. Obviously, these generators must be quite sophisticated and some, such as the input-output generator, are extremely powerful. The latter may deliver as a result of a single input sentence between 300 and 2000 coded computer instructions. It may be informed that an output tape is to be prepared for the High-Speed Printer or that the "preselection" rather than "stand-by" technique of handling input is desired. It is essentially a small compiler including generators within itself. Other generators such as JUMP,

MOVE, and EXECUTE proved to be comparatively simple. Three generators HALT, IGNORE, and STOP are almost trivial but turned out to be quite essential.

THE FLOWMATIC SYSTEM

The input to the FlowMatic system consists of a library of generators (corresponding to verbs) and data descriptions (corresponding to nouns). A problem is written in English sentences - FlowMatic pseudo-code. Processing by the FlowMatic compiler produces a computer-coded program ready to be run on UNIVAC I or II (depending upon the library employed).



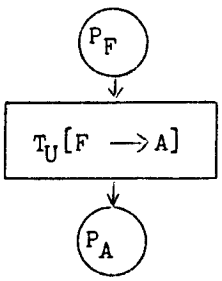
The compiler itself operates in three phases controlled by three routines called Translator, Selector, and Converter.

The translator (T_U) processes the pseudo-code and refers to the data descriptions. It produces an expanded problem description P_A. Each sentence of the English pseudo-code is expanded into a "file entry" in the form

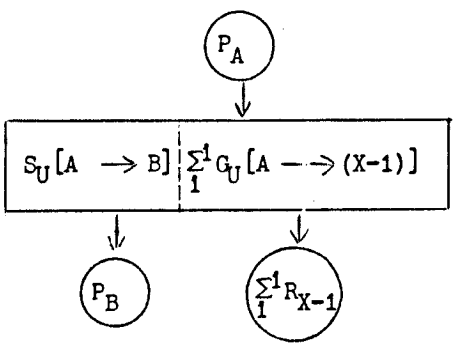
```
OP NUMBER      19
CALL WORD
COMPARE        (verb)
VARIABLES
INV-ON-HAND    (noun and abbreviated descriptor)
0053777716
INV-LIMIT
0063777776
PARAMETER
4              (sentence ended "if greater" - "otherwise")
JUMPS
J29
J23
END OF ENTRY
```

This file of expanded sentences can be considered to be a description of the problem in a computer-oriented language - if a systematic method of conveying information may be called a language. This "language" is curiously general. It is not based on the word structure of any computer

nor are its symbols dependent upon position. Each entry is preceded by an English word indicating the character of the information to follow. It uses the 26 letters of the alphabet, the 10 decimal digits, and characters for "space" and "ignore".



The selector accepts as input the problem description P_A and controls and provides services to the library of generators.

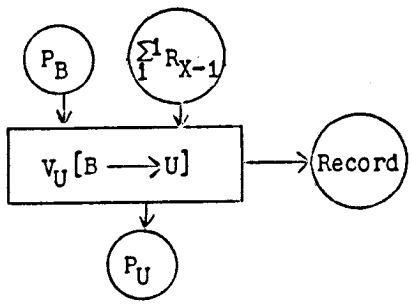


As the selector scans the file entries of P_A it summons the appropriate generator and supplies to it the information contained in P_A as required. The generator produces the subroutine, in X-1 code, required by the data descriptions. The generator also sends to the Selector information concerning the subroutine it has produced. The Selector adds this information to the P_A entry to produce a P_B file entry which in addition to P_A will now contain the start line, the end line, the entrance and exit lines, a list of needed constants, temporary storage and working storage, the transfer registers and other subroutines required by subroutine R_{X-1}.

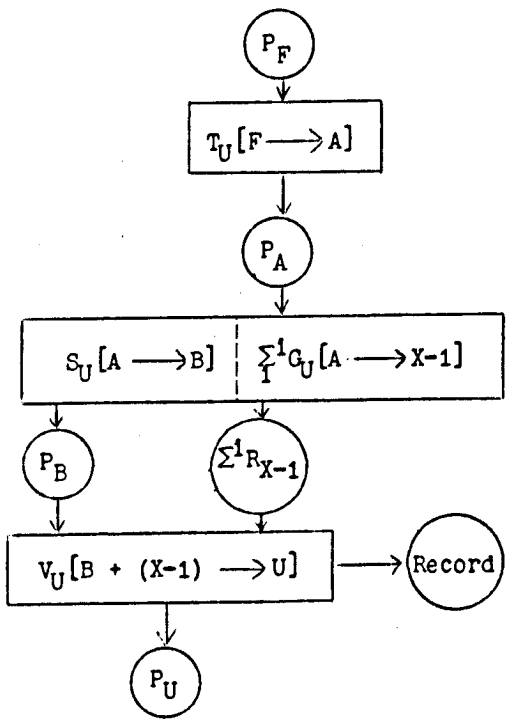
The file P_B contains listings of all the components of the final program, storage requirements for data, variables, constants, working and temporary storage, and instructions (subroutines). In addition, it contains the necessary information concerning relationships such as jumps and return jumps, entrances and exits.

The final phase of the compiler, the converter, allocates the storage in the most economical fashion assigning specific locations to all data and instructions and finally acting as an X-1 assembly system, produces a

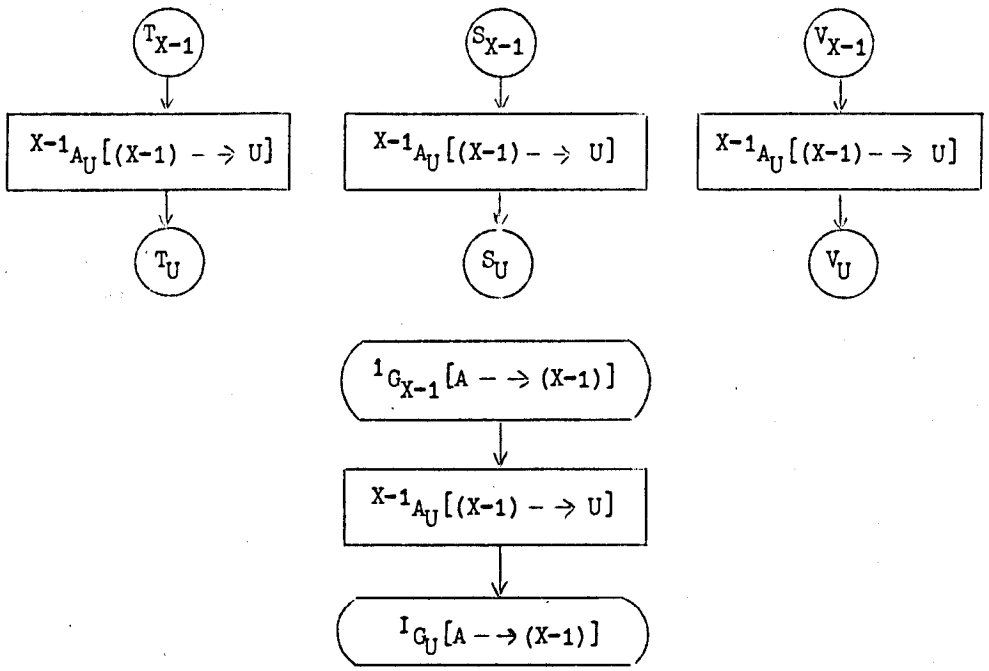
running program. Since at the time of allocation complete information was available concerning all data and instructions, it also produces a report called the "record". The record supplies a copy of the original pseudo-code and the data descriptions drawn from the library. It lists the storage assignments for input and output data and for variables. Corresponding to each pseudo-coded sentence it lists the relative X-1 coding produced by the generator together with the corresponding final fixed coding and its location.



Thus the compiler as a whole may be represented as:



The Translator, Selector, Processor, and generators of the basic compiler of the FlowMatic system were all produced by means of the X-1 Assembly System; i.e.,



A discussion of the application and use of this first FlowMatic Compiler must precede the discussion of further extensions of the system.

APPLICATION OF THE FLOWMATIC COMPILER

As with any new tool, experimentation was necessary to determine training time, operating time, and overall feasibility of using the compiler. Several experiments were carried out.

A man trained in transportation problems, but not in computers was selected and, forbidden to learn computer code, was trained in English language coding. He quickly learned the technique of preparing process charts or flow-charts in English using the assigned vocabulary, only to rebel at then writing out the English sentences. He felt that the logical design of the problem was complete and preparing the pseudo-code was a clerical job. Another experiment was started. A young girl, just graduated from high school, was taught the vocabulary and trained to write the pseudo-code from the process charts. "Division of labor" was rediscovered, a team consisting of

trained systems analyst and clerically trained pseudo-coder produced accurate programs faster. A new job title was listed - pseudo-coder.

Four installations agreed to try out the new system on real problems Installation A, (an old installation) was not able to make good use of the system. It was later determined that this was caused by the fact that the programmers feared that the compiler would put them out of a job. They could not see that its eventual effect would be to upgrade them to analysts. This was a management communications failure rather than a failure of the compiler.

Installation B was a brand new installation. The new programmers were more than willing to reach for any help that they could get. The compiler has been most useful. In their major application consisting of 80 computer runs, 76 were prepared using automatic techniques. Their programs were checked out before the computer was delivered and at considerably less cost than they had anticipated.

Installation C was an old installation whose programmers at first were not convinced that the computer could produce as efficient coding as they could produce by hand. However, because of extreme pressure to prepare a major problem consisting of over 700 computer runs in about a year, they adopted the compiling technique. Their evaluations and experiences proved extremely useful in providing the necessary suggestions for improving the efficiency of the system. Two years ago it required roughly 670 msec. per sentence (on UNIVAC I); to compile a program today, this figure is 480 msec. per sentence and for UNIVAC II it is about 235 msec. per sentence.

Installation C has prepared over 300 running programs. Their average compiling time per run is less than an hour and average debugging time per run about ten minutes, indicating that the majority of the runs require no debugging.

Installation D, the Air Materiel Command, Dayton, gave the compiler the most complete testing of all. It had been determined that a two weeks course was sufficient time in which to teach people unacquainted with computers how to write programs using FlowMatic. A course was set up for AMC personnel. They included three systems analysts, four programmers, two career sergeants and the major's secretary. All completed the course successfully. AMC then set up a project and, keeping careful account of time, determined that - including problem definition, analysis, coding and debugging, - the use of FlowMatic supplied a four to one improvement over manual techniques.

As the reports came in from these test installations, others tried out the compiler. Its use is well on the way to becoming general. Certain things had been proven:

1. the compiler reduced elapsed time by reducing
   (a) analysis time a little bit by supplying a vocabulary and a technique for process charting,
   (b) programming time by a factor of at least five,
   (c) by eliminating coding time and
   (d) by almost eliminating debugging time,
2. the use of the compiler reduced training time,
3. the compiler assisted in standardizing the formats of data and programs,
4. a new communication system was introduced extending from manager, through analyst, programmer, coder, and operator to the computer,
5. runs could be easily altered and updated by correcting the pseudo-code and recompiling,
6. a permanent record was maintained of the detailed content of each running program,
7. and finally, all of these contributed to a reduction in the cost of programming and operating a computer.

Such economies immediately posed the question "could the system be extended to other computers? At the same time the use of the pseudo-code language provided a badly needed means of communication.

COMMUNICATIONS IN A DATA PROCESSING OPERATION

The FlowMatic Compiler, designed to answer the need for more rapid coding techniques, surprised its creators by also fulfilling a long felt need for a better communication system throughout the Air Materiel Command. The following quotation is from a speech given by Colonel E. R. Miller, USAF at the USE Conference in August, 1958.

> ".....to a degree, we have made no appreciable progress in reducing or eliminating two very positive areas of difficulty."
>
> "The first area is that there is a lack of a common communication media between our Headquarters management personnel, our Depot operating management, and our computer programming group, particularly, during the development of EDPE systems.
>
> "The second area of difficulty is the difficulty of coding instructions for intricate equipment needed to produce the proper data processing. At the present time within the Air Materiel Command, we are still attempting to develop our EDPE systems in the following manner: each one of these is based upon a Command policy or operating philosophy established by a basic document, such as an Air Force Manual, a Command Regulation, Military Specifications, and Cataloging Directives. Frequently, these are not comprehensive and must be established by a steering committee composed of representatives from the three groups mentioned above. This source of basic guidance is finally completed to the point it is comprehensive but is not documented in the precise and sequential order needed for direct conversion to EDPE.

"At this point, a working group, task force, or staff element is composed which usually attempts to develop, concurrently, both the precise systems plan or strategy, and the machine statements and language, or machine tactics, to implement the systems strategy. At the outset, operating management plays an agressive part, but there is a gradual withdrawal of these people, which is due, perhaps, to their lack of understanding and the emphasis placed on data design and the functional machine runs, with a corresponding lack of emphasis on the management derivations, which should be considered.

"During the development phase, the computer programming group produces flow charts that are in a format for a specific computer with its own symbolic notations that are readily intelligible only to the skilled programmer. This is the blueprint our operating managers should review and evaluate but it is normally unintelligible to them. None of these managers doubt the arithmetic accuracy of a computer but they are unwilling to accept answers based upon the logical decisions employed by the program. A single example might be that a requisition becomes a backorder and the management personnel will examine, manually, stock available at other storage points, availability of substitute items or the possibility of directing a shipment from a contractor where the machine, itself has already exploited these possibilities.

"The third phase is the fabrication of the machine instructions themselves. As each equipment has its own specific code, relating to the internal functions being performed and not relating to the processing data function, the coding group may be diverted from the proper logic of the data processing system. Coding, itself, is a tedious and demanding type of work, subject to a certain amount of human errors. As these programs must be 100% precise, the debugging cost can be disproportionately high. It is not until the final products are obtained that we get an objective evaluation to our development effort. This usually calls for a realignment and reprogramming to incorporate the design changes that meet management requirements.

"In the final analysis, we find that the efficiency of our EDPE programs is largely a function of the effectiveness of the computer programming group, who today exercise a control that was never intended that they should have and is actually one they are reluctant to carry by themselves.

Having told of testing out the FlowMatic system as previously discussed, and of the institution of the AIMACO project (to be described later), Colonel Miller went on to list the benefits AMC expected to obtain by using the English language compilers:

"1. Communication between all offices will be enhanced by use of this English Type Pseudo-Code, which largely removes the barrier now existing between operating management and computer programming groups.

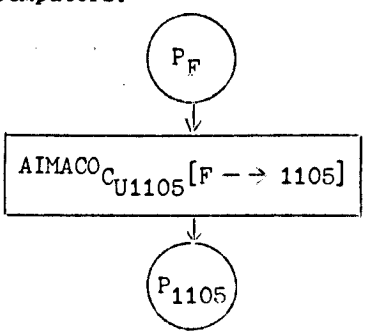
2. Provides AMC with an effective control of all systems -
   a. In developing stage.
   b. In processing stage.

3. Reduces time schedule -
   a. Problem definition - insures completeness and clarity.
   b. Permits "Pilot" Depot assignments without regard of equipment.
   c. Increases coding capability of present AMC staff - estimated to be 450.

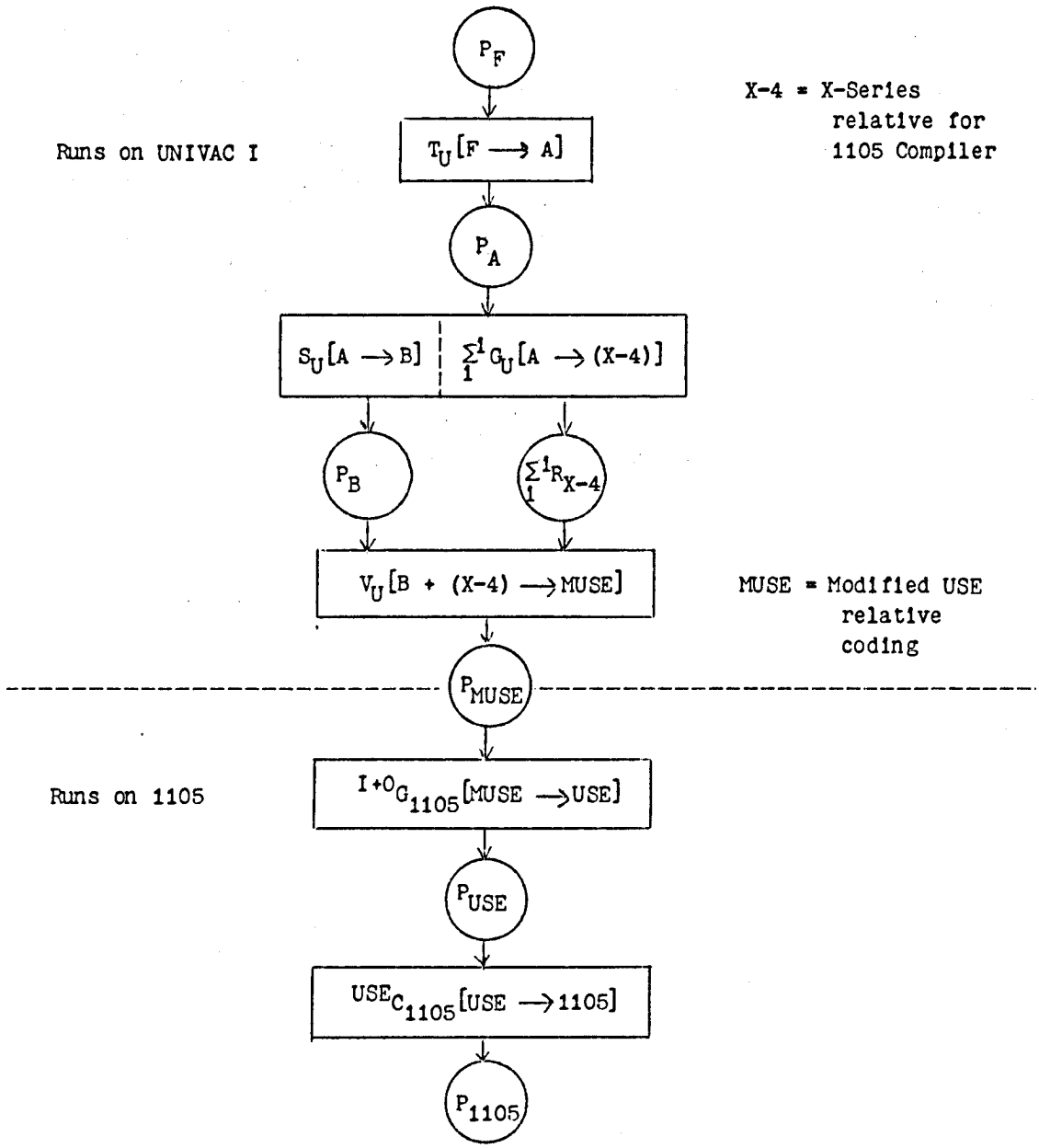
4. Reduces problem of systems modification - either in operational instructions or data.

5. Direct comparison of different EDPE now possible."

It is not immediately evident, but the first two phases of the FlowMatic compiler are independent of the computer to be programmed. The Air Materiel Command needed English language compilers for all of their computers. The next series of computers to be delivered were the Univac 1105's. It was decided to set up a joint project of AMC and Sperry Rand to produce a preliminary Air Materiel Command -AIMACO- Compiler making use of all existing routines as far as possible. These included the FlowMatic compiler (and AMC had a UNIVAC I in Dayton) and the USE compiler (Univac Scientific Exchange) as developed by Ramo-Wooldridge. The AIMACO compiler is then, a hybrid, running on two computers.



Diagramming it in all its phases:



Runs on UNIVAC I

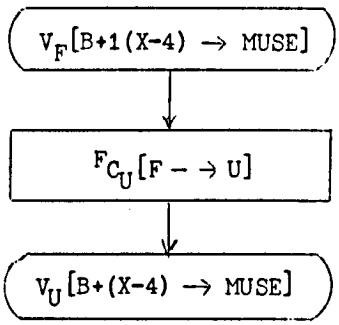
Runs on 1105

X-4 = X-Series relative for 1105 Compiler

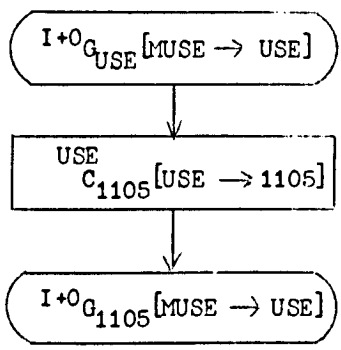
MUSE = Modified USE relative coding

Three components had to be prepared, the converter $V_U[B + (X-4) \rightarrow MUSE]$, the input-output generator $^{I+O}G_{1105}[MUSE \longrightarrow USE]$ and the set of generators $\sum_1^1 G_{1105}[A \longrightarrow (X-4)]$.

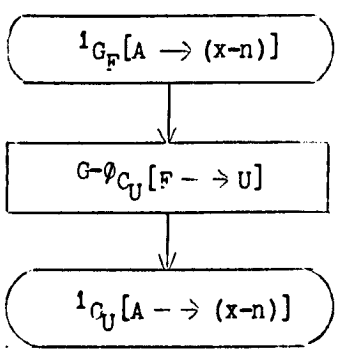
The converter was prepared by using the FlowMatic Compiler; i.e., it was treated as the data-processing routine that it is, and was written in English sentences:



The input-output generator could be prepared using the USE compiler; i.e.,



Finally, the FlowMatic compiler was modified into a new compiler - C-ϕ - designed specifically for the production of generators.

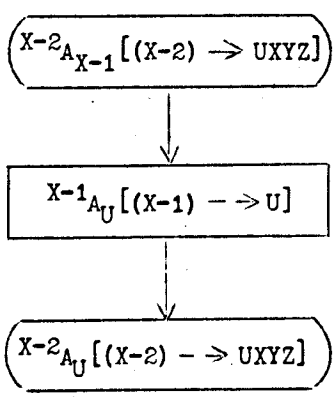


The components necessary to complete the AIMACO compiler were developed automatically. Colonel Miller's comments on the AIMACO project indicate not only the AMC interest in AIMACO but also their confidence in future developments:

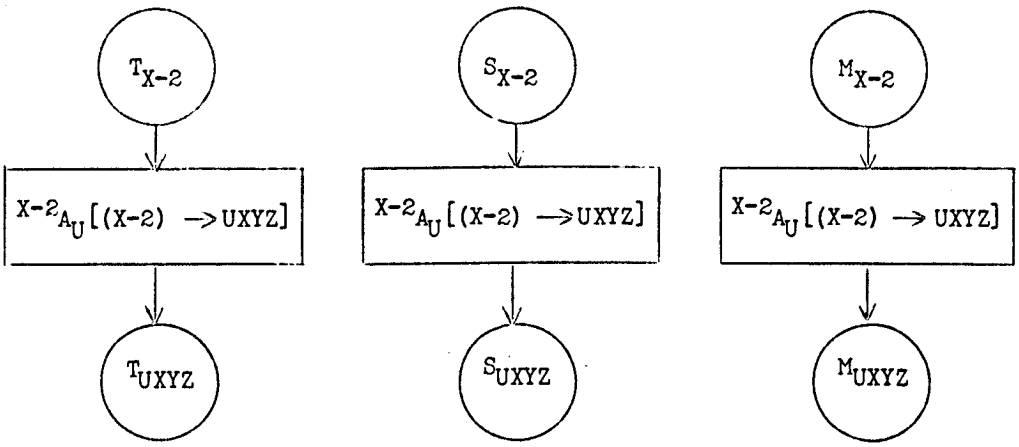
"Recently there has been considerable discussion and effort towards the development of a universal computer language. While Air Materiel Command could make excellent use of such language, it cannot wait several years for its development. Several articles have appeared which proposed the use of set theory notation as part of this language. To us, it doesn't appear to make much difference whether it is set theory notation or machine code which our people do not understand. The common language with us, in Air Materiel Command, is English. The advantage of using English and having an already proven technique in FlowMatic, seems to us to be overwhelming advantages in the AIMACO project. AIMACO is an attempt to provide more than just a new compiler for Air Materiel Command. It is an attempt to provide a complete method beginning at the first definition of the application through to the check-out and system refinements, that would inevitably come. It encourages use of English throughout the method AIMACO contains, within itself, the capability for growth in many directions. This capability for growth and change is extremely important to us."

THE FLOWMATIC SYSTEM FOR OTHER COMPUTERS

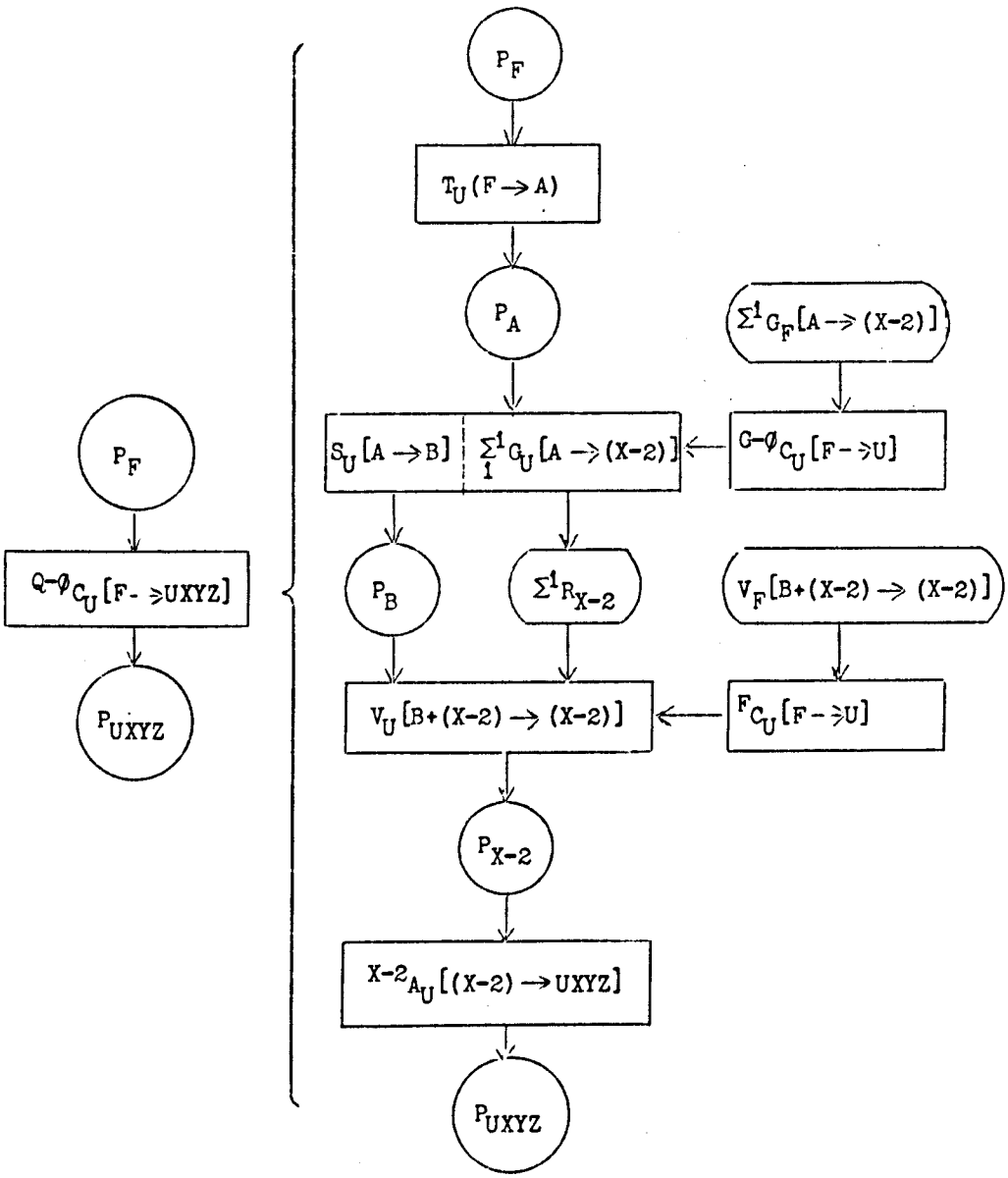
During this time, starting before FlowMatic was completed, a compiler had been independently developed for computer UXYZ. The first step was to create an assembly routine, $^{X-2}A_U[(X-2) \rightarrow UXYZ]$ by means of ^{X-1}A:



This X-2 Assembly system was then used to produce the translator, selector, and a minimizer-allocator (M) composing a new compiler, D-ϕ, to run on computer UXYZ.

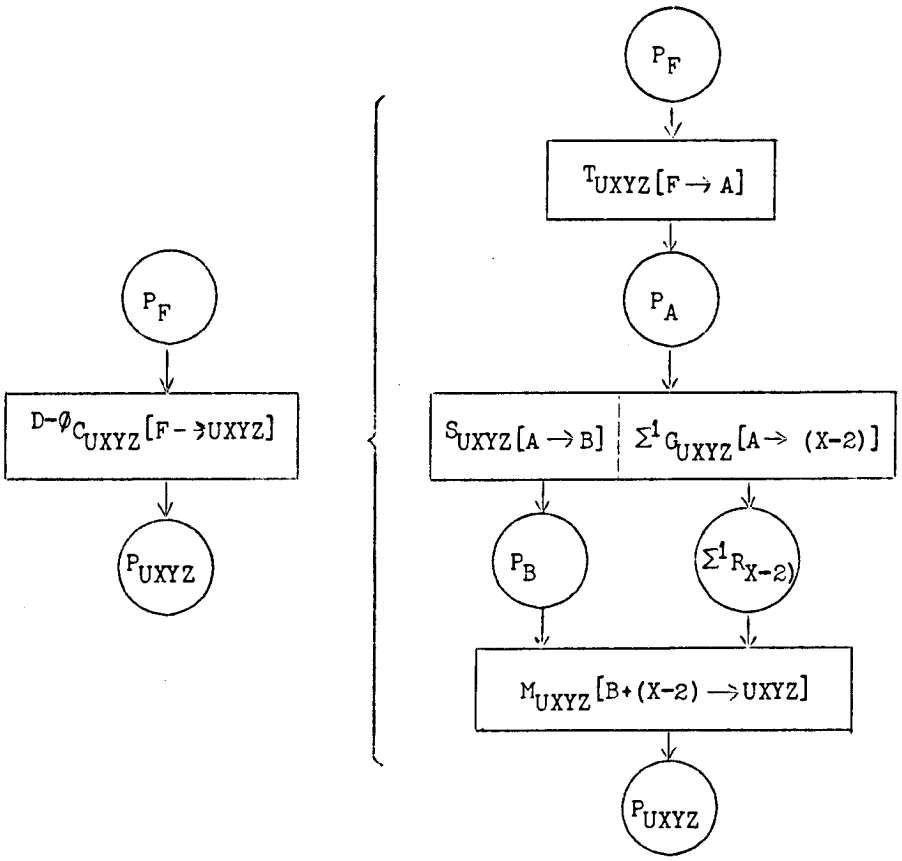


By the time the library of generators was required, another compiler Q-ϕ was developed which accepts FlowMatic pseudo-code and produces UXYZ code. Its production can be diagrammed as:



where both the FlowMatic and G-ϕ compilers were used to produce the lacking components, a library of generators and a converter. In turn, this

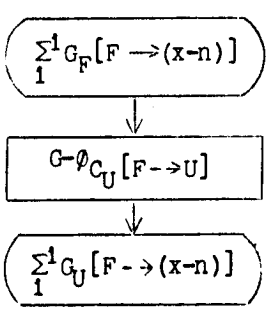
compiler could then be used to produce a set of generators and a minimizer to operate on UXYZ and D-$\emptyset$ was complete:



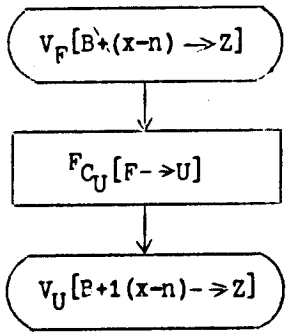
A degree of sophistication has now been achieved and it is possible to outline a technique for producing a FlowMatic compiling system for any computer Z.

(1) Define an X-series relative coding system X-n.

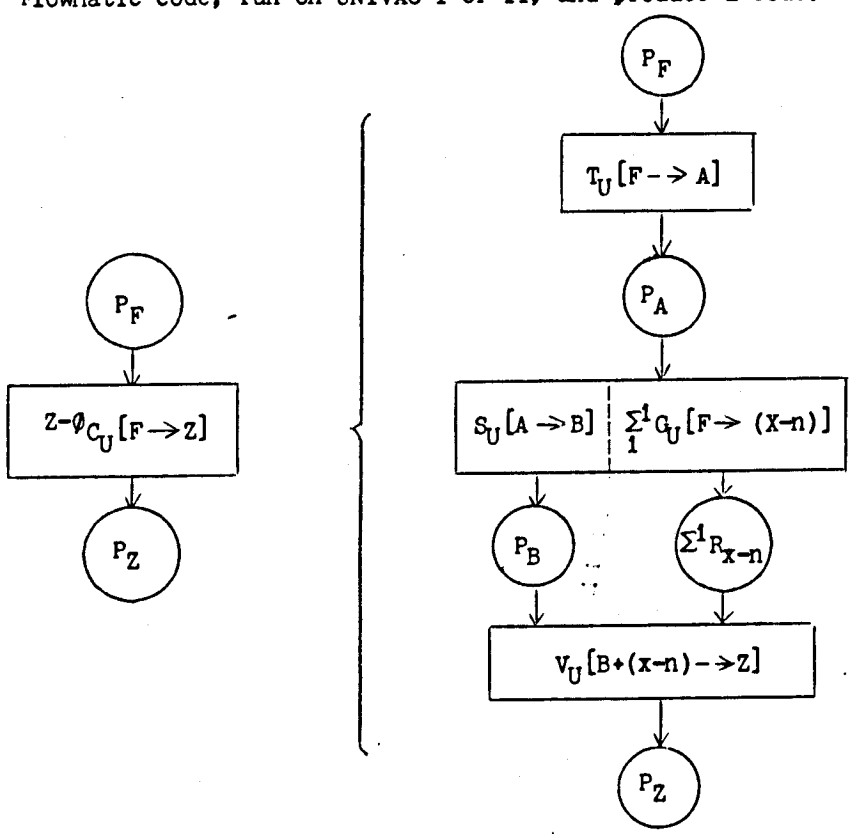
(2) Using G-$\emptyset$, produce a library of generators to run on UNIVAC I or II and produce X-n coded subroutines:



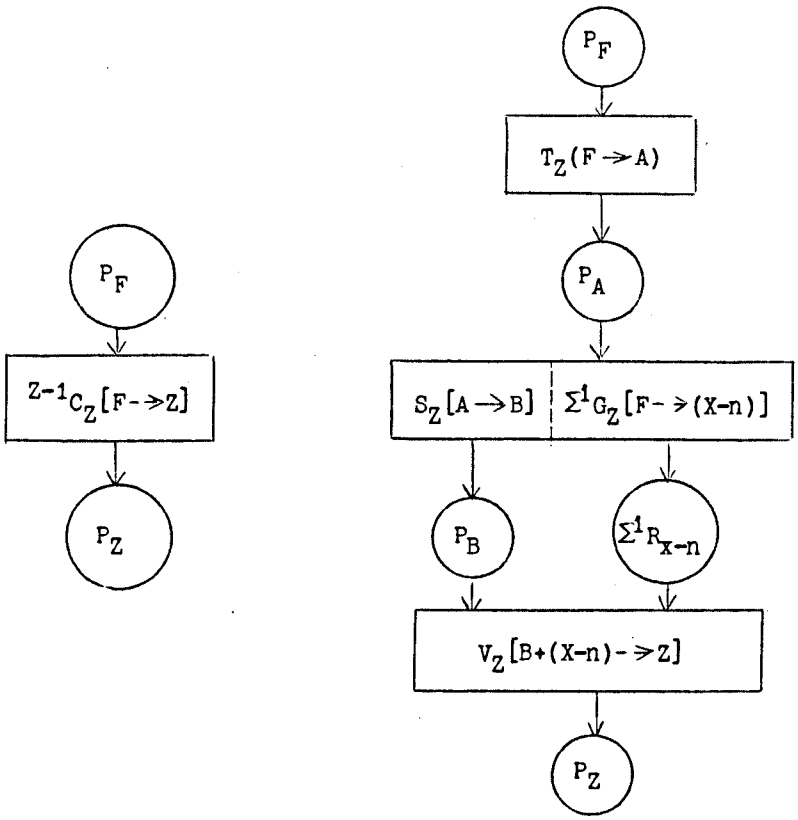
(3) Using the FlowMatic compiler produce a converter (includes allocating and minimizing as required) to run on UNIVAC I or II and convert X-n code into Z-coding.



(4) These elements then form the Z-$\emptyset$ compiler which would accept FlowMatic code, run on UNIVAC I or II, and produce Z-code:



(5) Using the Z-$\emptyset$ compiler, T_Z, S_Z, V_Z, and $\Sigma^1 G_Z$ can be produced to form the Z-1 Compiler:



Of course, once Z-1 is available, it can be used, as were FC, G-$\emptyset$C, Q-$\emptyset$C, etc. to produce compilers for still other computers.

It must be noted that this is not a recursive operation. Inefficiencies present in the original FlowMatic Compiler and generators are not transmitted, they act only as processing agents on the new efficient pseudo-codes. When G-$\emptyset$ is used to produce a new generator, the efficiency of the new generator depends on the pseudo-code defining it - not on G-$\emptyset$. Compilers may be used to successively re-code themselves to greater efficiency as the pseudo-code is altered and improved.

EXTENSIONS OF THE FLOWMATIC COMPILER

One extension of the FlowMatic Compiler consists in designing broader and more powerful generators. One such is the COMPUTE-FICA generator.

It covers about twelve elementary FlowMatic sentences in one comprehensive sentence:

(19) COMPUTE-FICA ON GROSS-PAY, DEDUCT FROM ADJUSTED-PAY AND STORE AS WEEKLY-FICA; ACCUMULATE ANNUAL-FICA, QUARTERLY FICA AND MONTHLY FICA; THEN GO TO OPERATION 27.

Not too many such compound generators can be distinguished for computer applications as a whole, but more can be defined within an installation or a group of installations in a single company or activity.

A second extension of the FlowMatic Compiler lies in the development of more flexible sentences than those presently available. Thus FlowMatic II pseudo-code will read as follows:

Sample Inventory Problem

Input inventory and transactions.
Output updated inventory, shipments, orders, and errors.
Prepare shipments, orders, and errors for printing.
(1) Compare inventory stock number and transaction stock number;
if greater go to operation 28, to process new item;
if equal go to operation 2, to apply transactions;
if less go to operation 12, no transactions to be applied.
(2) Test transaction code; if greater than 7 or less than 0, illegal code, go to operation 2a; otherwise go to operation 3.
(3) Test transaction code; if new item, illegal after match, go to operation 3a; otherwise go to operation 4.
(4) By-pass the next step after executing it once until processing of this item is completed;
(4a) Change operation 4 to go to operation 5;
operation 12 to go to operation 13;
operation 10c to go to operation 10d;
in order to compute a new unit price and an order request, if required, after transactions are applied.
(5) Test transaction code;
if change of re-order point, go to operation 5a;
if stock ordered, go to operation 5b;
if adjustment to quantity on order go to operation 5c;
otherwise go to operation 5d, in order to check for negative quantity on order.
(5a) Move transaction re-order point to inventory re-order point and go to operation 10 to read next transaction.
(5b) Add quantity ordered to quantity on order and go to operation 10 to read next transaction.
(5c) Add adjustment to quantity on order and go to operation 10 to read next transaction.
(5d) Test quantity on order; if positive, go to operation 6.
(5e) Insert 3, indicating negative quantity on order, in inventory error code.
(5f) Write inventory item in error file.
(6) Test transaction code;
if stock received, go to operation 6a;
if adjustment to stock on hand, go to operation 6f;
otherwise go to operation 6h, in order to check for negative stock on hand.

(7) Test transaction code;
if shipment made, go to operation 7a;
otherwise go to operation 7f.
(8) Test transaction code;
if delete item, go to operation 8a; otherwise go to operation 2a.
(8a) Insert D in transaction error code and in inventory error code to indicate item deleted.
(8b) Execute operation 5f. Write inventory item in error file.
(9) Write transaction item in error file.
(10) Read next transaction item; if end of data go to operation 10a;
otherwise go to operation 11.
(11) Jump to operation 1 until inventory file terminates.
- - - - etc.

Generators were prepared, using Q-$\emptyset$, to produce subroutines for a variety of non-Univac-family computers in order to prove the independence of the system.

Finally, in order to show that the system was actually independent of the FlowMatic English pseudo-code, the compiler was twice modified, once to accept and reply in German and secondly to accept and reply in French. Acceptable are such sentences as:

(1) COMPARE PRODUCT-NO (A) WITH PRODUCT-NO (B); IF GREATER GO TO OPERATION 10; IF EQUAL GO TO OPERATION 5; OTHERWISE GO TO OPERATION 2.
(1) COMPAREZ PRODUIT-NO (A) AVEC PRODUIT-NO (B); SI SUPERIEUR ALLEZ EN OPERATION 10; SI EGAL ALLEZ EN OPERATION 5; AUTREMENTS ALLEZ EN OPERATION 2.
(1) VERGLEICHE PRODUKT-NO (A) MIT PRODUKT-NO (B); WENN GROESSER GEHE ZU OPERATION 10; WENN GLEICH GEHE ZU OPERATION 5; ANDONSTEN GEHE ZU OPERATION 2.

(8) READ-ITEM A; IF END OF DATA GO TO OPERATION 14.
(8) LISEZ-PAQUET A; SI FIN DE DONNEES ALLEZ EN OPERATION 14.
(8) LESE-POSTEN A; WENN ENDE GEHE ZU OPERATION 14.

(11) JUMP TO OPERATION 1.
(11) SAUTEZ EN OPERATION 1.
(11) SPRUNG ZU OPERATION 1.

Obviously, the compiler can be extended to any language including an invented one, and translation is possible within the limited FlowMatic vocabularies.

AND NEXT?

With FlowMatic in hand to cope with the problem of coding, the next step must be to attack the problem of automatic *programming*, when the computer can be relied upon to specify the runs of a problem in an optimum manner, design the data layouts and in general behave like the "electronic brain"

it has so often been called. Ability to code rapidly and without errors should accelerate research programs in data retrieval, language translation and the many other problems urgently awaiting computer solution.

The group that I represent - the Automatic Programming Development Group foresees many, busy and we hope effective years of hard work ahead.

Great credit must be given to three groups of people for what has been accomplished so far:

to the members of the Automatic Programming Development group for their courage, initiative, imagination, and hard work in developing these systems;

to our superiors in Remington Rand UNIVAC whose vision granted us the support and opportunity to do this work;

and finally to our UNIVAC installations who have tested out, criticized, evaluated, used, misused, and re-used, suggested, requested, and demanded and finally implemented the results of our efforts.

REFERENCES

An Incomplete Bibliography on Automatic Coding

1. WHEELER, WILKES, and GILL, Preparation of Programs for an Electronic Digital Computer, *Addison Wealey Publishing Co., Cambridge Mass.*, (1951).
2. LANING, J. H. and ZIERLER, N., A Program for Translation of Mathematical Equations for Whirlwind I, *MIT Instrumentation Laboratory, Engineering Memorandum E-364*, (Jan. 1954).
3. Manual of Operation of the Automatic Sequence Controlled Calculator, *Harvard University Press, Cambridge, Mass.*, (1956).
4. Summary of Papers Presented at the Seminar on Data Handling and Automatic Computing, Office of Naval Research, Dept. of the Navy, Washington, D.C., March 1951.
5. Proceedings of the Symposium on Automatic Programming for Digital Computers, Office of Naval Research, Dept. of the Navy, Washington, D.C., May 1954.
6. Proceedings of the Symposium on Advanced Programming Methods for Digital Computers, Office of Naval Research, Dept. of the Navy, Washington, D.C., June 1956.
7. Proceedings of the Automatic Coding Symposium, *Monograph No. 3 of the Journal of the Franklin Institute*, Philadelphia, Pa., April 1957.
8. Proceedings of the Computer Applications Symposium, Armour Research Foundation of Illinois Institute of Technology, Oct. 1957:
   (a) HOPPER, G. M. - Automatic Programming for Business Applications.
   (b) BEMER, R. W. - The Status of Automatic Programming for Scientific Problems.

APPENDIX

Preceeding superscript indicates name of the routine:

$^{X\text{-}1}A$ = X-1 Assembly

$^{G\text{-}\emptyset}C$ = G-$\emptyset$ Compiler

Letter indicates type of routine

A = Assembly	P = Problem
C = Compiler	R = Subroutine
G = Generator	T = Translator
M = Minimizer	S = Selector
	V = Converter

Following subscript indicates code or pseudo-code in which routine is written

P_F = problem written in FlowMatic pseudo-code

A_U = assembly routine written in UNIVAC I or II

V_{1105} = converter written in 1105 code

Square brackets include statement of transition accomplished

[(X-1) $\rightarrow$ U] = X-1 relative coding to UNIVAC code

[USE $\rightarrow$ 1105] = USE symbolic coding to 1105 code

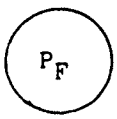
[F $\rightarrow$ U] = FlowMatic pseudo-code to UNIVAC code.

Rectangular boxes are used to represent the operating routine of a given configuration.

$\boxed{F C_U [F \rightarrow U]}$ = The FlowMatic Compiler which is written in UNIVAC I or II code and translates FlowMatic code into UNIVAC I or II code.

$\boxed{V_U [(X\text{-}4) \rightarrow MUSE]}$ = A converter which runs on UNIVAC I or II and converts X-4 code to MUSE code.

Circular boxes represent input or output data.



= Problem stated in FlowMatic pseudo-code



= The i subroutine in X-4 code that will make up a complete program



= The totality of i (task routines) written in UNIVAC code.

DISCUSSION ON THE PAPER BY DR. G. M. HOPPER

MR. J. A. GOSDEN: First of all there are two questions I should like to put to Dr. Hopper. The first is to ask her to outline in more detail the most difficult parts of her project that still have to be tackled. The work to date covers the area that can be called automatic coding, the future work which is automatic programming seems to me to be of a much higher order of complexity. The second question is to ask how far these techniques can be used on the small computers. The tasks outlined so far require several magnetic tapes, large tables in which to store nouns and verbs, and must include techniques such as pattern recognition and optimisation which require large programmes.

Many of the benefits and time savings quoted, stem from the fact that the planning as opposed to coding effort, was simplified. I think this reflects on the previous procedures. Even without adopting a coding system in English it is possible to do all the planning and charting in English with no restrictions at all on the vocabulary or grammar. In my own Company all planning can be done with no knowledge of the coding system and customers with no previous computer knowledge can immediately understand the flow charts. Too many flow charts say how they are doing something and not what they are doing, in fact they could be described as "coding in boxes".

MR. P. SHACKLETON: Anyone who has been concerned with producing a large commercial programme becomes almost obsessed with this problem of trying to make the task easier, as programming is more difficult for commercial jobs than for scientific problems. Your specification is vague and I think always will be: it continually changes as business is an organic activity: you are under pressure, and the board of directors often want results yesterday.

We are grateful to people like Dr. Hopper and her team who spent years in devising means to reduce the clerical side of programming. The most valuable contributions are the full use of generating techniques and "symbolic addressing"; this means that the programmer can refer to each element of data by name, independently of structure or representation in the machine. But I would like to challenge the assertion that this scheme really touches the problem of general data processing in business. I suggest three reasons why it fails to do so.

Firstly, it is still based on a particular machine or, at best, class of machines. There is a single level working store and magnetic tapes used for input and output as well as storage. The machine does - or at most, initiates - one thing at a time and is centralised onto an accumulator or other logico-arithmetic registers.

Secondly, the activity which it tackles is very simple file processing which, although of great importance in "bread and butter" jobs, is only a small part of business data processing. There is a pre-occupation with the large file of which only a small part is active. There is a tendency to streamline the system and avoid exceptions, which may be more feasible in the American than the British climate of opinion. A file seems to be regarded as merely a number of items as like as peas in a pod and jobs are broken down as a matter of course into repeated runs of the whole standing data. Really, for something like production control, a file would be very complex: some operations might refer to whole groups or sub-groups within a file, and each item might itself contain sub-files, lists, arrays, etc. This autocode pays little attention to logical relations, although many business decisions have a more complicated logical expression than in say, numerical analysis. Finally, nothing has been done in the deep and central problem of matching pieces of information from several long coded messages and controlling the repetition of this "giving access" in many different ways.

The third point I wish to make is an objection to introducing this pidgin English which pretends to be a common symbology but in fact may retard the effort to form a real one. I give an example of expressing a quite commonplace decision:-

in English

"If the difference between actual and standard cost is less than 10% of standard cost, do XYZ; if not, do ABC".

in Floumatic

(1) COMPARE ACTUAL-COST (A) WITH STANDARD-COST (B); IF GREATER GO TO OPERATION 4; OTHERWISE GO TO OPERATION 2.

(2) SUBTRACT ACTUAL-COST (A) FROM STANDARD-COST (B); STORE THE REMAINDER IN VARIANCE (W).

(3) JUMP TO OPERATION 5.

(4) SUBTRACT STANDARD-COST (B) FROM ACTUAL-COST (A): STORE THE REMAINDER IN VARIANCE (W).

(5) MULTIPLY STANDARD-COST (B) BY 0.1 GIVING VARIANCE-LIMIT (W).

(6) COMPARE VARIANCE (W) WITH VARIANCE-LIMIT (W); IF GREATER GO TO OPERATION 8; OTHERWISE GO TO OPERATION 7.

(7) --- XYZ ---

(8) --- ABC ---

Specified in data designs:- STANDARD-COST (File A), ACTUAL-COST (File B), VARIANCE and VARIANCE-LIMIT (W-Storage).

in Shacktonic (imaginary)

(1) $|C_A - C_S| < C_S/10$? (3)
(2) --- XYZ ---
(3) --- ABC ---

Operation (1) is to be read as a question similar to the original statement in English; if the answer is "yes", carry down the page to operation (2), if "no", go to operation (3) numbered to the right of the question mark.

I submit that the third expression is concise and clear to anyone willing to learn a few symbols and conventions. The second expression is stilted and computer-like (as present day computers follow only one path). I do not believe that volumes of it would be readily understood by non-technical managers. In spite of the remarks about how to spell "compare", each of the Flowmatic operations had to be written very carefully and in a rigid format, down to the last semi-colon (e.g. LESS cannot be used, but GREATER can). Although it has been claimed to have a multi-address code, you cannot even form (A + B - C) in one operation. A business autocode ought really to include the whole of an algebraic autocode as one part of its idiom.

The basic problem is to bridge a chasm in language between the man of affairs who talks a sort of English with a lot of jargon and idiosyncracy and a precise symbology such as logic; and even "exact" mathematics is often very sloppy from a logician's point of view. Of course logical expression may be more long-winded than English, e.g. the notion of sorting a set of data, but more often it is the other way around. For example, many factories have their own curious terminology for the way in which piece parts are ordered - according to forecast demand or actual requirements, in batches or whenever stock plus orders minus requirements falls below a certain level etc. etc. All these methods are most clearly expressed in logical formulae and apparent special cases may disappear, e.g. ordering to exact requirement is the same as ordering batches of minimum size 'one'. This is difficult to explain to a practical man but there is no doubt he must learn, if the computer is to be properly used. It is running away from that problem to propose a half-baked English which avoids difficulties supposed to have been solved.

After all civilisation itself is based on the use of symbols to an increasing degree of abstraction. We cannot go on assuming that the managers of its business and government are incompetent to use a precise symbolism. On the other hand those who are familiar with the accurate use of symbols must be less secretive and adapt them more to practical affairs.

MR. N. M. BLACHMAN: I would appreciate it if Dr. Hopper could explain more fully how a compiler can improve itself.

DR. G. M. HOPPER:* I used to be a mathematics professor. At that time I found there were a certain number of students who could not learn mathematics. I then was charged with the job of making it easy for businessmen to use our computers. I found it was not a question of whether they could learn mathematics or not, but whether they would. If I said so much as write II and then said I + I will get you the next record they said "throw those symbols out - I do not know what they mean, I have not time to learn symbols". I suggest a reply to those who would like data processing people to use mathematical symbols that they make their first attempt to teach those symbols to vice-presidents or a colonel or admiral. I assure you that I tried it.

An automatic programme which is now in process of being made has got rid of abbreviations. We have got rid of abbreviations and some of the staccato phonetic effect of the first FlowMatic which I admit was almost as bad as compiling. It is not really a language, it is actually another coding system. It is using English words to make a coding system. Anything other than this, I think, will for some years to come, be unacceptable to the business and commercial data processing. One of the things we found was they did not give us the full picture of what they wanted to do. We walked in to do a payroll once. We were given a book and told this is the way we do our payroll, everything is written down here. Of course they were not operating with the book. What is more, the book did not cover all the cases and it was tremendously difficult to make them write down all awkward cases. I dreamed up an awkward case and said "The book does not tell us how to pay him. What do you do?" The answer was "We ask Bill". I had to point out that the computer could not call Bill at 2 o'clock at night and ask what is he going to do.

Actually here it is not nearly so much the coding that is of importance as to get those people who want to use the computer at least to tell us what they are doing now. While we often wish to revise the system when we put it on a computer, we at least have to find out what is being done and I am afraid we cannot communicate with those people with mathematical symbols all we know about their compiler. In the United States there is a movement saying more people ought to be studying mathematics. I am in whole-hearted agreement but if I put that before anyone of the people in charge of our future installation I think they would throw me out very rapidly.

It is a limited language and it is growing, that is the only thing I can say in defence of it. We are making it smoother. We now can use it without all that jumpiness and the lines have gone from it and it is easier to read.

---

* An edited version was not available.

The round off is taken care of in the subroutines. For one of the computers you may ask to operate such as multiplication with or without round off and with or without checking. The type of round off is specified with the verb, carried with the verb and divided by the generator.

As well as the big computer there is a small one which punches cards on a 5,000 word drum as good as a fast reader and a slow reader and it does exactly the same compiling system as those Univac I or 1105. It is the same language and same statement of problem for all three computers from the smallest to the very largest. These are all running at the present time. So the knowledge is not limited by the size of the computer.

As far as using other computers is concerned Air Material Command asked if we could make a programme for T.R.C. They questioned the cost. We ran full tilt against the concept that in future instead of its own programme and its own ear, so long as anyone can state what has to be done, a computer can be made to do it and it does not matter what computer you use to do it with; any computer can be used to programme any other computer. This is a non-hardware concept, there is no relationship. The computer has no consciousness and has no conscious thought of its own instructions, no storage or anything.

As to what next. What we have done so far should only be called automatic coding because that is all it is. We have hardly begun to tackle the job of automatic programming. This is the future - the day when we can use the computer to analyse our problems and tell us who to do it for us, when we can get the computer to optimize the data handling, when the computer can separate the entire programme from the origination of the document, hand-written document of whatever kind, down to what is the best format to give the man a report.

There is a long way to go in the design of the system. The bulk of the time should be spent on system design, not auto coding. Coding we have shown at least in part to be a clerical job and to be able to be turned over to the computer. This should free our department to do system design. With a good system design you can make a poor job look extremely good while an excellent computer may do a perfectly lousy job.

We should turn our attention not to the tiny loops of technical coding but to design. A good system with a good flow of data, with a good motion of data and proper things done at the proper time and I hope this is the field we will be going towards and will be able to use the computer to optimize some of this work. I think we have taken only the first step in trying to design good systems and to make good use of our computers.

This was very much called to my attention because we are still discovering brand new things that we can do with Univac I and remember this design was frozen in 1948 and we are still discovering abilities in that computer that we have not previously made use of so we can spend time in trying to learn to use the computer rather than wasting time doing

de-coding work. I think everybody should go through it with at least three years of coding so they can appreciate having a computer do it.

Page 179, please. When I wrote the paper I have to say that I took a short cut. I used a pair of scissors and cut pieces out of Col. Miller's report and inserted them and the glue did not get on all of them. The paragraph on page 179 belongs to Col. Miller and not to me though I think I can justify what he said, if necessary. The communication system as viewed in the Air Force was largely a question of giving to the people designing the system a vocabulary and asking them to stay within that vocabulary. It made it far, far easier to do a system design and to interchange and communicate concerning the system when people agreed to stay within a certain level of vocabulary. Also it meant that they standardized their problems so when they referred to a certain quantity they were all referring to the same quantity.

Finally, Mr. Blachman's remarks, I think, are an attack upon this whole area and I still must insist this is only a first step and we have much yet to do. We felt that we could not tie ourselves to a recursive system because we would never have enough programmers to re-write these systems all the time. We must be able to have some form of aim and we must be able to use the computer to advance that aim. It is for that reason we want a non-recursive system in which each time each generator, each compiler, has written angle in English so we could use the computer and yet not be forced to fall back on the recursive system which would steadily make our routine useless and puerile. You can use an English definition of generation and feed it through a feeble compiler and get out a better generator because it rests upon the generator and not upon the controlling system at all. That is nothing but a service instruction and has no concern with the actual programme put out.

I think it would be lovely if we could use symbols all the time and everybody would be able to understand them.

SESSION 2

PAPER 2

SOME TECHNICAL FEATURES OF THE MANCHESTER MERCURY AUTOCODE PROGRAMME

by

R. A. BROOKER

BIOGRAPHICAL NOTE

Mr. R. A. Brooker graduated at the Royal College of Science, London, then spent two years designing and partly building a relay computer; this was followed by two years at the University Mathematical Laboratory, Cambridge, working first on the Differential Analyser, and later on the EDSAC I.

Since 1951 he has been at the Computing Machine Laboratory, Manchester University, where he is now Senior Lecturer in the Theory of Computing. He has worked mainly in the field of numerical analysis, programming, and logical design of machines.

SOME TECHNICAL FEATURES OF THE MANCHESTER MERCURY AUTOCODE PROGRAMME

by

R. A. BROOKER

INTRODUCTION

THE Manchester Mercury Autocode is a large scale conversion programme for translating "programmes" written in non-machine language and converting them to the precise instructions which are eventually obeyed by the machine. The subject language of Autocode was chosen to facilitate the numerical analysis of scientific and engineering problems, and has been described elsewhere (Brooker, 1958, *ref. 1*). The purpose of the present paper is to give some idea of how the conversion programme works behind the scenes. For this purpose we assume that the reader is familiar at least with the general ideas regarding conversion and intepretive routines and with the logical structure of two level storage machines of the magnetic drum type. The reader who is also familiar with the aforementioned account and with Mercury programming techniques, will find this paper easier to follow. Nevertheless it is not intended to give a detailed technical account of the programme - since this extends to approximately 3000 instructions it would obviously be out of the question - but mainly to indicate the organisation of the storage facilities, which in the author's opinion has been mainly responsible for its success in making Mercury easy to use. For the benefit of the general reader we shall outline the more important logical features of the Mercury computer before passing on to the main subject matter.

THE MERCURY COMPUTER

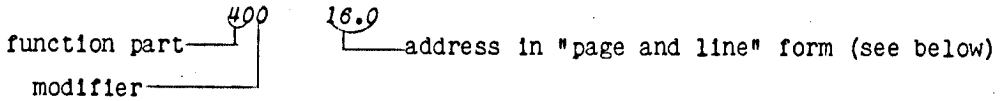
Mercury is a two level storage machine employing ferromagnetic cores for the working (fast) store and a magnetic drum for the auxiliary store. The working store extends to 2048 registers each of 20 bits, and the magnetic drum consists of 512 *sectors* each of 64 registers. The arithmetical unit is designed to operate in number expressed in floating binary form (40 bits). The instructions (20 bits) are of the single address variety and contain 3 "B" digits referring to one of eight modifier registers (10 bits). The input-

output medium is 5-hole paper tape. Altogether the following items of information are handled.

1. *Floating point numbers, $a.2^p$.* Here a ($1 > a \geqslant 1/2$, $-1/2 > a \geqslant -1$) has 30 bits, and the exponent p ($255 \geqslant p \geqslant -256$) has 10 bits (one being redundant). Access to numbers is restricted to the 1024 *double* registers

$$0, 2, 4, \ldots\ldots, 2046 \qquad (1)$$

2. *Instructions.* The 20 bits of an instruction are made up as follows: function digits (7); modifier digits (3); address digits (10). They are represented on programme sheets in decimal code, e.g.,



means "transfer contents of double register *1024* (*16* x *64* + *0*) to the fl.pt. acc." Access to instructions is restricted to the *1024* registers

$$0, 1, 2, \ldots\ldots, 1023 \qquad (2)$$

3. *10-bit words.* These are equivalent to the address half of an instruction, but may be used to represent miscellaneous pieces of information. Access to 10-bit words is restricted to the 2048 *half* registers denoted by

$$0, 0^+, 1, 1^+, \ldots\ldots, 511, 511^+ \qquad (3a)$$

$$512, 512^+, 513, 513^+, \ldots\ldots, 1023, 1023^+ \qquad (3b)$$

The interpretation of the address digits of an instruction depends on the nature of the instruction. If the operand is a 40-bit number, then the 10 address digits automatically refer to the 1024 *double* registers (*1*); while if the instruction is a transfer of control, then the address part refers to the 1024 (*medium*) registers (*2*). Finally if the operand is a 10-bit word (e.g. in setting a modifier register) the address part refers to one of the 1024 *half* registers (*3a*) or (*3b*), a duplicate set of instructions being provided to deal with the second set of 1024 half registers. In this way (i.e., by effectively including one of the address digits in the function) access to 2048 half registers is attained. Individual access to the (medium) registers *1024, 1025,* ....., *2047* is not possible however.

The Magnetic drum

Information is transferred to and from the fast store in blocks of 64 medium words, the contents of a *sector*. For this purpose the fast store is divided into 32 *pages*, corresponding to registers $64n$ to $64n + 63$ inclusive, $n = 0\ (1)\ 31$, and it is possible to transfer the contents of any *sector* to any *page*, and vice versa. This requires two instructions, the first

instruction

670 m

means "select *sector* number m ($0 \leq m \leq 511$)", while a second instruction

680 n (read to *page n*) or *690 n* (write from *page n*)

effects the transfer proper. The time for a magnetic transfer is approximately 17½ ms. If a transfer involving an even (odd) numbered *sector* is followed immediately by a transfer involving an odd (even) *sector*, then the time for the second transfer is only 8½ ms.

Times of operation (in beats of 60 μs)

Transfers of 40-bit numbers to and from the fl.pt. acc.	2 beats
Addition to or subtraction from the acc.	3 beats
Multiplication by the contents of the acc.	5 beats
All other instructions	1 beat

Input: 200 5-hole characters per sec.
Output: 33 5-hole characters per sec.

THE AUTOCODE LANGUAGE

The Autocode language employs the following symbols

$a\ b\ c\ d\ e\ f\ g\ h\ u\ v\ w\ x\ y\ z\ \pi$

$i\ j\ k\ l\ m\ n\ o\ p\ q\ r\ s\ t$

$.\ 0\ 1\ 2\ 3\ 4\ 5\ 6\ 7\ 8\ 9$

$+\ -\ \neq\ =\ >\ \geqslant\ *\ (\ ,\ )\ \longrightarrow\ /\ \emptyset\ '$

which are represented on the input medium by 5-hole teleprint characters following a "LETTERS" or "FIGURES" switch. For ease in checking only those teleprinter combinations which actually mark the paper are regarded as significant so that SPACE and LINE FEED are ignored, as also are LETTER SHIFT on "LETTERS" and FIGURE SHIFT on "FIGURES". A CARRIAGE RETURN symbol however is used to denote the end of an instruction.

The Autocode "instructions" are of two kinds, the *arithmetical* instructions which perform the calculation proper, e.g.,

$$y = \cos\ (2\ \pi\ x\ /\ a)$$

and the *control* instructions which "organise" the calculation, e.g.,

$$i = 1(1)\ 15$$

or arrange to print results, e.g.,

$$\text{print}\ (x_i,\ 1,\ 10)$$

It has not been possible to conceal the two level nature of the storage facilities of Mercury and for this reason an Autocoded programme is partitioned into *chapters*, a *chapter* being the largest bit of the programme which can operate in the fast store. The *chapters* are normally stored on the magnetic drum and as each *chapter* is called in it is transferred to the fast store, (*pages 2-14* inclusive). The general idea is that each *chapter* will keep the machine sufficiently occupied with computation so that the time spent in changing from one *chapter* to another (approximately 120 ms) is not significantly large.

Just as a *chapter* has to be transferred to the fast store before it can be used, the immediate data requirements of the calculation are kept in the second half of the fast store (*pages 16-31*). These numbers are referred to as the *working variables*. Numerical information in excess of immediate requirements has to be kept on the drum, on that part of it not taken up by the *chapters* and the permanent material. These extra numbers are referred to as the *auxiliary variables*. The (508) working variables will be sufficient for many programmes, and for larger programmes the transfer of numer numerical material from the drum to the working store, and vice versa, is under the control of the user. There are 508 working variables corresponding to the double registers,

16.0, 16.2,....., 31.54

They are partitioned as follows:

The Special Variables

The 14 locations

31.28, 31.30,....., 31.54

are denoted by the letters

a b c d e f g h u v w x y z

These symbolic addresses can be employed with the basic machine instructions, thus, e.g., *400 (h)* would be understood to mean "transfer the contents of (double) register *31.42* to the accumulator".

The 14 locations

31.0, 31.2,....., 31.26

are similarly denoted by the primed letters

a' b' c' d' e' f' g' u' v' w' x' y' z'

The remaining locations on *p.31* contain frequently needed constants, thus

31.56 contains π (i.e., *3.14159*.....)

31.58 contains 0.2^{-256} (absolute zero)

31.60 contains -1.2^{0} (minus one)

31.62 contains $(-\frac{3}{4})2^{29}$ (used in converting 40-bit fl.pt. integers to 10-bit form)

The 480 locations on *pp.16-30* inclusive are named by the programmer, by partitioning them into groups associated with the above (unprimed) letters. This is done by writing suitable *directives* at the head of the chapter, thus e.g.,

chapter 1

a ⟶ *9*

b ⟶ *14*

x ⟶ *44*

associates the storage locations

16.0, 16.2, ..., *16.18* with a_0, a_1, ..., a_9,

16.20, 16.22, ..., *16.48* with b_0, b_1, ..., b_{14},

16.50, 16.52, ..., *16.138* with x_0, x_1, ..., x_{44}.

Although the above directives envisage that the *a*'s extend only up to a_9, and that the *b*'s extend only up to b_{14}, nevertheless b_0 is equivalent to a_{10}, b_1 to a_{11}, and x_0 is equivalent to b_{15}, and so on. Up to 14 directives may be set corresponding to each of the variable letters.

A machine instruction following one of these directives, e.g., *400* (b_3), would mean "transfer contents of (double) register *16.26* to the floating point accumulator". Details of the allocation of the working storage locations are recorded with each *chapter* in the magnetic store.

Indices

There are 12 *index* or pseudo-modifier (half) registers associated with the letters

i j k l m n o p q r s t

Since access to 10-bit words is restricted to the first half of the store, these quantities are recorded in *p.0*, in the (half) registers

58, 58+, 59, 59+....., *63, 63+*

As an example of such a symbolic address employed with a true machine instruction, the instruction *006 (m)* means "transfer contents of (half) register *60* to *B*-register number *6*". The indices take integral values in the range $-512 \leq n \leq 511$. They are primarily intended to be combined with variables in the form, e.g.,

$$a_i \text{ or } b_{(n-1)} \text{ or } x_{(s+50)}$$

to represent a free suffix. Thus for example, if $n = 4$, then $b_{(n-1)}$ would refer to b_3. The last two expressions illustrate the most general form which a suffix may take, namely (index ± integer). Employed with an actual machine instruction, e.g., *400* $(x_{(s+50)})$, this may be regarded as equivalent to the pair of machine instructions $\begin{matrix} 006\ (s) \\ 406\ (x_{50}) \end{matrix}$

The Auxiliary Variables

These are kept on the magnetic drum on the space not occupied by the *chapters*. As explained in the sequel *(layout of material on magnetic drum)* the address system for the 40-bit locations on the drum is regarded as starting at *sector 64: sector 64* contains locations *0-31*, *sector 65* contains locations *32-63*, and so on. Pseudo "transfer" instructions are provided which allow the user to transfer a block of consecutive variables from the auxiliary (drum) store to the working store, or vice versa, starting at an arbitrary location in the auxiliary store and an arbitrary location in the working store, thus, e.g., ϕ_2 *(4000,* x_1*, 50)* copies the 50 numbers,

$$x_1, x_2, \ldots\ldots, x_{50},$$

to the auxiliary locations

$$4000, 4001, \ldots\ldots, 4049.$$

The corresponding "read" instruction is ϕ_1 *(4000,* x_1*, 50)*. Since the basic machine instructions for effecting magnetic transfers provide only for *sector/page* transfers, it is necessary to use a buffer store to arrange the generalised transfer. Thus in a "reading" transfer the *sectors* involved are transferred one at a time to *p.15*, meanwhile the relevant material is transferred from there, by copying, to its final destination in the working store. In the case of a "writing" transfer, the first and last *sectors* of the group of *sectors* involved are first copied to *p.15*, so that the "flanking" material on these *sectors* is preserved when the contents of *p.15* are ultimately transferred to the magnetic drum. The routine for carrying out both kinds of transfer operates in *p.1* during the execution of the programme.

THE ARITHMETIC INSTRUCTIONS

These will be illustrated by means of examples giving the equivalent machine instructions (see appendix) employing the symbolic addresses already introduced. (Here *A* denotes the fl.pt. acc., and *D* denotes a double length register).

AUTOCODE	*MERCURY CODE*	
$a = 0$	*400, 14.62* *410 (a)*	simple transfer via *A*
	D 14.62 contains 0.2^{-256}	
$x_k = x_k + 1$	*006 (k)* *406 (x_0)*	transfer x_k to *A*
	420, 14.60	add 1
	006 (k) *416 (x_0)*	plant result
	D 14.60 contains $(0.5)2^1$	
$x_n = x_0 + nh$	*400 (x_0)* *410, 38*	dump x_0 in *D 38*
	200 (n) *210, 57+* *400, 56* *440, 31.58*	convert index *(n)* to 40-bit fl.pt. number in *A*
	500 (h)	multiply by *h*
	420, 38	add x_0
	006 (n) *410 (x_0)*	plant result
$v = 2\pi u/n$	*400, 14.58* *500 (π)* *500 (u)* *410, 36*	form and dump $2\pi u$ in *TS 36*
	200 (n) *210, 57+* *400, 56* *440, 31.58*	convert index *(n)* into 40-bit fl.pt. number in *A*
	101 (1) *590, 6*	call in division subroutine: form *36/A* in *A*
	1) 410 (v)	plant result

The instruction *101(1)* sets the link prior to calling in the division subroutine, which for time economy reasons is permanently available in p.0. The link is the control number corresponding to the micro-label *1)* attached to the next instruction *410 (v)*, to which it is desired to return.

AUTOCODE		*MERCURY CODE*	
a = log *(xx + yy)*	*400 (x)*		form
	500 (x)		$x^2 + y^2$ in *A*
	410, 38		
	400 (y)		
	500 (y)		
	420 38		
	670, 14		transfer subroutine
	680, 15		to *p.15*
	101 (1)		and enter it
	590 15.0		
1)	*410 (a)*		plant result

This example illustrates the technique for dealing with functions involving a subroutine. The subroutine is transferred from the magnetic drum to a page of the fast store *(p.15)* when it is required. For functions of a single variable the subroutine is designed to replace the argument in *A* (on entry) by its function value (on exit). Later on arrangements can be made to accommodate the more frequently used subroutines into any space that is left over in the 13 *pages* available for the *chapter*. This has obvious time economy effects and subroutines treated in this way are known as *quickies*. In the case of a quicky the assembly programme has to search for references to that subroutine and amend them accordingly, thus

670 -		*570 -*	dummies
680, 15	becomes	*570, 15*	
101 (1)		*101 (1)*	
590 n		*590 (n - 15.0 + m)*	
1)		*1)*	

where m is the address of the first instruction of the quicky. This operation is carried out at the end of each *chapter* before it is finally transferred to the drum.

y = mod *(xx - yy)*	*400 (x)*	form and plant
	500 (x)	x^2 in *TS 38*
	410, 38	
	400 (y)	form $-y^2$ in *A*
	500 (y)	
	520, 31.60	
	490 (1)	change sign of *A* if
	520, 31.60	negative plant result
1)	*410 (y)*	

Certain functional operations do not involve a subroutine and may be made by means of a short open sequence in the programme, as in the above example.

THE CONTROL INSTRUCTIONS

The most important instructions under this heading are the *jump* instructions, which allow the machine to select alternative courses of action as in the following example:

Calculate

$$f(x) = \begin{cases} e^x, & x < 0 \\ 1 + \sin x, & x \geq 0 \end{cases}$$

```
      jump 1, x ≥ 0
      f = exp (x)
      jump 2
1)    f = sin (x)
2)    f = 1 + f
```

The first instruction is a *conditional jump*, and means "jump to the instruction labelled *1* (i.e., the fourth instruction) if the condition $x \geq 0$ is satisfied, otherwise obey the next instruction in the usual way". Any instruction can be labelled with an integer in the range 1-127 inclusive. The labelling system of each *chapter* is independent and the *control numbers* corresponding to the labelled instructions (i.e., the register address of the first of the equivalent set of machine instructions) are kept in a directory, which is recorded along with the *chapter* on the magnetic drum. As in all floating address systems the actual control numbers can only be substituted for labels after the complete *chapter* has been read, and this is one of the functions initiated by the *close* directive, with which each *chapter* terminates. The label directory for each *chapter* is also referred to during the subsequent execution of the programme when passing from one *chapter* to another. This is done by means of *across* instructions, e.g.,

across *1/2*

which means "jump to the instruction labelled *1* in *chapter 2*". This is a relatively time consuming operation (120 ms.), since it involves the replacement of 13 *pages* of programme from the drum. It is not inconvenient, therefore, to have to refer, as well, to the *sector* bearing the label directory of the new *chapter* to locate the actual control number corresponding to the label in question; and is preferable to keeping a list of cross references between *chapters*, which would mean postponing the completion of the *across* instructions until the whole programme had been translated. *Chapter* changing is done by means of a *chapter* changing sequence which is normally kept on the drum and transferred to *p.1* when required. In addition there are the instructions such as *down 1/2* and *up* which allow the programmer to call in *subchapters*, ultimately returning to the original *chapter*.

Cycles of operations

The second important class of control instructions are those which arrange to execute cycles of instructions for values of an index (say i running from p by increments of q (or $-q$) to r, where p, q, r denote any indices or whole numbers. Nesting of cycles within cycles is permitted to a depth of 8 deep. The following example illustrates an application of these instructions.

Evaluate: $f = a_0 + a_1 x + \ldots + a_{10} x^{10}$

```
                        400  (a10)   ┐ f = a10
                        410  (f)     ┘
                        300,  9      ┐ r = 9 (-1) 0
                    1)  210  (r)     ┘
                        400  (f)     ┐
                        500  (x)     │
                        006  (r)     │ f = fx + ar
                        426  (a0)    │
                        410  (f)     ┘
f = a10                 200  (r)     ┐
r = 9 (-1)0             370,  0      │ repeat
f = fx + ar             137,  1      │
  repeat                280  (1)     ┘
```

NUMERICAL INPUT-OUTPUT

Instructions are provided for input and output of numerical data with 5-hole tape.

For example,

```
                    670, 9      ┐ transfer READ routine
                    680, 15     ┘ to p.15
read (x)            101  (1)    ┐ and enter it
                    590 15.1    ┘
                1)  410 (x)       copy result from A
```

means "read the next number on the tape and set the working variable x to this value". Similarly

```
                    006  (i)     ┐
                    406  (y0)    ┘ transfer yi to A
                    105,  2      ┐
                    106,  6      ┘ set parameters
print (yi, 2, 6)    670,  7      ┐ transfer PRINT routine
                    680, 15      ┘ to p.15
                    101  (1)     ┐
                1)  590, 15.15   ┘ and enter it
```

punches (it is subsequently printed on off-line equipment) the working variable y_i in tabular form with 2, 6 digital positions allowed for integral part and fractional part respectively.

Alternatively it is possible to obtain a printed record (to a fixed number of decimal places) by placing a ? symbol by an arithmetical instruction, e.g.,

$x_k = x_k + 1$?

```
006  (k)
406  (x0)
420, 14.60      (see p.209)
006  (k)
416  x0
101  (1)        calls in PRINT routine.
590, 1.38
```

1) ———

the effect of which will be to print the value of x_k on a new line. This is primarily a "debug" print, and by throwing a switch on the console the ?'s can be ignored during conversion. Thus it is possible to proceed to a production run with the final trial tape.

A ? print involves only two extra instructions in the assembled programme. This is achieved by arranging for the instructions which actually transfer and enter the PRINT routine, namely

```
680, 7
680, 15
590, 15.0
```

to stand in *p.1*, and to include only the link setting instruction in the main programme.

LAYOUT OF MATERIAL IN THE MACHINE

At this point it is appropriate to sketch the layout of material in the working store and the auxiliary drum store.

Layout of material in fast store during execution of programme

p.0	division subroutine, miscellaneous working space, details of *chapter* levels, etc., needed by *chapter* changing sequence, indices.
p.1	auxiliary (magnetic) transfer subroutine and miscellaneous sequences (used in connection with ? print and complex sq.rt. routine).
pp.2-14	*chapter* being obeyed.

p.15 used as temporary store for function subroutines, and as buffer store in auxiliary transfers.

pp.16-30 main (indexed) variables.

p.31 special variables and miscellaneous constants.

Layout of material on magnetic drum

sectors 0,1 basic binary input-output sequences (these *sectors* are isolated and have no connection with the Autocode programme other than to read it in to the machine).

sectors 2-15 contain subroutines relevant to the execution of the translated programme, namely,

2 tangent routine
3 page *0* material (see previous table)
4 part of read routine (mainly negative powers of 10)
5 *across* and *down chapter* changing sequences
6 *up chapter* changing sequence
7 print routine
8
9 read routine
10 arctangent routine
11 auxiliary transfer sequences and miscellaneous material
12 sine-cosine routine
13 exponential routine
14 logarithm routine
15 square root routine

sectors 16-31 These are used to record the contents of *pp.16-31* (the working variables) should they need to be preserved when calling in a *subchapter* which makes independent use of these *pages*. Thus *sectors 16-31* are used in this way by the master *chapter* when calling in a *subchapter*, and *32-47* by a *subchapter* when calling in a *sub-subchapter*.
32-47

sectors 48-63 *Chapter 0*,
64-79 *Chapter 1*, and so on.

sectors 432-479 Autocode conversion programme.

sectors 480-511 These are normally occupied by the engineering test routines unless they are required for the Matrix Algebra Programme.

It will thus be seen that, although a *chapter* occupies 13 *pages* in the fast store, 16 *sectors* are allocated to each *chapter* on the magnetic drum. This is because some extra information about each *chapter* is retained on the drum. More precisely, if the 16 *sectors* were transferred to *pp.0-15*, then the extra material would be found as follows:

$p.0$ wasted.
$p.1$ variable directives.
$p.15$ label directory.

The Label Directory

The 127 (half) registers

15.0+, 15.1, 15.1+, ... contain the control numbers corresponding to labels *1, 2, 3,* etc. The label zero is not used.

Each control number is, of course, that of the first of the equivalent set of machine instructions. The label directory will be referred to during the execution of the programme, as well as during conversion.

The Variable Directives

These are not referred to during the execution stage but only during input. This sector contains details of the storage locations associated with the letters, *a, b, c,* .....*z*. A 40-bit code word associated with each letter contains:

storage location corresponding to a_0 } specified by the programmer
" " " " a_n }
" " " " a } fixed for all programmes
" " " " a' }

The code words are recorded along with each *chapter* after it has been assembled, in case of a reference back to them later on; e.g., the variable directives employed in *chapter 1* may also be used in *chapter 2* and in this case the programmer writes:

chapter 2
variables *1*
———
———
———
close

The Auxiliary Variables

For a programme extending up to and including *chapter N*, the first available sector for the auxiliary variables is *64 + 16N*. Now the numbering of the auxiliary locations is regarded as starting at the 1st location on *sector 64*, so that the first available location is *512N*, and extends up to *11,775* (last location on *sector 431*). The storage can be extended from *11,775* to *13,311* (another *1536* variables) by overwriting the conversion routine if this is no longer needed; and still further space can be obtained by overwriting the engineering test routines - if the Matrix Algebra programme is not used.

THE CONVERSION PROCESS

So far we have only described the nature of the converted programme, and nothing has been said about the conversion process itself. This has two phases: (1) the recognition of the Autocode formula, and (2) the assembly of the equivalent machine instructions. With the exception of the cycling instructions, the Autocode instructions are translated individually without regard to what has gone before or comes after. Only the labels and the quicky references need to be filled in later on, and this is the function of the close directive (see below: the *main assembly process).*

It is not possible to give a complete account of the conversion process without assuming a detailed knowledge of Mercury programming techniques, but it is worth noting here the main features, in case they should find application on other machines. The main steps are as follows.

I. The individual characters of the instruction are read from the tape (taking proper account of LETTER SHIFT and FIGURE SHIFT combinations) and listed as "6-bit" symbols in consecutive half registers of the store.

II. If the instruction is labelled - recognised when the first item is a *number*, and the next item is a *right bracket* - the address of the next available register in the object programme is entered in the label directory.

III. The symbol list is scanned and condensed into an *item* list, the items being, *quantities* (i.e., variables, indices, or numerical constants) *words*, and *separator symbols.*

Thus, e.g., the instruction print $(x_i, 2, 8)$
which is listed initially as

$$p|r|i|n|t|(|x|i|,|2|,|8|)$$

is "itemised" in the form

$$\text{print}|(|x_i|,|2|,|8|)$$

Each of these items is represented by a 40-bit code word which contains some information pertaining to the item in question and also a recognition tag. For example the code word for x_i specifies the (double) register x_0 and the (half) register i, and carries as a recognition tag the integer 3. The code word for a purely numerical item is the number itself expressed in standardised floating binary form. In this case there is no need for a recognition tag as such a quantity can be recognised otherwise. The *item* list provides for up to 55 items, which number is only likely to be approached in the case of an algebraic formula, e.g.,

$$y = 2\ mna_{(m+1)} + a_m\ n + ma_n + 0.01\ m + 0.01\ n$$

which has 18 items.

IV. The recognition process proper. The general idea is that from an examination of the recognition tags the programme selects the appropriate assembly routine. As it turns out quite a lot can be learned from an examination of the first three items.

V. The micro-assembly process. A large class of instructions can be dealt with by selecting a 'skeleton' sequence of instructions from a 'cupboard' and filling in certain details from the item list - the *micro*-assembly process. The resulting instructions are then copied into the main programme - the *main* assembly process. In the case of the *print* instruction above, the code word for *print* gives the location of the skeleton sequence within the cupboard. The skeleton sequence itself and the final sequence are as follows:

skeleton sequence	final sequence	
	006 (i)	set A to x_i
7 400 [2]	*406 (x_0)*	
5 005 [4]	*005 (2)*	set $B5$ to *2*
5 006 [6]	*006 (8)*	set $B6$ to *8*
1 670, 7	*670, 7*	transfer PRINT
1 690, 15	*690, 15*	routine.
4 101, (1)	*101, (1)*	set link and
1 590, 15.15	*590, 15.15*	enter it.
1) ———	*1)* ———	

In the skeleton sequence the addresses in square parentheses refer to entries in the item list,

0	1	2	3	4	5	6	7
print	(	x_i	,	2	,	8	)

and the *tags* on the left of the instructions describe how each presumptive instruction is modified. Thus

tag 1 : the instruction is unaltered.

tag 4 : the address part refers to a microlabel.

tag 5 : the item referred to in this case is a numerical integer. However it could also be an index letter. The instruction is modified so as to transfer the integer, or the contents of the index register, to the B line in question.

tag 7 : the item referred to in the present example is x_i, but it might equally have been (say *0.125*, *y*, or a_3. In all these cases the presumptive construction *400* - would be modified as follows:

item	*0.125*	*y*	a_3	x_i
final instruction	*400 S* *	*400 (y)*	*400 (a_3)*	*006 (i)* *406 (x_0)*

* The 40-bit constant *0.125* is stored in *S*, the next available location at the end of the *chapter* (see below: the main assembly process).

In the case of arithmetical instructions the algebraic formula on the right hand side has to be treated rather differently, since there is virtually no limit to the complexity of form. It is not proposed to discuss this problem here except to mention that the skeleton technique can still be applied in a piecewise manner. For instance when treating a functional operation

$$y = \phi\ (x)$$

the instructions for calling in the function subroutine, namely

670, --
680, 15
101, (1)
590, --
1) ————

are conveniently assembled by means of a skeleton to which are added the details of sector number and entry position, taken from the code word associated with the particular function name.

THE MAIN ASSEMBLY PROCESS

The recognition and assembly routines operate in the first half of the working store *(pp. 0-14)*, while the *chapter* is assembled in the second half *(pp. 18-30)*. The equivalent machine instructions for each Autocode instruction are first built up in *p.15* (where access to half registers is possible) before being copied into the *chapter* proper. The 40-bit item list is stored in *pp. 16, 17*. The *chapter* is assembled from both ends, the instructions being listed in registers *18.0, 18.1,* ....., while any necessary 40-bit constants are stored *backwards* in double registers *30.62, 30.60*, etc. The conversion programme prints a warning should the two lists overlap. (In this case the user has to recast his programme). Each *chapter* terminates with a *close* directive, the function of which is to complete those details of the *chapter* which remain to be filled in, namely, the labels and the *quicky* references, if any, and then to transfer the *chapter* to the auxiliary store, together with the label directory and the variable directive details.

All 13 *pages* of the *chapter* are transferred regardless of the actual size of the *chapter*, even though this may occupy less than a *page*. The converted programme is eventually entered in *chapter 0*, which must always be the last *chapter* on the programme tape, and is automatically entered at the first instruction.

CONCLUSIONS

In a conventional programming manual it is necessary to present a complete logical picture of the machine if the user is to enjoy complete flexibility in his approach to it. However, many programming features are concerned with stretching the machine to the limit of its resources, and are of marginal utility where the general run of programmes are concerned. They can therefore be dispensed with in a system such as Autocode, with consequent saving of explanatory material and simplification of the rule book. This is an important aspect of recruiting since the number of potential users is very markedly affected by the size and complexity of the manual which they have to digest. Autocode has been devised primarily for the numerical analysis of scientific and engineering problems, and we have been able to concentrate on streamlining the accepted units of computation in this field, e.g., evaluation of elementary functions, cycling operations, etc. The resulting programme will not necessarily be as efficient as a hand tailored programme, but for those who wish to do so it is possible to use machine language to polish up the inner loop of a programme - the equivalent machine instructions in the foregoing examples are written in a form acceptable to Autocode.

One of the principal problems in devising a simplified coding scheme for Mercury has been to overcome the difficulties arising from the two level storage system. It has been proved possible to entirely *conceal* this feature (and still preserve efficiency in the object programme), but it has been possible to *alleviate* its consequences.

Thus as regards the instructions, all that the user is required to do is to partition his programme into reasonably large, but not too large (a few rules are given) *chapters*, and to ensure that the *chapter* divisions avoid the inner loops of the calculations. As regards the numerical information, the auxiliary store is regarded as consisting of a set of consecutively numbered storage locations regardless of *sector* divisions. The generalised transfer instruction is both easy to understand and convenient to use; and although the access time per word is considerably longer (about 2½ times) than that of the basic *sector/page* transfers, fewer words are transferred in general. There is also less wastage of storage space, particularly in the fast store, than is usually the case with transfers of fixed block size.

Finally, one of the most tedious features of Mercury, the multiple address system associated with half, medium, and double words, is avoided by the use of symbolic addresses, i.e., "variable" and "index" letters for double words and half words, and floating addresses (labels) for instructions.

ACKNOWLEDGEMENT

I am particularly indebted to Dr. B. Richards who has been of considerable assistance at various stages of this work.

REFERENCE

1. BROOKER, R. A. The Autocode Programs developed for the Manchester University Computers, *The Computer Journal*, 1958, **1**, 15.

APPENDIX

SUMMARY OF MERCURY INSTRUCTION CODE

All instructions are of the form *XYB, n* where

XY are the function digits *(00 - 99)*
B denotes one of the modifier registers *(0 - 7,* but *B0* contains zero)
n is the address part

In what follows

D(n) denotes contents of double length register *n*
M(n) " " " medium " " *n*
H(n) " " " half " " *n*

In *"B"* modified instructions the address part *n* becomes *(n + b)*, where *b* represents the number in the specified *B*-register.

A denotes the number in the floating point accumulator
SAC " " " " " short (10-digit) accumulator

1. *Instructions for doing floating point arithmetic*

These are defined by the following relations which give the new contents of the register altered.

40 $D(n + b) \longrightarrow A$

41 $A \longrightarrow D(n + b)$

42 $A + D(n + b) \longrightarrow A$ *44* (unrounded form of *42*)

43 $A - D(n + b) \longrightarrow A$ *45* (unrounded form of *43*)

50 $A.\ D(n + b) \longrightarrow A$ *52* (unrounded form of *50*)

51 $-A.D(n + b) \longrightarrow A$ *53* (unrounded form of *51*)

In all cases it is the most significant 29 digits of the result which is calculated.

2. *Instructions for using the 10-digit accumulator (SAC)*

These are of the form *2YB,n* and *3YB,n*

In the left hand form the address part is interpreted in the usual way (except that now it refers to a half register), while in the right hand form it is the address part *itself* which is used as the operand. In both cases the address part is *B*-modified.

2YB,n	*Y*	*3YB,n*
$H(n + b) \longrightarrow SAC$	0	$(n + b) \longrightarrow SAC$
$SAC \longrightarrow H(n + b)$	1	
$SAC + H(n + b) \longrightarrow SAC$	2	$SAC + (n + b) \longrightarrow SAC$
$SAC - H(n + b) \longrightarrow SAC$	3	$SAC - (n + b) \longrightarrow SAC$

3. *Instructions for altering the modifier registers (N.B. SAC ≡ B7)*

Again there are two forms *0YB,n* and *1YB,n* according as to whether n specifies an address or is used directly. However the *B* digit now specifies the *B*-register which is to be altered and *not* a modifier register.

0YB,n	*Y*	*1YB,n*
$H(n) \longrightarrow b$	0	$n \longrightarrow b$
$b \longrightarrow H(n)$	1	
$b + H(n) \longrightarrow b$	2	$b + n \longrightarrow b$
$b - H(n) \longrightarrow b$	3	$b - n \longrightarrow b$

4. *Jump instructions*

The phrase "jump to *n*" in the following description means "jump to the instruction standing in *M(n)*"

59B,n means "jump to *(n + b)*"
49B,n means "jump to *(n + b)*, if $A \geq 0$
29B,n means "jump to *(n + b)*, if SAC ≥ 0
09B,n means "jump to *(n + b)*, if the number in the *B*-register *last altered* is ≥ 0"
08B,n is similar to *09B,n* but jumps in the case $\neq 0$
28B,n means "jump to *(n + b)* if *SAC* $\neq 0$"
38B,n means "jump to *(n + b)*, if *SAC* $\neq 0$, and in either case increase *SAC* by 1"

All these instructions are B-modified in the usual way, but the following instruction is an exception.

18B,n means "jump to *n*, if number in *B*-register *last altered* is $\neq 0$, and in either case increase *b* by 1"

Here *b* is the number in the *B*-register specified (which could also be the register tested).

5. *Comparison instructions*

These are similar to the *03, 13, 23, 33* (subtraction) instructions, except that the *B*-register in question (or *SAC*) is not actually altered, although the subsequent effect of the test instructions *09, 08, 18; 29, 28, 38* would be the same. They thus enable the programmer to compare a *B*-register (or *SAC*) with a given quantity without actually having to form the difference proper.

instruction	effect	
07B, n	*b − H(n)*	→ "*B*-test" register
17B, n	*b − n*	
27B, n	*SAC − H(n)*	→ "*SAC*-test register
37B, n	*SAC − n*	

6. *Magnetic transfer instructions*

67B, n means "*select* sector *(n + b)*"
68B, n means "copy *sector* last *selected* to *page (n + b)*"
69B, n means "copy *page (n + b)* to *sector last selected*"

7. The only other instruction referred to in the text is the *dummy* instruction

570, 0

but instructions also exist for various other purposes, e.g., operating the input/output devices.

APPENDIX 2*

The layout of the programme on the magnetic drum has been altered since this paper was written, and the table given on page 214 (lines 5 - 34) should now be replaced by the following version. (On page 208, line 16 it is also necessary to alter the references to *sectors 64, 65* to *128, 129* respectively).

Layout of material on magnetic drum

sectors 0, 1 basic binary input-output sequences (these *sectors* are isolated and have no connection with the Autocode programme other than to read it in to the machine).

sectors 2 - 25

2	*up* chapter changing sequence and *restore*
3	*page 0* material (see previous table)
4	part of *read* routine (mainly negative powers of 10)
5	*down* and *across chapter* changing sequences, and *preserve*
6	*tan*gent routine
7,8	*print* routine
9	*read* routine
10	*arctan*gent routine
11	auxiliary transfer sequences and miscellaneous material
12	*sine-cos*ine routine
13	*exp*onential routine
14	*log*arithm routine
15	*square root* routine
16	pseudo *random* number generator
17	*test* routine
18	special input routine (for Sunvic-logger)
19	starting routine
20	*read* routine (floating decimal extension)
21	*read* routine (extension to provide ϕ_{10})
22	*print* routine (extension to provide ϕ_8, ϕ_9)
23	*print* routine (floating decimal extension)
24	translates *programme* directives) part of
25	translates *across* and *down* directives) assembly programme

sectors 32 - 47

A "secret" dump used to record the contents of *pp. 16 - 31* (the working variables) when executing the *rmp* instruction (read more programme) and certain matrix operations.

---

*Added after the Symposium.

sectors 48 - 63 & 64 - 79
Further dumps used to record *pp. 16 - 31* should they need to be preserved when calling in a *subchapter* which makes independent use of these *pages*. Thus *sectors 48 - 63* are used in this way by a *preserve* instruction given in a master *chapter*, and *64 - 79* by a *preserve* instruction in a *subchapter*.

sectors 80 - 127 Autocode conversion programme.

sectors 128 - 463 Auxiliary variables/*chapters* (see below)

sectors 464 - 479 *Chapter 0*.

sectors 480 - 511 These are normally occupied by the engineering test routines unless they are required for the Matrix Algebra programme.

The auxiliary variables/chapters

Starting at the first location on *sector 128*, the auxiliary locations are numbered continuously from *0 - 10,751*. Not all of these are available, however, since the *chapters 1 - N* occupy the last *512 N* locations. Thus in a programme extending up to and including *chapter 4*, the highest available auxiliary location is 8703.

Sectors 32 - 127 correspond to the negatively numbered locations *-3072* to *-1*, and these can be used as further auxiliary storage under appropriate restrictions. Thus e.g., if the Autocode conversion programme is no longer required (which means dispensing with the *rmp* facility) then locations *-1* to *-1536* are available, together with those corresponding to the "secret" dump *(-2561* to *-3072)* unless this should be required for any matrix operations. Similarly *sectors 480 - 511* corresponding to locations *10,752 - 11,775* can be used in appropriate circumstances.

Certain *additional* facilities have also been provided since the paper was written, namely, the ϕ_6 and ϕ_7 auxiliary transfer instructions, the *preserve* and *restore* instructions, and the generalised form of the *print* instruction. Nevertheless these do not invalidate any of the techniques described here.

DISCUSSION ON THE PAPER BY MR. R. A. BROOKER

MR. P. M. WOODWARD: I really want to ask quite a simple question as a mathematician. I want to know what is of fundamental mathematical interest in this game of writing compilers. It always seems to me that a compiler is impressive because of its sheer magnitude, but that its microscopic details are really quite trivial. I would have liked to have seen one of the speakers on the subject of automatic programming define some technical terms for the use of people who *write* compilers in addition to terms for the use of the "customer". We have a few terms in programming which I think are generally understood, such as interpretative routine, floating address and so forth, but I wonder whether we have yet quite enough of these terms (or the concepts to back them up) to talk about programming in a really satisfactory way. We know how difficult it is to write a paper about programming, and I think it is for the reasons I have described. I would like to ask the speaker whether he shares my views.

DR. J. PATRY: We talk about coding and explaining programmes. The best thing to do, I think, is to use the flow chart. If you wish to add the squares of all numbers up to N, for example, the flow chart is given in the figure. At first you add 1 to x, which is 0 at the beginning, take the square of the result and add it to Σ, which is also 0 at the beginning. At this point a choice must be made. If x is equal to N (or bigger) the programme must go further on. If x is smaller than N the loop must be run at least once more. This symbolic language will be understood by everybody. The programmer and the reader know what they are doing and another user can apply the code, he wants. I think, that the flow chart simplifies the papers on programming. It has also the advantage that the same programme can be used for every, or almost every computer.

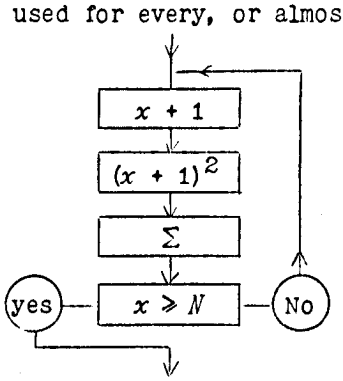


Fig.: Example of a flow chart

MR. D. W. BARRON. I think that in many cases Autocodes are used to disguise the shortcomings of a machine, and that it is possible to obtain many of the advantages of an Autocode from the basic machine design, as is done on EDSAC 2 at Cambridge. To achieve this, three things are required: a systematic order code with absolutely no exceptions, subroutines for common operations such as sine, cosine and the like wired in a permanent store so that they are available to the programmer as secondary machine instructions, and a comprehensive input routine providing floating address facilities, also wired in. With these facilities it is possible to have a simple programming scheme that can be taught in one or two weeks, and no time is wasted in compiling every time a programme is run.

I am surprised that with the Mercury Autocode it is not quicker to compile once and for all. With the scheme I have outlined, it is possible to have finalised programmes punched out in binary, which reduces the length of the programme tape and can be read at approximately twice the speed of normal programme.

DR. S. DOUGLAS: May I anticipate Dr. Brooker's answer to the last speaker? Firstly, he has raised the question of design of the machine, in particular the absorption of a suitable code into the design. It is true that this can be done, but it may nevertheless be cheaper to do it by programming. He does not appear to have given this due consideration. It is not a question necessarily of concealing the design of the machine. It is, if you like, supplementing it, and Dr. Barron's built-in routines are no more than this. The provision of these amounts to substituting a programme for some hardware. We may regard the choice of method used as a matter of straightforward cash and nothing else. Secondly, I would remark that two weeks is a long time for initial programme training. Just recently we held a course for beginners in the use of Autocode on Pegasus lasting one afternoon and at the end of this time a number of those attending were able to write programmes for it. I do not doubt a similar training in Mercury Autocode, which is rather more difficult than curs, could be given in less than three days.

DR. J. McCARTHY: I would like to ask how many programmers are using the computers at Manchester and Cambridge? We have an IBM 704 at M.I.T., and about 400 programmers use it - people from all departments of M.I.T. and other New England schools. About half the machine-time is spent either on Fortran compilations or running programmes that were produced by Fortran. For the rest of our work we use the standard SHARE assembly programme which is a symbolic address scheme. The assembly and operation of SHARE language programs is controlled by the M.I.T. automatic operation system.

It would be interesting to compare the sizes of the programming groups using various machines. It makes a considerable difference whether you have twenty people programming for a machine, or whether you have several hundred people using it and you have little control over their programming practice.

DR. M. L. MINSKY: The question of whether to build your compiler into the machine or realize it as a programme is not just one of initial cost. The art of automatic programming is advancing at such a rate that it would be folly to commit any particular scheme to wiring, for we cannot rewire our computers every year.

I would like to add a remark in connection with a point raised by Dr. McCarthy. This concerns the problem of sharing methods and routines with people in other computing centres. It is a good thing to learn how to design machines, and to design programming systems. But one pays a large price if one owns a one-of-a-kind machine or programming system. Unless it is your specific goal to learn these particular arts, and actually doing computation is secondary, it might be wise to design your machine (and programming language) so that it will be compatible with programmes written for one of the more widely available machines. One might as well take advantage of the fact that there has been no system of copyright protection for programmes and programming systems---they may be regarded as much like mathematical formulations---and do things so that your laboratory will be able to use and contribute to the pools of exchangeable programmes.

It is true that at this time there is no convenient form in which to publish the description of a programme, as Dr. Woodward has noted. This problem may be considerably alleviated as more powerful automatic programming systems are developed, because these languages are becoming more expressive. But there will have to be some standardizations before the benefits of this expressiveness can be used in public communications.

DR. J. W. BRAY: As a user of Autocode I could perhaps answer the question about numbers of people using it. In ICI I think we have something like fifty people who at present are writing programmes in Autocode, and these, of course, are not all mathematicians, but mostly chemists and engineers. Then I would like to ask Mr. Brooker whether he expects to put in further programme checking facilities? There are some already, but Autocode still takes a great deal of debugging, at any rate for most of us, and there are many further checks which could be written in. The only one I would mention is the overlapping use of the stores. The other point is whether MERCURY could not be used to translate its own Autocode into machine codes of all the other computers which are in use in this country at present, the IBM.650, the Elliott machines, and so on.

MR. R. A. BROOKER:

(In reply to MR. WOODWARD)

I do not think there is any fundamental mathematical interest in assembly routines; they are simply programmes of a logical nature concerned with recognising items of information and taking appropriate action, e.g., recording related information in a directory. As regards programming terms, I can only add "generator" and "directory", - the former being, I understand, the American term for what I have referred to in my paper as a "skeleton".

(In reply to DR. PATRY)

As regards flow diagrams, I agree that these are very useful aids to understanding programmes. Part of their usefulness, however, lies, I am sure, in the fact that they are two-dimensional, whereas a programme as usually understood is an ordered list of instructions following the left-hand margin of a page. This need not necessarily be the case, however, and there is no reason, for example, why we should not, when arriving at a branch, place the instructions to which we want to jump somewhere on the right-hand side of the page opposite the branch point. This would make it easier to follow and, if we are using floating addresses, it could be left in this form until it came to be punched, when those sequences "out on a limb" can be added to the main sequence following a convenient unconditional branch instruction.

(In reply to MR. D. W. BARRON)

I agree with Mr. Barron that in many cases the motivation for writing Autocode programmes is to disguise the shortcomings of a machine, but it would be very unwise to say that the story ends there. A systematic order code with absolutely no exceptions would be very desirable, but it would still be limited to translating those concepts which can be specified with a single address (in the case of one-address machines) and one operation code. An Autocode bears the same relation to such a code as sentences to do words, and I think most people would be happier to express their thoughts in the former and less disjointed units of information.

As regards the duration of a programming course, I agree with Dr. Douglas when he says that two weeks is a long time for initial programme training. A Mercury Autocode course can take three days, but it depends how much you want to put across. If you merely want to teach someone to code the sort of problem that uses only the fast store, then it need take only an afternoon.

As to Mr. Barron's last point concerning the economics of compiling once and for all, I can only say that we very seldom find it convenient to

punch finalised programmes in binary form. This is possibly because there is such a large turnover of one-off jobs, but also because there will be a very much greater amount of information to be handled in this form. Moreover, this mode of operation assumes a reliable punch.

(In reply to DR. J. McCARTHY)

How many programmers have we got? Dr. Bray threw some light on this question, insofar as he revealed that in I.C.I. alone there are about fifty people writing Autocode programmes, and almost entirely catered for by postal service. I think altogether Mercury serves between 100 and 150 programmers, of whom probably half are dealt with in this fashion. About 60% of our work is done in Autocode, and in other Mercury installations the figure is still higher. We have no automatic operator system of the kind which I think Dr. McCarthy refers to, but the tapes are nevertheless dealt with in a fairly automatic fashion by the operating personnel, and only a few decisions have to be left to them.

(In reply to DR. BRAY)

A programme was under development which would scan an Autocoded programme to detect instructions in which the permissible format had been violated. Obviously this could have been incorporated into the assembly routine itself, but was not. Unfortunately the man in charge of this project has recently died, and it may be some time before it can be restarted. At the moment violation of the format is sometimes detected by the assembly routine during input, and sometimes gets translated into faulty programme, which of course is most undesirable. Surprisingly, however, mistakes of this sort are less common than one would think; probably because the format is a natural one, and the rule book short and simple to understand. The particular check which Dr. Bray refers to, the possibility of overlapping use of the stores, is not possible to incorporate because such use of the stores is an unorthodox mode of operation.

As regards translation into other machine languages, I do not think this could be done efficiently for the machines available in this country, because they are all two-level storage machines of very diverse kinds. Interchange of programmes, and the idea of a universal code, is all very well if you have at your disposal a race of giant one-address machines, but is hardly a feasible proposition at the present stage of development in this country.

SESSION 2

PAPER 3

AUTOMATIC PROGRAMMING PROPERTIES AND PERFORMANCE OF FORTRAN SYSTEMS I AND II

by

J. W. BACKUS

BIOGRAPHICAL NOTES

John W. Backus received his A.M. degree (mathematics) at Columbia University in 1950. Since 1950 he has been with International Business Machine Corporation working on programming, computer systems design, and automatic programming systems design. He is at present manager of the Programming Research Department.

He is author of "The IBM 701 Speedcoding System" (1953) and joint author (with G. M. Amdahl) of "The System Design of the IBM Type 704" (1955) presented to meetings of the Association for Computing Machinery.

AUTOMATIC PROGRAMMING:

PROPERTIES AND PERFORMANCE OF FORTRAN SYSTEMS I AND II

by

J. W. BACKUS

SUMMARY

A BRIEF general discussion of the goals and methods of automatic programming techniques is followed by a somewhat detailed description of the input languages of FORTRAN Automatic Coding Systems I and II. The statements of these input languages provide (in part) concise means for writing algebraic expressions, for specifying iterative repetitions of portions of a procedure, for referring to one, two and three dimensional arrays of data, for defining new functions and procedures, and for specifying input and output procedures. Recent extensions of the input language are described.

FORTRAN I has been in use for a year and a half. Over half the instructions being written for some sixty 704 installations are thought to be produced by FORTRAN. The cost of programming and "debugging" is reduced by about 4-to-1. A few other statistics and applications are cited. The final section discusses the relationship of automatic programming and the mechanization of thought processes.

INTRODUCTION

AUTOMATIC programming techniques seek to make it easy to do a desired job on a computer. This is accomplished by providing a language in which the statement of the desired process is natural and concise, together with some means of causing a machine to carry out the stated process. Given a machine equipped to accept statements of procedures in a language other than the "hardware language", one may ask the same questions about this "synthetic" machine that are used to evaluate a real machine: (1) how easy is it to prepare programs for, and (2) how rapidly does it execute the programs for

which it will be used? In general, automatic programming systems provide synthetic machines which are far superior to their real counterparts with respect to question (1) and range from slightly inferior to greatly inferior to their real counterparts on question (2).

Automatic programming systems have been widely used in the United States for a number of years. The purpose of the early systems was to provide synthetic machines which had floating-point operations and often index registers (B-tubes), since the real machines did not. In many instances the synthetic machines were only one-half to one-tenth as fast as their real counterparts, but although their input languages were machine-like, they were considerably easier to program.

In the past eighteen months a number of new automatic programming systems have been completed, and others are being developed, which have strikingly similar input languages. These languages use mathematical notation and a variety of statements which make the writing of most procedures of numerical computation a natural and concise process. The synthetic machines provided by these systems range in speed from virtually the speed of the corresponding real machine to perhaps one-half of that speed. FORTRAN is one of these systems. Considerable effort was devoted to making the speed of the synthetic FORTRAN machine as close as possible to that of the IBM 704, its real counterpart. This paper will describe the language and some of the recent developments of this system and report on some of the effects which it has had on the computing work at a number of 704 installations.

In the field of commercial data processing a number of automatic programming systems have recently been under development which provide considerable simplifications in specifying jobs in this area. Automatic programming systems for data processing, however, still face several difficult problems:

(1) Existing languages are often too rigid to describe various desirable procedures and situations and, in many instances, are not sufficiently concise.

(2) The need for synthetic machines to be virtually as fast as their real counterparts is much greater in the data processing area than in computing and the difficulty of achieving this is greater.

PROPERTIES OF FORTRAN

GENERAL

The FORTRAN synthetic computer is realized by a 704 program, called the FORTRAN translator, which accepts programs written in FORTRAN language and produces corresponding 704 machine language programs. The translator

program was written over a period of two and one-half years with an expenditure of about 18 man-years of effort; it comprises a program of about 25,000 instructions. The translator program was made available to all 704 installations in April, 1957. A report of the usage of the system since that time appears in a subsequent section of this paper.

Since the initial distribution of the FORTRAN system, certain extensions have been made to the input language and the translator has been considerably modified to accept this extended language. The resulting system is known as FORTRAN II; it has recently been distributed to 704 installations†.

FORTRAN I INPUT LANGUAGE

The following description-by-example of the FORTRAN I input language is extracted from a previously published paper *(ref. 1)*.

Arithmetic Statements

Example 1: Compute:

$$\text{root} = \frac{-(B/2) + \sqrt{(B/2)^2 - AC}}{A}$$

FORTRAN Program:

```
ROOT = (-(B/2.0) + SQRTF((B/2.0)**2 - A*C))/A
```

Notice that the desired program is a single FORTRAN statement, an arithmetic formula. Its meaning is: "Evaluate the expression on the right of the = sign and make this the value of the variable on the left." The symbol * denotes multiplication and ** denotes exponentiation (i.e., A**B means A^B). The program which is generated from this statement effects the computation in floating-point arithmetic, avoids computing (B/2.0) twice and computes (B/2.0)**2 by a multiplication rather than by an exponentiation routine. Had (B/2.0)**2.01 appeared instead, an exponentiation routine would necessarily be used, requiring more time than the multiplication.

The programmer can refer to quantities in both floating-point and integer form. Integer quantities are somewhat restricted in their use and serve primarily as subscripts or exponents. Integer constants are written without a decimal point. Example: 2 (integer form) vs 2.0 (floating-point form). Integer variables begin with I, J, K, L, M, or N. Any meaningful arithmetic expression may appear on the right-hand side of an arithmetic statement, provided the following restriction is observed: an integer quantity can appear in a floating-point expression only as a subscript or as an exponent or as the argument of certain functions. The functions which

† The new sections of the FORTRAN II translator for the 704 were written by G. E. Mitchell, P. B. Sheridan, B. Brady and L. May with the assistance of the authors of the original FORTRAN system.

the programmer may refer to are limited only by those available on the library tape at the time, such as SQRTF, plus those simple functions which he has defined for the given problem by means of function statements. An example will serve to describe the latter.

Function Statements

Example 2: Define a function of three variables to be used throughout a given problem, as follows:

```
ROOTF(A,B,C) = (-(B/2.0) + SQRTF((B/2.0)**2 - A*C))/A
```

Function statements must precede the rest of the program. They are composed of the desired function name (ending in F) followed by any desired arguments which appear in the arithmetic expression on the right of the = sign. The definition of a function may employ any previously defined functions. Having defined ROOTF as above, the programmer may apply it to any set of arguments in any subsequent arithmetic statements. For example, a later arithmetic statement might be

```
THETA = 1.0 + GAMMA*ROOTF(PI, 3.2*Y + 14.0, 7.63)
```

DO Statements, DIMENSION Statements, and Subscripted Variables

Example 3: Set Q_{max} equal to the largest quantity $P(a_1+b_1)/(P(a_1-b_1)$ for some i between 1 and 1000 where $P(x) = c_0 + c_1x + c_2x^2 + c_3x^3$.

FORTRAN Program:

```
1     POLYF(X) = C0+X*(C1+X*(C2+X*C3))
2     DIMENSION A(1000), B(1000)
3     QMAX = -1.0E20
4     DO 5 I = 1, 1000
5     QMAX = MAXF(QMAX, POLYF(A(I) + B(I))/POLYF(A(I) - B(I)))
6     STOP
```

The program above is complete except for input and output statements which will be described later. The first statement is not executed; it defines the desired polynomial (in factored form for efficient output program). Similarly, the second statement merely informs the executive routine that the vectors A and B each have 1000 elements. Statement 3 assigns a large negative initial value to QMAX, $-10x10^{20}$, using a special concise form for writing floating-point constants. Statement 4 says "DO the following sequence of statements down to and including the statement numbered 5 for successive values of I from 1 to 1000." In this case there is only one statement 5 to be repeated. It is executed 1000 times; the

first time reference is made to A(1) and B(1), the second time to A(2) and B(2), etc. After the 1000th execution of statement 5, statement 6 -- STOP -- is finally encountered. In statement 5, the function MAXF appears. MAXF may have two or more arguments and its value, by definition, is the value of its largest argument. Thus on each repetition of statement 5 the old value of QMAX is replaced by itself or by the value of POLYF(A(I)+ B(I))/POLYF(A(I) - B(I)), whichever is larger. The value of QMAX after the 1000th repetition is therefore the desired maximum.

Example 4: Multiply the $n \times n$ matrix $a_{ij} (n \leq 20)$ by its transpose, obtaining the produce elements on or below the main diagonal by the relation

$$c_{i,j} = \sum_{k=1}^{n} a_{i,k} \cdot a_{j,k} \qquad (\text{for } j \leq i)$$

and the remaining elements by the relation

$$c_{j,i} = c_{i,j}.$$

FORTRAN Program:

```
      DIMENSION A(20,20), C(20,20)
      DO 2 I = 1, N                                        P
      DO 2 J = 1, I                                        Q
      C(I,J) = 0.0
      DO 1 K = 1, N                                        R
1     C(I,J) = C(I,J) + A(I,K)*A(J,K)
2     C(J,I) = C(I,J)
      STOP
```

As in the preceding example, the DIMENSION statement says that there are two matrices of maximum size 20x20 named A and C. For explanatory purposes only, the three boxes around the program show the sequence of statements controlled by each DO statement. The first DO statement says that procedure P, i.e., the following statements through statement 2 (outer box) is to be carried out for I = 1 then for I = 2 and so on up to I = N. The first statement of procedure P(DO 2 J = 1, I) directs that procedure Q be done for J = 1 to J = I. And of course each execution of procedure Q involves N executions of procedure R for K = 1, 2,..., N.

Consider procedure Q. Each time its last statement is completed the "index" J of its controlling DO statement is increased by 1 and control

goes to the first statement of Q, until finally its last statement is reached and J = I. Since this is also the last statement of P and P has not been repeated until I = N, I will be increased and control will then pass to the first statement of P. This statement (DO 2 J = 1, I) causes the repetition of Q to begin again. Finally, the last statement of Q and P (statement 2) will be reached with J = I and I = N, meaning that both Q and P have been repeated the required number of times. Control will then go to the next statement, STOP. Each time R is executed a new term is added to a product element. Each time Q is executed a new product element and its mate are obtained. Each time P is executed a product row (over to the diagonal) and the corresponding column (down to the diagonal) are obtained.

READ, PRINT, FORMAT, IF and GO TO Statements

Example 5: For each case, read from cards two vectors, ALPHA and RHO, and the number ARG. ALPHA and RHO each have 25 elements and ALPHA(I)≤ ALPHA(I+1), I = 1 to 24. Find the SUM of all the elements of ALPHA from the beginning to the last one which is less than or equal to ARG [assume ALPHA(1)≤ARG<ALPHA(25)]. If this last element is the Nth, set VALUE = 3.14159*RHO(N). Print a line for each case with ARG, SUM, and VALUE.

FORTRAN Program:

```
      DIMENSION ALPHA(25), RHO(25)
1     FORMAT(5F12.4)
2     READ 1, ALPHA, RHO, ARG
      SUM = 0.0
      DO 3 I = 1, 25
      IF (ARG - ALPHA(I))4, 3, 3
3     SUM = SUM+ALPHA(I)
4     VALUE = 3.14159*RHO(I - 1)
      PRINT 1, ARG, SUM, VALUE
      GO TO 2.
```

The FORMAT statement says that numbers are to be found (or printed) 5 per card (or line), that each number is in Fixed-point form, that each number occupies a field 12 columns wide and that the decimal point is located 4 digits from the right. The FORMAT statement is not executed; it is referred to by the READ and PRINT statements to describe the desired arrangement of data in the external medium.

The READ statement says "READ cards in the card reader which are arranged according to FORMAT statement 1 and assign the successive numbers

obtained as values of ALPHA(I) I = 1, 25 and RHO(I) I = 1, 25 and ARG." Thus "ALPHA, RHO, ARG" is a description of a list of 51 quantities (the size of ALPHA and RHO being obtained from the DIMENSION statement). Reading of cards proceeds until these 51 quantities have been obtained, each card having five numbers, as per the FORMAT description, except the last which has the value of ARG only. Since ARG terminated the list, the remaining four fields on the last card are not read. The PRINT statement is similar to READ except that it specifies a list of only three quantities. Thus each execution of PRINT causes a single line to be printed with ARG, SUM, VALUE printed in the first three of the five fields described by FORMAT statement 1.

The IF statement says "If ARG - ALPHA(I) is negative go to statement 4, if it is zero go to statement 3, and if it is positive go to 3." Thus the repetition of the two statements controlled by the DO consists normally of computing ARG - ALPHA(I), finding it zero or positive, and going to statement 3 followed by the next repetition. However, when I has been increased to the extent that the first ALPHA exceeding ARG is encountered, control will pass to statement 4. Note that this statement does not belong to the sequence controlled by the DO. In such cases, the repetition specified by the DO is terminated and the value of the index (in this case I) is preserved. Thus if the first ALPHA exceeding ARG were ALPHA(20), then RHO(19) would be obtained in statement 4.

The GO TO statement, of course, passes control to statement 2, which initiates reading the 11 cards for the next case. The process will continue until there are no more cards in the reader. The above program is entirely complete. When punched in cards as shown, and compiled, the translator will produce a ready-to-run 704 program which will perform the job specified.

Other Types of FORTRAN Statements

In the above examples the following types of FORTRAN statements have been exhibited.

Arithmetic statements
Function statements
DO statements
IF statements
GO TO statements
READ statements
PRINT statements
STOP statements
DIMENSION statements
FORMAT statements

The explanations accompanying each example have attempted to show some of the possible applications and variations of these statements. It is felt that these examples give a representative picture of the FORTRAN language; however, many of its features have had to be omitted. There are 23 other types of statements in the language, many of them completely analogous to some of those described here. They provide facilities for referring to other input-output and auxiliary storage devices (tapes, drums, and card punch), for specifying preset and computed branching of control, for detecting various conditions which may arise such as an attempt to divide by zero, and for providing various information about a program to the translator. A complete description of the language is to be found in "Programmer's Reference Manual, the FORTRAN Automatic Coding System for the IBM 704".

Preparation of a Program for Translation

The translator accepts statements punched one per card (continuation cards may be used for very long statements). There is a separate key on the keypunching device for each character used in FORTRAN statements and each character is represented in a single column of the card. Keypunching a FORTRAN program is, therefore, a process similar to that of typing the program.

Translation

The deck of cards obtained by keypunching may then be put in the card reader of a 704 equipped with the translator program. When the load button is pressed one gets either 1) a list of input statements which fail to conform to specifications of the FORTRAN language accompanied by remarks which indicate the type of error in each case; 2) a deck of binary cards representing the desired 704 program; 3) a binary tape of the program which can either be preserved or loaded and executed immediately after translation is complete; or 4) a tape containing the output program in symbolic form suitable for alteration and later assembly.

FORTRAN II INPUT LANGUAGE

The principal shortcoming of the FORTRAN I input language is that the programmer is unable to define new statements in terms of the given ones. He can introduce only functions which he can define by an arithmetic expression. The input language of FORTRAN II is identical to that of FORTRAN I except that certain additional statements are provided. With these the programmer may create a FORTRAN language subprogram, assign a name to it, and indicate those variables in the subprogram which are to be regarded as the inputs and/or outputs. He may invoke a subprogram in any

other program by giving its name and supplying appropriate parameters. There is provision for indicating that a subprogram defines either a function (which may be employed in arithmetic expressions in other programs) or a subroutine (which may be invoked as a statement in other programs).

Example 1: Write a subprogram which clears the first *N* elements of a vector.

FORTRAN II Subprogram:

```
      SUBROUTINE CLEAR(A,N)
      DIMENSION A(1000)
      DO 1 J = 1, N
1     A(J) = 0.0
      RETURN
```

The first statement indicates that the entire program following is a subroutine, that its name is CLEAR, and that its parameters are the vector A and the integer variable N. The last statement indicates that the process which constitutes the subroutine is complete.

Example 2: Write a subprogram which realizes the function:

$$g(a,x) = \sum_{i=1}^{10} a^i \left(\sum_{j=1}^{10} x_{ij} \right)$$

FORTRAN II Subprogram:

```
      FUNCTION G(A,X)
      DIMENSION X(10,10), R(10)
      CALL CLEAR (R,10)
      G = 0.0
      DO 2 I = 1, 10
      DO 1 J = 1, 10
1     R(I) = R(I)+X(I,J)
2     G = G+A**I*R(I)
      RETURN
```

The first statement says that the following subprogram defines a function G of A and X. The CALL statement invokes the subroutine of example 1 to clear the vector R. Notice that the next statement and subsequent ones refer to a variable whose name is that of the function. The value of the variable G when RETURN is executed is the value of the function.

The above subprograms may be translated independently. The output in each case will be a deck of punched cards having the machine language program, with the appropriate prologue for linkage, in a form which may be easily adjusted by a loading program to operate anywhere in the store. In this form the names themselves of called subprograms are retained in tne deck representing the calling program and are used by the loading program to establish the appropriate references of each program to its subprograms as they are being loaded. This complete independence of a program from its calling and called programs until the point of loading for execution has several benefits:

(1) Each program and subprogram may be checked out and/or recompiled independently. Thus each correction in the FORTRAN language statement of a process will involve retranslation of only that subprogram in which the change is made.

(2) Each subprogram is produced in precisely the form suitable for inclusion in the library. Inclusion is accomplished simply by placing the output deck in the library deck. When loaded this deck will produce a new library tape which is available to the translator. Library programs, of course, require no translation.

(3) Since each subprogram may be obtained without the inclusion of the subprograms which it calls, the use of two subprograms, which both call a third, need not result in a copy of the third accompanying each of them in the store.

(4) Since the machine-language calling sequences resulting from input language references to functions or subroutines are quite standard, it is immaterial whether a required subprogram is generated by FORTRAN or is written in machine-language. Thus, for example, a special hand-coded input routine may be invoked by a CALL statement.

Example 3: Compute $g(sin(g(\pi,r))^2,s)$ for pairs of 10x10 matrices r, s read from cards, and print the result for each case, where g is the function of example 2.

FORTRAN II Program

```
      DIMENSION R(10,10), S(10,10)
3     READ 1, R, S,
1     FORMAT (10F6.3)
      RESULT = G(SIN(G(3.14159,R))**2,S)
      PRINT 2, RESULT
2     FORMAT (E12.4)
      GO TO 3
```

When this program is translated one may request that any called subprograms which are on the library tape plus a copy of the loading program be incorporated in the output deck. In this case the given program and the subprogram for the sine function might be contained in the output. When the outputs from translating the programs of the first two examples are added to this deck, it is ready to be loaded and run.

FURTHER DEVELOPMENTS IN FORTRAN

In addition to the facilities provided by FORTRAN II, work on the translator program for the 704 has been almost completed to provide the following facilities†.

Use of Symbolic Machine Instructions as Statements in a FORTRAN Program

This facility, of course, makes it possible for the programmer to use operations and to deal with units of information not provided for in the FORTRAN language. Certain restrictions are necessary to effect an appropriate relationship between machine language references to information and references in FORTRAN statements within the same FORTRAN program.

Example: Write a subroutine which sets B(I) = 0.0 or B(I) = 1.0 according as the Ith bit of cell A is 0 or 1 for I = 1, 2,..., N where N ≤ 36.

FORTRAN III Program:

```
      SUBROUTINE EXPAND (A,B,N)
      DIMENSION B(36)
S     LDQ A
      DO 1 I = 1, N
S     CLM
S     LGL 1
S     TZE*1
      B(I) = 1.0
S     1 STO B(I)
      RETURN
```

In the above program those statements preceded by an "S" are symbolic 704 instructions. For example, the first such statement is "Load the Q-register from cell A". The asterisk in the fourth S-statement indicates

† The principal additions to the translator to provide these facilities were planned and written by I. Ziller and R. A. Nelson.

that the number following is a statement number, which in this case is that of the fifth S-statement. Note that the address part of the latter is a subscripted variable. For those who care to understand this not-very-good program the operation codes have the following meanings:

CLM: clear the accumulator.

LGL: shift the contents of the 73 bit register comprising the accumulator (left 37 bits) and the Q-register (right 36 bits) one bit to the left.

TZE: transfer control when contents of accumulator are zero.

STO: store contents of accumulator.

Note the services provided by the translator even for essentially non-numerical processes, in this case: generation of a prologue, assignment of storage, loop formation, referencing of arrays.

Use of Boolean Expressions

Provision has been made to permit writing "expressions" in which the operators are "and", "or" and "not" rather than the usual arithmetic operations. When two words are combined under the operation "and" the resulting word has a zero-bit in every position except those in which the two operand words both have a 1-bit. Analogous results are given by "or" and "not", the latter operation having a single operand. "And", "or" and "not" are designated respectively by "*", "+" and "-".

Example: Let A, B, C, D be four quantities which are zero or one according as conditions a, b, c, d respectively, hold or not. Write a statement which makes E zero or one according as the condition

$$a + (-(b*c)*(-d))$$

holds or not.

FORTRAN III Statement

```
B     E = A+(-(B*C)*(-D))
```

The "B" preceding the statement indicates that the operators are to be interpreted as Boolean operators. "B" may be applied in this way to any statement which contains an expression.

Thus,

```
B     IF(A*B) 0,5,14
```

will cause control to pass to statement 5 only if A*B (A "and" B) is zero; otherwise control will pass to statement 14.

Use of Function or Subroutine Names as Parameters of Functions or Subroutines

FORTRAN II permits a function *value* to appear as a parameter of a function or subroutine, [e.g., SIN(COS(X)) meaning the sine of the number, cos(X)]. FORTRAN III has provision for specifying the name of a particular function or subroutine as a parameter of another function or subroutine which refers to one or more *arbitrary* functions or subroutines.

Example: Write a subprogram, R, which, for any arbitrary functions f and g, replaces each element, A_i, $i = 1, 2, \ldots, N$, of a vector A, by

$$f(A_i)^2 + g(A_i+1)^2$$

FORTRAN III Subprogram:

```
      SUBROUTINE R(A,N,F,G)
      DIMENSION A(1000)
      DO 1 I = 1, N
1     A(I) = SQRT(F(A(I))**2+G(A(I+1))**2)
      RETURN
```

Write a subprogram, S, which replaces X_i, $i = 1, 2, \ldots 10$, by

$$sin(x_i)^2 + cos(x_i+1)^2$$

FORTRAN III Subprogram:

```
      SUBROUTINE S(X)
      DIMENSION X(11)
F     SIN, COS
      CALL R(X, 10, SIN, COS)
      RETURN
```

The first of the subprograms above is unusual only in that the names of two functions, F and G, are listed as parameters. The second subprogram calls the first and specifies that SIN and COS are to be used for F and G respectively. The statement preceded by "F" makes it possible to correctly interpret the CALL statement by indicating that "SIN" and "COS" are names of subprograms rather than names of variables, and, therefore, must be retained in symbolic form (up to loading time).

In addition to the above mentioned facilities for stating procedures which the FORTRAN III translator will accept [namely, (1) use of symbolic machine instructions as statements, (2) use of Boolean expressions, (3) use of subprogram names as parameters] there are facilities provided for writing procedures to manipulate alphabetic and other symbolic information. No further discussion of the latter is given here.

Operation of Programs Which Require More than the Available High-Speed Storage Space

The preceding paragraphs describe certain new properties of the FORTRAN translator program. The present section concerns a program called SLAM† (Subroutine Loader And Monitor). SLAM is concerned with the subprograms from which a program is built. At load time it splits these subprograms into groups, each of which will fit into the core space available after the data area has been reserved. It writes these blocks onto tape and at the same time rearranges the linkage between subprograms. Then at execution time the monitor part of SLAM puts in cores, at the appropriate moment, that group of subprograms necessary for the execution of the program. The monitor takes care of the transfer of arguments between subprograms.

EXPERIENCE WITH FORTRAN

Since its distribution over a year ago, the use of FORTRAN I has been steadily increasing. A survey in April of this year of twenty-six 704 installations indicates that over half of them use FORTRAN for more than half of their problems. Many use it for 80% or more of their work (particularly the newer installations) and almost all use it for some of their work. The latest records of the 704 users' organization, SHARE, show that there are some sixty installations equipped to use FORTRAN (representing 66 machines) and recent reports of usage indicate that more than half the machine instructions for these machines are being produced by FORTRAN. SHARE recently designated the FORTRAN language as the second official medium for transmittal of programs within the organization (the other is symbolic machine language).

Programming Economies

Most 704 installations cannot provide accurate information regarding programming costs. However, the Service Bureau Corporation, which does a lot of contract programming, has kept careful records of programming time, machine time used in debugging, and so forth, for many jobs coded in symbolic machine language and many others in FORTRAN. Their analysis of these costs shows that the average cost per debugged instruction obtained from FORTRAN programs is about one-fourth the corresponding cost per instruction obtained from symbolic programs. (A FORTRAN-produced program and a symbolic program for a given job are about the same length.) Since users have found that a FORTRAN-coded program runs at about the same speed as the corresponding hand-coded program, considerable savings are

† Written by M. deV. Roberts.

effected. (Where symbolic programming is used, the cost of programming and debugging is usually 1/2 to 3/4 the total cost of solving a problem.) Elapsed time for programming has often been reduced by factors of 10 to 1.

The average ratio of the number of FORTRAN statements in a program to the number of resulting machine instructions (exclusive of all subroutines) is 1 to 7. The translator produces about 100 instructions per minute.

Scope of Applications

Calculations have been programmed using FORTRAN in all the major areas of scientific computation (e.g., jet engine design, airframe design, reactor design, numerical weather prediction, X-ray diffraction analysis, and a multitude of others). It is obvious from the nature of the input language that FORTRAN offers numerous conveniences to programmers in these areas. It is interesting to note the conveniences offered, and hence, the use of the system, in the case of processes which are essentially non-numerical. For example, FORTRAN programs have been or are being written (with considerable payoff in simplicity) to do the following: simulate the behaviour of a control mechanism for a proposed computer (J. Cocke, IBM), play chess (J. McCarthy, MIT), print isobaric weather maps (H. A. Zartner, Jr., U. S. Weather Bureau), and prove theorems in geometry (H. Gelernter and N. Rochester, IBM). Many users who have not made much use of FORTRAN I report that FORTRAN II will greatly enlarge the system's usefulness for their work.

FORTRAN Translators for Various Machines

The preceding paragraphs have referred to various FORTRAN systems for the 704. There are also translator programs for other machines which accept FORTRAN-language programs and produce programs for the machine in question. FORTRANSIT I, II and III are translator programs for three configurations of the IBM 650. Some 265 different 650 installations have requested some of the FORTRANSIT translator programs. The actual extent of use which they make of the system is not known.

Work on FORTRAN translators for the IBM 705 and for the IBM 709 is nearing completion.

RELATIONSHIP OF AUTOMATIC PROGRAMMING TO THE MECHANIZATION OF THOUGHT PROCESSES

In the following remarks "mechanization of thought processes" or "thinking" for a machine will be taken to mean simply "information processing of such a nature and complexity that it leads to behaviour

which in a person would be called thinking". In these terms the technique of automatic programming has three roles in relationship to the study of machine thinking: one as recipient, two as contributor.

Certainly the process of devising an efficient machine program to perform some previously stated complex task may be taken to require thinking. As more general and more effective techniques are found for the mechanization of various thought-like processes, their application in the automatic programming area will undoubtedly permit the users of machines to communicate requests for programs with increasing ease and brevity and obtain more efficient programs to perform each desired task. Thus automatic programming will benefit from progress in the mechanization of thought processes.

Conversely, as a special topic in the field, efforts to find better languages for stating procedures and better techniques for translating them into machine language will contribute to the general understanding of complex information-processing methods. Already various translators such as FORTRAN perform some of the more complex information-processing tasks presently done by machines.

Almost by definition the mechanization of thought processes requires the construction of very large and complicated programs. Therefore, it would appear that progress in this area will depend heavily on the availability of automatic programming systems which will make it possible to construct many experimental programs of great length. Some of the people who have used FORTRAN for writing programs, such as the geometry program or the chess program mentioned earlier, feel that the task of writing them would be virtually impossible if they had to concern themselves with the complexities of machine coding in addition to those of the process itself. For example, it is expected that the geometry program will comprise about 16,000 machine instructions and that these will result from 3,000 to 4,000 FORTRAN statements. In this way automatic programming will enlarge the horizon of research in the mechanization of thought processes by permitting us to study processes which would otherwise be beyond our power to realize.

REFERENCE

1. BACKUS, J. W., BEEBER, R. J., BEST, S., GOLDBERG, R., HAIBT, L. M., HERRICK, H. L., NELSON, R. A., SAYRE, D., SHERIDAN, P. B., STERN, H., ZILLER, I., HUGHES, R. A., and NUTT, R. The FORTRAN Automatic Coding System. *Proceedings of the Western Joint Computer Conference, Los Angeles, California,* (Feb., 1957).

DISCUSSION ON THE PAPER BY J. W. BACKUS

DR. J. V. GARWICK: My interest in automatic programming is partly caused by the fact that I have for several years watched and to some extent worked on an autocoding project at the Norwegian Defence Research Establishment for the Ferranti Mercury computer. The project has much in common with Fortran, but at the time the project was started nothing was known of this system.

One great difference from Fortran is that nearly all the work has been done by one man, Mr. O. J. Dahl *(ref. 1)*. A great difficulty during the early stages of this work was the extreme complexity of the flow diagrams. We strongly felt the need of a method to simplify these diagrams. As this type of flow diagram only treats logical problems, we thought that maybe the mathematical language we used was unsuited and that a logical language should be used. A study of a book on mathematical logic was therefore undertaken.

The outcome of this study showed that apart from the elementary signs belonging to the domain of Boolean algebra, the two operators "for all x" written (x) and "there exists an x such that..." written (Ex) were very useful. We changed the notation of the latter to $E(x)$ and let it include in its meaning the operator "the value of x for which...", usually written Cx.

In what follows P is a logical statement depending upon the (free) variable x and which can be true or false. The two statements $(x)P$ and $E(x)P$ are then statements independent of x which may be true or false. Flow diagrams for the determination of their truth or falsehood are given in *fig. 1*

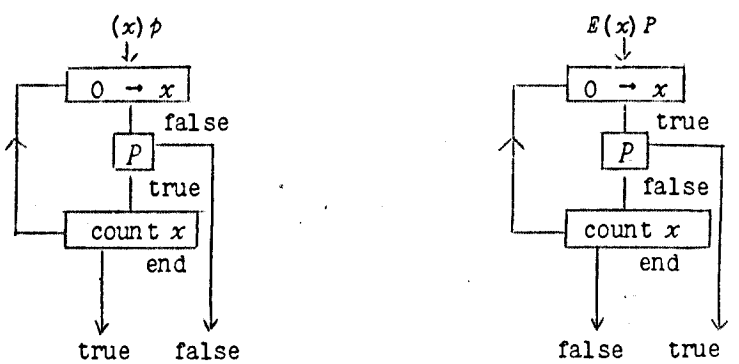


Fig. 1.

The power of the method is clearly seen if one wants to programme more complicated problems. As an example we take the statement in chess: "White can make a mate in two moves". Here convenient subroutines must define "white move no k" denoted w_k and similarly for black. We must also have subroutines for testing the statement "white is check" denoted C_w and similarly for black. We use $\sim$ as the sign for "not" and form the composite statement.

$$E(w_1)\ \{\sim C_w\ \&\ \{C_b v E(b_1)\ \sim C_b\}\ \&\ \{(b_1)E(w_2)\ \{\sim C_w\ \&\ C_b\ \&\ (b_2)C_b\}\}\}$$

$\sim C_w$ insures that white does not move into a check,

C_b v $E(b_1)$ $\sim C_b$ insures that black is not stalemate.

$(b_1)E(w_2)$ $\{\sim C_w$ & C_b & (b_2) $C_b\}$

insures that the problem has a solution. With the more convenient dot notation we obtain

$$E(w_1)\ .\cdot.\ \sim C_w \cdot C_b\ \text{v}\ E(b_1)\ \sim C_b\ :\ (b_1)E(w_2)\ \sim C_w \cdot C_b \cdot (b_2)\ C_b$$

With our convention that $E(w_1)$ not only decides the possibility of a white's first move but also fixes one of them if one or more exists, we see that this statement contains the entire logic of the chess problem.

When this method had shown its power to handle flow diagrams, it was but a small step to allow these operators to be among the symbols understood by the autocode.

As an example, we can take the multiplication of an 5 x 3 matrix stored by rows (5 rows each of 3 elements) with a 3 x 7 matrix also stored by rows to give a 5 x 7 matrix stored by columns. Our proposed system would write

FA = 5 x 3 (read: Field A is 5 by 3)
FB = 3 x 7
FC = 7 x 5 (observe interchange of factors)
$(i)(k)..0 \rightarrow C_{ik}$
$(s)\ .\ C_{ik} + A_{ks}\ B_{si} \rightarrow C_{ik}\ ..$

Further generalisations of the symbols like $(i < 10)$, $(5 \leq k < s)$, $E(s > k \geq 5)$ (count *down* from $s - 1$ to 5 in this case) etc. are obvious.

The slight experience we have, clearly indicates that when a logical problem is handled like running a war game or making an autocode, the language of logic is very superior to the language of mathematics.

REFERENCE

1. DAHL, O. J., Autocoding for the Ferranti Mercury Computer. Report No. 24 *Norwegian Defence Research Establishment.*

MR. E. A. NEWMAN: As I see it, all automatic programming aids have one or more of the following features.

1. Groups of instructions - subroutines - are available as part of a large program, and can be called in by use of a code name for the set.
2. Machine operations are specified in a language other than the machine code. The language used is prescribed but is hoped to be more natural for man to use.
3. Instructions are presented in a definite prescribed manner which is wrong for the machine, and are sorted automatically to the right order for the machine. The hope is that this wrong order will be more natural for man to use than the right machine one.
4. Labels can be used for variables which are not those of the locations where the variable will be stored, an automatic aid determining suitable locations.
5. If a group of operations has to be performed on a set of groups of variables, only the group of operations on one group of variables, and the set of groups of variables are labelled, the system looking after any necessary organisational instructions.
6. Some use can be made automatically of context.

One is tempted to suggest also

7. Instructions can be written down side by side instead of one under the other.

Does Mr. Backus agree that this list is comprehensive?

One does not have to have automatic coding facilities to have a comprehensive set of small and large subroutines. At N.P.L. our subroutines are stored under code names. These code names only are specified by the programmers, and the routines assembled very cheaply clerically. Operation 2 in my classification could also readily be done clerically. Items 3, 4, 5 and 6 are however essentially jobs for a machine. Two questions arise therefore, which Mr. Backus might answer. First in FORTRANS how are the savings one obtains using programming aids divided between the items I have given, and secondly, if, as I suspect, most of it arises from item 1, would it not be cheaper to use filed subroutines assembled clerically?

DR. J. McCARTHY: I have used FORTRAN for logical work in connection with a chess programme which has grown to the size of perhaps 1,500 FORTRAN statements which in this case runs to 5,000 machine instructions and I would agree with John Backus that in this case it has turned out that the automatic book-keeping features and the features for keeping track of tables which are in the chess programme were more important than the algebraic features.

I was also to some extent responsible for the use of FORTRAN in the IBM geometry programme. This was done, with the expectation, which I think has been realised, that an algebraic notation would turn out to be important in the symbolic work. In fact, some quite lengthy algebraic expressions are used. Single statements run to four or five lines of algebraic expression.

It turns out, of course, that the functions which are used are not mainly addition and multiplication but are compositions of certain basic subroutines, and it turns out to be quite useful to have the compiler generate these compositions automatically.

One feature of FORTRAN II which has proved very important - I am not sure whether all the other compiler systems have it - is the ability to compile subroutines separately and debug them separately and then load them into the machine together.

DR. A. S. DOUGLAS: It seems to me that Mr. Garwick is falling into the same mistake that Alan Turing did a long time ago. Way back in 1951 there was a conference in Manchester at which the base 32 was proposed as being far better than base 10. For some purposes this is no doubt true. However, in suggesting the immediate adoption of such a scheme Turing was ignoring the difficulty of retraining several million people. In order to work base 10 out of our system it would take more than 15 years and I rather think the same is true of algebraic notation.

The notation Garwick has suggested may well be successful in the long term but I do not see that it is any use in the short term. I feel strongly that we ought to consider what is the best notation, and start training students in the universities and schools to change over to this when we are sure we have a good system. But in the short term I agree with Grace Hopper that the right thing to do is to take the existing state of affairs as we find it. If Bill likes to write programs in English, let him write it in English or anything else that he can understand easily, rather than insist that he code in some new form now. It seems to me that the importance of schemes such as Aimaco, Fortran and Autocode is that they are useful short term solutions of the problem.

MR. R. A. BROOKER: I have a very brief question about a specific point in the paper, namely section 4 on page 242. Could you say something more about the treatment of programme parameters. How are they selected from the main programme and planted in the subprogramme?

DR. G. M. HOPPER: I feel there is a lack of full understanding of the systemisation of the Fortran. After the translation phrase Fortran is very sophisticated in its arrangement in automisation of the programme and its allocation of the storage in various components of the computer. It is far

above anything that could be considered an assembly system and much more sophisticated than that and when Mr. Backus has said it is housekeeping, it means far more than just housekeeping, I wish he would emphasise that.

It seems that the degree of systemisation of the Fortran system has not been fully stated. In addition to and following the translation phase, the Fortran system is extremely intelligent in the arrangement of the segments of the program, in the allocation of both program and data storage, and in the assignment of the various components of the computer system. It should be understood that it is far more sophisticated than a mere assembly routine. When Mr. Backus referred to the "housekeeping" perferred by the Fortran system, he was referring to more than is normally considered the housekeeping within a program. In Fortran, housekeeping includes several levels of allocation and program management.

MR. C. STRACHEY: There is one point which I think might be emphasised to people in this country anyhow about automatic coding systems. That is that the feasibility of an automatic coding system depends on having a very large single level store and that all the viable automatic coding systems have, if not very large, at least fairly large single level store.

The problem of writing a satisfactory automatic coding system for a machine which essentially has a two level store is a very difficult one. The difficulty is that in order to make use of the full speed of the machine, it is necessary to take account of the structure of the store in the strategy of the programming. If this step is left to the programmer, I do not think we can be said to have a satisfactory automatic coding system as it is still necessary to know a great deal about the machine in order to programme efficiently. It has not proved possible to perform this strategic reorganisation of the program automatically, and it does not seem to be the sort of problem which is yet amenable to a mechanised solution. The only remaining possibility is to treat the machine as if it only had a single level store, and this generally involves slowing the machine down by such a large factor that the automatic coding system loses much of its appeal except for very small or occasional problems.

MISS A. D. WOOLNER: I think that I must disagree with the last speaker about the problem of two level storage; it does exist in this country, but as Dr. Hopper has said, if one can make rules for solving a problem a computer can handle it.

The first stage in compiling a program for a two level store machine is to build a skeleton of instructions for the whole program, and at the same time to make an index of the cross references between instructions and data. This index can then be used to show which data and sections of program are required most frequently, and this allows storage to be allocated

in the quick access and backing stores according to priorities indicated by the index.

It should be possible for an autocode to produce a more efficient program than a human programmer, because it is very difficult for an average programmer to arrange the timing of instructions for operating the multiplicity of auxiliary equipment in a complex computer system to the best advantage. This is particularly true when magnetic tape equipment is involved and large tape files are being processed.

MR. J. W. BACKUS (in reply):ϕ With regard to the remarks of Dr. Garwick, while I like the notation that he presented, in the case of the first example he stated was rather a declarative one in directions of what to do, it seemed, and secondly, its use in imperative statements, an example of which he gave, I think that this simply illustrates one basic thing about languages for studying procedures, that you can have greater preciseness at the cost of freedom of expression and in that particular example it is not perhaps necessary that one would like the range of the variables over the entire dimensions of the various arrays referred to.

I think it has been pointed out adequately that variables in Fortran need not be floating point. It is only that the operations which are split in arithmetical expressions usually refer to floating point operations but if you have functions and routines which deal with different types of functions you are perfectly at liberty to treat them as such.

With regard to the remarks of Mr. Newman I think I will have to pass, showing how these various things, seven points which he listed, how the benefits are distributed over the seven points. I think he omitted a few like the ability to create routines and so on.

Again with regard to his second point I just simply have to state that I believe that mechanical processing is cheaper than human processing if you do it properly.

Then, Dr. McCarthy made some remarks about geometry programme and I want to thank him for clarifying that. I did mistate the question about the geometry programming as far as expressions. I am glad that he pointed out that. There were many expressions involving functions of functions of functions of various things that brought into play some features that I had not mentioned.

Mr. Brooker brought up a question about point 4 on page 242 and I think the point raised is mentioned here. The machine-language calling sequences resulting from input language references to functions and so on, in a source language programme, are standard type calling sequences. They could not be used to refer to programmes that are not generated by the system but

---

ϕ Edited version by the author was not available.

rather written by hand and these calling sequences are standard Wells, Wheeler and Gill calling sequences described in the book by those gentlemen.

In regard to the question of automatic programming for machines with two level stores and small stores, I only want to mention that we have some hope that something can be done for we do regard this as a very, very difficult part of it. We have now a project the goal of which is to take a Fortran programme which mentions no auxiliary store and produce a programme which does mention auxiliary store in such a way that the resulting Fortran programme when compiled would operate a machine with a small store and do so rather efficiently.

SESSION 2

PAPER 4

THE WORK OF THE COMPUTING CENTRE OF THE ACADEMY OF SCIENCES OF THE USSR IN THE FIELD OF AUTOMATIC PROGRAMMING

by

DR. A. P. ERSHOV

THE WORK OF THE COMPUTING CENTRE OF THE ACADEMY OF SCIENCES OF THE USSR IN THE FIELD OF AUTOMATIC PROGRAMMING

by

Dr. A. P. ERSHOV

1. PRELIMINARY RESEARCHES

THE real beginning of automatic programming development in the Computing Centre (as in the rest of the Soviet Union) was in the year 1954. In that year the first preliminary works appeared which contained the main ideas and their realization which I shall discuss first.

The main necessity, which arose in programming from the very beginning was the separation of the programming process into two main stages. In the first stage the programmer wrote a detailed computing plan for solving the problem. In the second stage the direct programming (coding in your terminology) was carried out, in which the detailed computing plan was rewritten in the computer's language.

Thus, in the course of programming, the necessity of constructing an intermediate language arose. This language is the language, by means of which the results of the first stage of programming (the detailed computing plan of problem solving) are written down.

In the beginning, im the Soviet Union as in other countries such an intermediate language was the flow-charts language, which was used by John von Neumann and Goldstine.

But in the years 1952-53 Prof. A. A. Ljapunov developed a new programming method, which was first described in his lectures at the Moscow University *(ref. 1)*. This method which had a great influence on the development of programming in the USSR, is now well known and called *the operator programming method.*

The main conception of the operator method is the consideration of the program as a complex operator which acts on the initial data and consists of elementary operators of various types. The main types of elementary operators, acting on the initial data, are *arithmetical operators* which fulfil the direct transformation of the data, and *logical operators* which determine the next direction of computation.

The conception of the dependence of the operator on some integer parameter (for instance the dependence of the address of the indexed number on the values of these indexes) was also very important.

Besides this, *control operators* were considered. These operators do not act on the initial data but deal with the program itself. I shall mention first the *re-addressing operators* which change the addresses which depend on parameters when the values of the parameters change.

The *restoration operator* restores the operator which depends on the parameter to its initial value according to the initial value of the parameter.

The record of the detailed computing plan written as the sequence of operators of various types is called the *operator scheme* of the program (*or the logical scheme* of the program).

This conception was first briefly described by Kitov *(ref.2)*, p.193.

Ljapunov's operator method was very important for the development of automatic programming for four main reasons:

1. The components of the program (elementary operators) were formally classified for the first time according to their function, independent of the type of problem programmed. This made it easier to formalize a language of logical schemes as a universal language for writing problems.
2. The language of logical schemes of programs was easy to algebraize. This was very important for the direct input of the logical scheme of the program into the computer.
3. The logical scheme of a program as an intermediate result of the programmer's work, successfully divides up the programming process. The "intellectual" work of the programmer is in composing the logical scheme while the most laborious and mechanical work is in the transition from the logical scheme to the object program. Thus the stage of the programming process, which is easiest to automatize, was separated.
4. The partition of the whole problem into a sequence of separate operators permits the programmer to divide his programming process in the same way. This also makes the coding easier.

Also in 1953, Prof. Ljapunov proposed that some parts of the second stage of programming could be automatized.

The first work in the field of automatic programming in the Soviet Union was carried out in the Mathematical Institute of the AS of the USSR and in the Computing Centre of the AS of the USSR, which at that time was a division of the Institute of Exact Mechanics and Computing Technique of the AS of the USSR.

In summer 1954 in the Computing Center V. M. Kurochkin composed a program for the BESM computer, which made symbolic programming possible. L. N. Korolev composed the first version of the program for BESM, which carried out the programming of arithmetical formulae.

G. S. Bagrinovskaja from the Mathematical Institute described a project which achieved the programming of re-addressing operators. In the summer of 1954 E. Z. Ljubimsky and S. S. Kamynin constructed the programming program (PP-1) for the STRELA computer, which performed the programming of re-addressing, restoring and logical operators.

All these works were only experimental but it was a necessary stage on the way to the design of big programming programs.

2. THE FIRST PROGRAMMING PROGRAMS

The first project of the programming program (PP-2), which fully automatizes the transition from the logical scheme to the object program, was described by E. Z. Ljubimsky and S. S. Kamynin in Prof. Ljapunov's seminar in October 1954 *(ref. 3)*.

Source information for this PP consists of the logical scheme of the program and the table of the storage allocation. The logical scheme may include arithmetical and logical operators, re-addressing and restoring operators and so-called non-standard operators, which represent some parts of the object program written in computer instructions with symbolic addresses.

In February 1955 a group of the Mathematical Institute staff-workers constructed a working PP for the STRELA computer on the basis of this project. This PP was described in a series of articles *(ref. 4)*.

In December 1954 a project for another PP was carried out. Some versions of the PP were developed on the basis of this project; the final one was constructed in March 1956. This PP is described in my book *(ref. 5)*.

I want to point out the main differences between the PP for BESM and the PP-2 for STRELA.

(a) In the PP for BESM the hierarchy of arithmetical operations when programming arithmetical formulae is taken into consideration.

(b) In PP for BESM the logical operators used are of a different type from those in the PP-2.

(c) There are no re-addressing or restoring operators in the source information for the PP for BESM, but some special notations are provided for specifying cyclic repetition of some groups of operators.

(d) In the PP for BESM the storage is allocated automatically.

The first public communications about the first programming programs were made by the authors in their reports on the All-Union Conference "The course of development of Soviet mathematical computing machinery" held in Moscow on the 12-17th of March, 1956.

3. THE PROGRAMMING PROGRAM FOR THE STRELA-3 COMPUTER

The main result of the Computing Centre's work in the field of automatic programming is the programming program for the STRELA-3 Computer. This computer is installed in the Computing Centre.

In the autumn of 1956, work began on the construction of the programming program for the STRELA-3 (PPS). It was carried out by staff-workers of the theoretical programming department, and finished in the autumn of 1957.

When the requirements for the source information were developed we had the following aims:

(a) The representation of the source information should be close to the mathematical formulation of the problem.

(b) The size of the auxiliary and technical work, not connected with the mathematical formulation, should be reduced to a minimum.

(c) The source information must give full information about the structure of the object program.

(d) The source information should be as compact as possible, yet easily understandable.

The source information prepared by the programmer consists of five parts:

1. The scheme of the program (the main part of the information)
2. The operators to be transferred to the second part of the source information.
3. Information about variables.
4. Information about storage blocks.
5. Blocks.

Some parts of the source information may be absent.

A little about terminology. *A variable* is a letter symbol which is contained in the source information. It may denote some mathematical variable or constant or, in general, any machine word. *A storage block* means any group of storage locations which have to be consecutive.

In the 3rd part all the variables, which require some additional information, are written down. Such additional information may be:

(a) A constant, if the variable represents a constant. Four types of constants may exist: decimal floating-point constant, binary floating-point constant, instruction with symbolic addresses and instruction with real addresses.

(b) An address, if required, to prescribe the location of the variable in the storage.

The 5th part includes any data arranged in blocks and introduced into the storage by the object program itself.

The scheme of the program. In the scheme of the program all the operators of the object program are written down one after another. The order in which the operators are placed in the scheme corresponds to the order in which

they are placed in the object program. The scheme may include operators of the following types:

Name of the operator	*Symbol denoting the type of operator*
1. Arithmetical	А
2. Logical	Л
3. Restoring	В
4. Non-standard	Н
5. Re-addressing	П
6. Repeat and check	ДС

As a rule, every operator is represented by a letter denoting its type, after which follows (in parenthesis) the information about the operator. The letter symbol and parenthesis are omitted for the arithmetical operators only.

It is possible to specify in the scheme of the program the cyclic repetition of some sequence of operators. For this purpose such a sequence is put into a brace: { denotes the opening of the *cycle* or *loop*, and } the close of the cycle.

Beside this, special symbols may appear in the source information; they denote the instructions which transmit information between parts of the computer.

Every operator may possess a number, which is inserted as a subscript to the symbol denoting the type of operator.

Every operator can be transferred from the scheme of the program into the second part of the source information. All the non-standard operators and some complicated arithmetical ones are transferred to the second part. In this case in the place of the operator transferred only a letter denoting the type with the number is left in the scheme.

The main feature which makes the source information easily understandable is the successful solution of the problem of recording the arithmetical formulae while retaining use of the obvious symbolism.

Arithmetical operators. Arithmetical operators are used in sequences of formulae such as

$$F(x_1,\ldots,x_n) \Longrightarrow y,$$

where the function F is expressed in terms of operations from some fixed list of operations and $x_1,\ldots,x_n$ are variables and constants. In particular, this list contains all the operations of the STRELA-3. y denotes the result of evaluating the formula. The constants in formulae are represented either

by letters or by numbers. The variables may have any number of letter subscripts. In this case every subscript must be a parameter of some loop.

I shall give (with some abbreviations) the list of operations used in the arithmetical operators.

List of operators

Name	Symbol
operations with two arguments	
addition	+
subtraction	-
addition modulo 2 (digit by digit)	$\dot{+}$
logical addition (digit by digit)	υ
multiplication	x or $\cdot$
logical multiplication (digit by digit)	$\wedge$
operations with one argument	
integer part	E
fractional part	D
absolute value	Mod
sign	sign
sine	sin
cosine	cos
arcsine	arcsin
arccosine	arccos
arctangent	arctg
exponent	exp
natural logarithm	ln
binary to decimal conversion	Dec
decimal to binary conversion	Bin
logical negation (digit by digit)	Π

powers:

$$x^{-1},\ x^{2},\ x^{3},\ x^{\frac{1}{2}},\ x^{\frac{1}{2}},\ x^{\frac{+p}{2q}} \quad (1 \leq p \leq 511,\ 0 \leq q \leq 7)$$

The usual hierarchy of operations useful in mathematics is used for determining the order in which the operations are performed. There is only one difference between the marked and non-marked multiplication ($\times$ and $\cdot$), which is obvious from the following example:

$$\underline{\operatorname{tg}\cos 2x^2} + \sin\underline{(a+b)(a-b)} + \sin\underline{(a+b)} \times (a-b) -$$

$$-\operatorname{Mod}\underline{y^{-i}}\ \operatorname{sign}\underline{(x+y)}\operatorname{tg}\underline{a^2}\ .$$

In this example the terms, which are arguments of the corresponding operations, are underlined. Besides this, the following is allowed: $\sin^m x$, which denotes $(\sin x)^m$.

Our experience shows that the only transformations of formulae which need be carried out in the preparation of source information are:- the omitting of radixes, and lines denoting the division, and replacing them by appropriate powers.

Logical operators. One logical operator is used for verifying one of five standard logical relations. The standard logical relations are the following:

$$a < b$$
$$a \leq b$$
$$|a| < |b|$$
$$|a| \leq |b|$$
$$a = b\ ,$$

where a and b are variables or constants.

In the scheme of the program logical operators are represented as

$$\prod\left(a \sim b \quad \begin{matrix} N_1 \\ \square \\ N_2 \end{matrix}\right)$$

where N_1 and N_2 are the numbers of the operators and $a \sim b$ is one of the five logical relations. This operator acts as follows: if the relation $a \sim b$ is true then we transfer to the operator N_1, otherwise we transfer to the operator N_2. If one of the numbers of the operators is omitted, this means that we transfer to the operator next after the logical operator.

Non-standard operators. A non-standard operator is any part of the object program written in computer instructions with symbolic addresses. There are many rules of writing down the non-standard operators but they are of no interest, and I shall omit them.

As mentioned above, the non-standard operator is always transferred to the second part of the source information (transferred operators). In the scheme of the program the non-standard operator is represented by a letter H with a number. In the second part a "head" is placed before the non-standard operators, which has a form

N	n		50

where N is the number of the non-standard operator and n is the number of words the operator contains.

The loops. A loop denotes the repeated execution of some sequence of operators for all given values of the *parameter* of the loop.

In the scheme of the program the loops are represented in the form

$$\{A\}$$

where {and} are the opening and closing braces of the loop and A is the sequence of operators to be repeated. The parameter i of the loop and its initial value i_H, if it is not equal to zero, are placed under the opening brace:

$$\underset{i=i_H}{\{} \quad \text{or} \quad \underset{i}{\{} \qquad (\text{if } i_H = 0)$$

If the number of repetitions of the loop is determined by the final value of the parameter i_k then the latter is placed over the opening brace:

$$\overset{i_k}{\underset{i=i_H}{\{}} \quad \text{or} \quad \overset{i_k}{\underset{i}{\{}}$$

In such a case the loop will be repeated from the value i_H up to the value i_k inclusive by steps of the size 1 ($i_H \leqslant i_k$).

i_H and i_k are either variables or non-negative integer constants. If i_H and i_K are variables, then it is supposed that they are evaluated by the object program as non-negative integer normalized numbers before the loop is reached.

It is also possible to set the number of repetitions of the loop by means of one of the five logical relations $a \sim b$.

Re-addressing operators. A re-addressing operator is recorded in the scheme of the program as

$$(N,\ n,\ h),$$

where N is the number of an operator; n is either a number of an instruction of a non-standard operator or a variable; h is either a variable or an integer constant.

The re-addressing operator is carried out thus:

(a) If n is a number of an instruction of the non-standard operator N, then the re-addressing constant h is added to the instruction.

(b) If n is a variable and h is a constant, then h is added to all the addresses of the variable n in the operator N.

(c) If n is a variable and h is also a variable, then it is supposed that the value of h needed is evaluated in the third address of the location for h as an integer number.

Restoring operators. A restoring operator is recorded in the scheme of the program as

$$B(N)$$

where N is the number of an operator.

If there are some variable instructions in the operator N which are transformed by re-addressing operators, the restoring operator will transfer the initial values of these variable instructions into their places in the operator N.

The repeat-and-check operator. While making up the program with the help of the PPS we may prescribe control of the computation by repeating and checking certain parts of the program. For this purpose symbols denoting repeat and check are written in the scheme of the program. The computer will obey every part of the program between two such symbols (following one after another) and accumulate all the contents of the internal storage. If the two sums are equal, then the whole content of the storage is recorded on the magnetic tape, and the computation is carried on further. If the control sums are not equal, the computer stops and, after a new start, repeats the computation once more and compares the three sums.

Special symbols for transfer of information between parts of the computer. The following notation, which denotes the transmission of information between the different parts of the computer, is used:

$$A \overset{n}{\rightarrow} B$$

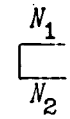
where A is the symbolic address of the beginning of the place from which information is going to be transferred, B is the symbolic address of the beginning of the place to which it is going to be transferred and n is the number of words transferred.

Unconditional transfer is denoted by



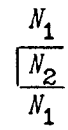
where N is the number of an operator to which we shall transfer.

Conditional transfer is denoted by



If the signal ω for the preceeding instruction is equal to 1, we shall transfer to the operator N_1. Otherwise we shall transfer to the operator N_2.

For transfer to subroutines the following notation is used:



which means: carry out the subroutine beginning from the operator N_1 and up to but not including the operator N_2.

Information about storage blocks. This information, at first, shows how many blocks there have to be in the storage and how long they should be. If required, it is possible to specify the position of some blocks in the storage. Such information about every block has the following form:-

$$M_k,\ \lambda \text{ words},\ \nu = \ldots$$

where k is the number of the block, λ is its length and ν is the address of the first location of the block (if it is given).

Then the information about the variables concerned in this block is written down.

If the address of a variable a or $a_{ij}\ldots$ depends on parameters $i,j,\ldots$, it is recorded in the form

$$a(\text{or } a_{ij}\ldots) = H(\text{or } K) + \Delta + h_1 i + h_2 j + \ldots$$

where H denotes the beginning of the block (its first location) and K denotes the end of the block (its last location). Number Δ is a "shift" in our terminology, h_1, h_2,... are the steps of re-addressing according to parameters $i,j,\ldots$. Δ and h may be either integer constants or variables.

If the address of the variable a from the block does not depend on any parameter then the information about its place in the block is recorded by

$$a = H \text{ (or } K) + \Delta.$$

If some variable appears in the scheme of the program with different sets of equal number of subscripts then the dependence of the address of the variable on the subscripts is mentioned in the information only once. For example, if the variables a_{isk}, a_{kli}, a_{ksl} appear in the scheme of the program, then in the information about blocks only one row is written down

$$a_{ijk} = H(\text{or } K) + h_1 i + h_2 j + h_3 k$$

That is all about source information.

I shall briefly describe the structure of PPS. The PPS as a whole consists of 30 separate blocks recorded on the magnetic tape. Each block is used only once during the running of the programming program. The information transformed is placed in the internal storage. Every block has the length of about 240 locations. A part of the storage (about 130 locations) is intended for some working data, which the PPS works out and passes from one block to the next. Thus the length of the simultaneously transformed information may be up to 1500 words. The blocks of the PPS change one after another automatically. The PPS punches on punched cards the completed object program and some additional information. Besides this, it is possible (if desired) to punch some intermediate information.

PPS is able to discover automatically many formal mistakes in the source information.

In conclusion of this part let us consider an example of preparation of the source information.

Suppose we have to integrate a parabolic partial differential equation

$$\frac{\partial z}{\partial t} = 0{,}75 \sqrt{x(1-x)\,(t^2 x+z)}\,\frac{\partial^2 z}{\partial x^2}$$

$$z(x,0) = 0;\quad z(0,t) = 0;\quad z(1,t) = t$$

from $t = 0$ to $t = T$, where $M = h^{-1}$ and τ are given (h is the step of x, τ is the step of t).

The evaluation formulae are the following

$$z_{0,n+1} = 0$$

$$z_{m,n+1} = z_{m,n} + \tau \cdot 0{,}75 \sqrt{x(1-x)\,(t^2 x + z_{m,n})}\;\frac{z_{m+1,n} - 2z_{mn} + z_{m-1,n}}{h^2}$$

$$m = 1, 2, \ldots, M-1$$

$$z_{M,n+1} = (n+1)\tau$$

$$n = 0, 1, 2, \ldots$$

Let us pick out one block of the storage, containing 1000 numbers for storing all the values of z for each layer of τ. The new z we shall store in place of the old ones with a delay of one step, because the old values of z will be needed for the evaluation of the new ones. Every time three points z_m, $z_{\overline{m}}$ and z_m^+ will be used, which correspond to $z_{m,n}$, $z_{m-1,n}$, $z_{m+1,n}$.

The source information is thus:

1. The scheme of the program is:-

$$\rightarrow M \rightarrow \tau \quad \Pi c(5) \quad \left\{ \underset{m=0}{\overset{M}{}} \; 0 \Rightarrow z_m \right\} \left\{ \underset{n}{} \; n\tau \Rightarrow t \right. \quad \left\{ \underset{m}{\overset{M}{}} \Pi (m = 0 \sqsubset \underset{N_1}{}) \right.$$

$$0 \Rightarrow r \sqsubset \underset{N_2}{\overset{N_2}{}} \quad \Pi_{N_1} (m = M \sqsubset \underset{N_3}{}) \; (n+1)\, \tau \rightarrow z_m \sqsubset \underset{N_4}{\overset{N_4}{}} \; \sqsupset \underset{N_3}{} \; mM^{-1} \Rightarrow x$$

$$z_m + 0{,}75(x(1-x)(t^2 x + z_m))^{\frac{1}{2}} (\overset{+}{z_m} - 2z_m + \overset{-}{z_m}) M^2 \Rightarrow r$$

$$t<T \; M$$

$$\sqsupset \underset{N_4}{} \; s \Rightarrow \overset{-}{z_m} \quad \sqsupset \underset{N_2}{} \; r \Rightarrow s \Big\} \Pi C \Big\} \quad \left\{ \underset{i=1}{} \text{Dec } z_i \Rightarrow z_i \right\} \; a \overset{1000}{\Rightarrow} \text{STOP.}$$

The logical operators here pick out the cases $m = 0$ and $m = M$, which are evaluated separately. The sign $\sqsupset$ is used to show the number of the next operator, if the latter does not possess a letter denoting its type.

2. Transferred operators are absent.
3. Information about variables is absent.
4. Information about blocks of storage is as follows:

M_1 , 1000 words

$z_m = H + 1 \cdot m$

$\overset{-}{z_m} = H - 1 + 1 \cdot m$

$\overset{+}{z_m} + H + 1 + 1 \cdot m$

$a = H.$

5. Blocks of storage are absent.

4. SOME PROBLEMS OF AUTOMATIC PROGRAMMING

The problems I am going to discuss are only a small part of the problem of automatic programming, but they characterize our interests clearly enough. I shall consider two types of problems:

(A) The problems of the source information and

(B) The problems of constructing new programming algorithms.

A. THE PROBLEMS OF THE SOURCE INFORMATION, This name covers all the problems whose solution simplifies the construction of the source information and approximates its form to the mathematical form of the problem, using all the richness of modern mathematical symbolism.

The natural first step in this direction is the expansion of the symbolism which is admissible in the source information, by adding some widespread mathematical terminology. The latter may include some sorts of mathematical operators, such as summation, multiplication, differential operators, and so on, or some functionals, such as min $f(x)$, max $f(x)$, $\int_a^b f(x)dx$ and so on, and some widespread special functions. Lastly, we can include in the source information not only real scalar numbers but more complicated things such as complex numbers, vectors or matrices and so on.

The method of subschemes (already used in some programming programs) may be used for decoding new symbols which appear in the source information. The essence of this method is as follows. A special block is added to the PP which may "understand" new symbols in the source information. For every symbol of this sort, which denotes some mathematical algorithm D, this block constructs a subscheme which represents the record of the algorithm. This record is the algorithm D in terms of standard operators (arithmetical, logical and so on). The subscheme constructed is substituted for the symbol, denoting the algorithm D which appears in the source information. After this the transformed scheme of the program, in which all the special symbols have been replaced is transformed by PP in the usual way.

Briefly this is the unification of the PP programming method with the method of compiling programs. Though the possibility of such a unification was discussed long ago, little has been done in this direction.

B. THE PROBLEMS OF CONSTRUCTING NEW PROGRAMMING ALGORITHMS

From the broad field of problems of that kind I shall consider only two, perhaps the most important from the practical view.

1. *The analysis and transformations of the scheme of the program*

The term "analysis of the scheme of the program" denotes the construction of algorithms which are able to find out from the scheme of the program the interrelations between the operators forming the scheme.

Previously programming programs were constructed in such a way that every operator from the scheme was programmed as far as possible independently of the others. For example, the economy of instructions is achieved only in one operator during the programming of arithmetical operators. In programming loops all variable instructions which depend on the parameter of the loop must be placed in the scheme between the first operator and the last one of the loop. The independent programming of various operators

allows us to simplify the programming algorithms but excludes some opportunities for improving the program constructed.

I shall give an example of the improvement of the programming algorithms requiring an analysis of the scheme of the program.

Let us suppose that the same term $F(x_1, \ldots, x_k)$ appears in each of the arithmetical operators $A_1, \ldots, A_n$. It is obvious that, in some cases, $F(x_1,\ldots,x_k)$ may be programmed in one operator only, while in the others the term $F(x_1,\ldots,x_k)$ is replaced by the symbol of the result of evaluating the term. This may be done in the case where there is an operator A_i among $A_1,\ldots,A_n$, which is always carried out before the others and if $x_1,\ldots,x_k$ do not appear as the results in the computation between the use of the operator A_i and A_j (where A_j is one of $A_1,\ldots,A_n$). It is clear that it is impossible to discover the possibility of such a generalized economy of instructions without analysing the scheme of the program.

The problem of analysing the scheme of the program may be formulated in a particular case more exactly as follows:-

To construct an algorithm, which for any two arbitrarily chosen operators A_i and A_j from the scheme, determines which of the following possibilities takes place

(a) A_i is always carried out before A_j;

(b) A_j is always carried out before A_i;

(c) A_i and A_j are exclusive, that is, if one is fulfilled, the other is not;

(d) A_i and A_j can be carried out in any relative order.

The construction of such an algorithm would be very important for programming.

In connection with the inclusion of such notations as the summation term Σ and the multiplication term Π in the source information, there is an interesting problem of transforming the schemeoof the program.

Let us consider as an example the problem of evaluating the series

$$u = \sum_{k=0}^{n} f(k)$$

with some given accuracy. It is obvious that if we replace this notation directly by the subscheme

$$0 \Rightarrow u \left\{ \begin{matrix} n \\ k \end{matrix} \right. \quad u + f(k) \Rightarrow u \Big\},$$

the object program would be very bad. If we compute according to this program, we have to evaluate $f(k)$ for every k anew, and this is extremely unprofitable in most cases. In fact, we usually replace the direct evaluation of $f(k)$ by its evaluation by a recurrence relation.

For instance, the series

$$u = \sum_{k=0}^{n} \frac{x^k}{n!}$$

is always evaluated according to a scheme such as

$$1 \Rightarrow t,\ 0 \Rightarrow u \left\{ \begin{matrix} n \\ \\ k=1 \end{matrix} \right. \left\{ u + t \Rightarrow u,\ \frac{t \cdot x}{k} \Rightarrow t \right\} .$$

It seems possible to use to solve the following problem. For any arbitrary function $f(k)$ of an integer argument, represented by a combination of the functions k^n, n^k, $k!$ to find an algorithm which replaces the closed notation $f(k)$ by recurrence relations. The realization of such an algorithm in PP would save the programmer having to do the preliminary work of finding the recurrence relations; instead this work would be done by the computer. The programmer has only to write down the common member of the series or the multiplication in a closed form.

In connection with the programming of loops the following problem also concerning the transformation of the scheme of the program is interesting. In many problems the following appears:

"Evaluate the table of the values of the function

$$z = F(p, l, m, n)$$

(where p is a parameter) for all sets of values of the integer arguments, for which

$$l_0 \leqslant l \leqslant l_k,\ m_0 \leqslant m \leqslant m_k,\ n_0 \leqslant n \leqslant n_k."$$

(the number of arguments is of no importance and can be arbitrary).

It is very easy to write the scheme of the program in the form:

$$\left\{ \begin{matrix} l_k \\ \\ l=l_0 \end{matrix} \right. \left\{ \begin{matrix} m_k \\ \\ m=m_0 \end{matrix} \right. \left\{ \begin{matrix} n_k \\ \\ n=n_0 \end{matrix} \right. F(p, l, m, n) \Rightarrow z,\ \text{punch } z \right\} \Big\} \Big\}$$

But if, for instance, $F(p, l, m, n,)$ can be represented thus:

$$F(p, l, m, n) = \varphi(l, \psi(m, \xi(n, \zeta(p)))),$$

then the scheme

$$\zeta(p) \Rightarrow u \left\{ \begin{matrix} n_k \\ \\ n=n_0 \end{matrix} \right. \xi(n, u) \Rightarrow v \left\{ \begin{matrix} m_k \\ \\ m=m_0 \end{matrix} \right. \psi(m, v) \Rightarrow w \left\{ \begin{matrix} l_k \\ \\ l=l_0 \end{matrix} \right. \varphi(l, w) \Rightarrow z,\ \text{punch } z \Big\} \Big\} \Big\}$$

would be, of course, much more profitable. That is because, according to this scheme, in every loop only that part of the function F is being evaluated which is essentially dependent on the parameter of this loop. The evaluation of the left part of F is carried out in the external loops, which are used less often. Therefore the program as a whole works faster.

In this case it is also probably possible to find an algorithm which would perform the transformation of the scheme of the program mentioned above.

2. *Increasing the speed of the PP* The new PP contains more and more opportunities to improve the quality of the object programs by increasing the complexity of the programming algorithms. But the increased complication and the size of PP have their disadvantages in the increase of running time of PP. When the running of PP on the computer takes a long time, it reduces the effect of using PP (especially when the computer is not working well). Therefore I would like to stress the fact that it is necessary, while constructing new algorithms of the programming, to aspire in every possible way to increase the speed of the algorithm's work. One of my works *(ref. 6)* testifies some possibilities in this direction.

REFERENCES

1. LJAPUNOV, A. A. "O Logičeskikh skhemakh program" (On logical schemes of programs). *Problemy kibernetiki, Vypusk 1 (Cybernetics problems, issue 1). Moskva, Fiziko-matematiceskoje izdatelstvo,* 1958.
2. KITOV, A. I. "Elektronnye cifrovye mashyny" (Electronic digital computers), *Moskva, Sovetskoe radio,* 1956.
3. KAMININ, S. S., LJUBIMSKIY, E. Z., SHURA-BURA, M. R. "Avtomatizacija programmirovanija c pomošču programmirujuščej programmy" (Automatization of programming by means of a programming program). *Problemy kibernetiki Vypusk 1.*
4. LJUBIMSKIY, E. Z. "Arifmeticeskij Blok v PP-2" (The arithmetical block in the PP-2)
   KAMININ, S. S. "Blok pereadresacii v programme PP-2" (The readdressing block in the PP-2 program).
   LUKHOVICKAJA, E. S. "Blok obrabotki logičeskikh usloviy v PP-2" (The block of the treatment of logical conditions in the PP-2)
   SHTARKMAN, V. S. "Blok ekonimii rabočikh jačeek v PP-2" (The block of the economy of temporary locations in PP-2). *Problemy kibernetiki, Vypusk 1.*
5. ERSHOV, A. P. "Programmirujuščaja programma dlja bystrodejstvujuscej elektronnoj sčetnoj mashyny" (The programming program for the BESM). *Moskva, Izdatelstvo Akademii Nauk,* 1958.
6. ERSHOV, A. P. "O programmirovanii arifmetičeskikh operatorov" (On programming of arithmetical operators), *Doklady Akademii Nauk SSSR,* 1958, 118, No.3, 427. There is a translation: *Communications of the ACM,* 1958, 1 No.8.

(94009)

DISCUSSION ON THE PAPER BY DR. A. P. ERSHOV

MR. C. STRACHEY: I should like to ask Dr. Ershov one small point on his very interesting paper. I think he mentioned cycles; I should like to ask whether his machine has got what we call modifier registers, index registers, or B lines for the automatic modification of instructions?

DR. ERSHOV: No, "STRELA" doesn't have index registers.

DR. J. McCARTHY: Where does programming come in in the university curriculum?

DR. A. P. ERSHOV: Programming is carried out in 5 different universities, Moscow University, Kiev University, Leningrad University, Erivan University and Gorky University. Programming in all these universities.

DR. J. McCARTHY: What are the provisions for teaching programming to students - to how many students do you teach it and in what year?

DR. A. P. ERSHOV: I only know Moscow Institute myself. Moscow University turns out 250 to 300 young mathematicians, of which 100 usually specialise in computing mathematics, but the other mathematicians also do programming to a certain degree, although they have a wider syllabus, and there are a few other universities where they also do programming. In fact it is going to be extended to all the universities within a few years. Programming usually is given at the third year of learning.

DR. J. McCARTHY: Mathematical notation as it presently exists was developed to facilitate stating mathematical facts, i.e. making declarative sentences. A program gives a machine orders and hence is usually constructed out of imperative sentences. This suggests that it will be necessary to invent new notations for describing complicated procedures, and will not merely be able to take over intact the notations that mathematicians have used for making declarative sentences. Do you have any comments on what may be required in this direction?

DR. A. P. ERSHOV: If you deal with the usual kind of numerical analysis the problem is not very difficult because the imperatives needed in numerical analysis are rather simple. But for describing complicated non-arithmetical procedures it becomes a very important problem. I think that introducing new special notations has to take place in comparatively small quantities, but the main way is to describe the procedure in words.

But I have to say that this opinion arises more from the practical point of view than the theoretical one.

MR. M. WOODGER:

1. In the scheme of a program for the PP for STRELA-3, use is made of suffixes, superfixes, and other symbols occupying more than one line. How is this represented on the input medium to the computer?
2. If a program needs to use magnetic tape for the storage of intermediate results, how is this expressed in the scheme of the program?
3. Can you give examples of running time of PPS while producing object programs for useful work and the running times of these object programs, to show the proportion of computer time needed in typical cases?

DR. A. P. ERSHOV:

1. Before programming you must make up a numerical coding of the scheme of the programme (STRELA-3 does not have an alphabetical input). There is a special code for the constant symbols such as signs of operations, powers, parenthesis, braces etc. Every variable gets an arbitrarily choosen octal number from the interval 0100-1000 (conventional number in our terminology). If a variable appears as a subscript its conventional number is increased by 1000. The numerical coding is written into standard programming sheets for the STRELA-3.

For instance, the coding of the scheme of the programme of the example mentioned is ("N" denotes the conventional number of N):

A_1	A_2	A_3	Op
		"M"	" → "
		"τ"	" → "
	0005		" Дc "
"m"	"0"	"M"	" { "
"0"	" ⇒ "	"z"	01
"m"+1000			
			" } "
"n"			" { "
"n"	" τ "	" ⇒ "	01
"t"			
"m"		"M"	" { "
"m"	" = "	"0"	" Л "
	N1		20
"0"	" ⇒ "	"r"	01
N_2	N_2		20
N_1			50
"m"	"="	"M"	" Л "
	N_3		20
"("	"n"	"+"	01
"1"	")"	"τ"	
" ⇒ "	"z"	"m"+1000	
N_4	N_4		20
N_3			50
"m"	"M"	"pow-1"	01
" ⇒ "	"x"	"z"	

A_1	A_2	A_3	Op
"m"+100	"+"	"0.75"	
"("	"x"	"("	
"1"	"−"	"x"	
")"	"("	"t"	
"pow2"	"x"	"+"	
"z"	"m"+1000	")"	
")"	"pow1/2"	"("	
"z+"	"m"+1000	"−"	
"2"	"z"	"m"+1000	
"+"	"z−"	"m"+1000	
")"	"M"	"pow2"	
" ⇒ "	"r"		
N_4			50
"S"	" ⇒ "	"z−"	01
"m"+1000			
N_2			50
"r"	" ⇒ "	"s"	01
			" Дc "
"t"	" < "	"T"	" "
"i"	"1"	"M"	" "
"Dec"	"z"	"i"+1000	01
" ⇒ "	"z"	"i"+1000	
			" } "
"a"	1000		" → "
			"STOP"

end

Other parts of the source information are coded in a similar way. There are special auxiliary personnel for numerical coding.

2. Handling magnetic tape is expressed in terms of the notation for the transmission of information, $A \xrightarrow{n} B$. For instance, if you want to record an array of data from the internal storage on a magnetic tape you have to write

$$a \xrightarrow{n-1} N,$$

where a is the symbol of the first variable from the array, n is the length of the array and N is the number (address) of the block on the

magnetic tape. Reading from the magnetic tape is written in a similar way

$$N \xrightarrow{n-1} a.$$

3. The average speed of the PPS is the production of about 35 three-address instructions per minute. It is impossible to evaluate an average running time of object programs because this time essentially depends on the number of versions. We have solved various problems with running times from few minutes to several hours.

SESSION 2

PAPER 5

TIGRIS AND EUPHRATES - A COMPARISON BETWEEN HUMAN AND MACHINE TRANSLATION

by

R. H. RICHENS

BIOGRAPHICAL NOTE

R. H. Richens was born in Penge, near London, in 1919. He read natural sciences at Cambridge and is now Assistant Director of the Commonwealth Bureau of Plant Breeding and Genetics at Cambridge. He has been a member of the Cambridge Language Research Group since its foundation. His principal research interests have been the taxonomy and history of the elm, the history of Soviet genetics, and machine translation.

TIGRIS AND EUPHRATES - A COMPARISON BETWEEN HUMAN AND MACHINE TRANSLATION

by

R. H. RICHENS

SUMMARY

TRANSLATION is treated as a species of formal transformation of symbols. Everything symbolized by a set of symbols constitutes the domain of symbolization of the set. The ultimate elements of the domain which symbolize nothing further are designated the terminal indicatum. Most domains of symbolization comprise mediate symbols which are both symbolized by other symbols and themselves indicate further symbols. Mental concepts are treated as symbols.

In translation, a set of symbols is transformed to another set in another language, the two sets having terminal indicata that only differ within narrow limits. Different kinds of translation can be differentiated on the basis of the extent to which the mediate symbols of the domains of symbolization of the input and output passages are similar.

Human translators utilize many categories of symbols and display great flexibility in their choice of procedure for translating and in their choice of criteria for ascertaining the significance of symbols of multiple use. Machine translation (MT) is much less flexible though most of the symbols used by human translators, with the exception of auditory symbols, phonemes and uttered-word segments, have analogues in MT.

Different types of input passage require different translation procedures, in particular with reference to the relative roles played by syntactic and semantic analysis. Closer formal resemblances may occur between human translation and MT procedures for the same type of input than between the procedures of either the human translation or MT confronted with input passages of various types.

While the over-all performance of the human translator is unlikely to be approached in the near future by MT, the latter may accomplish certain individual translation operations more efficiently than its human counterpart, and, on occasion, produce a better translation.

MACHINE translation (MT) promises to provide a highly developed analogue to an intricate operation till recently an exclusive pursuit of human beings. It is not the object of this paper to discuss this analogy either in the light of psychology or the philosophy of language; this can be done effectively only by specialists in these fields. What will be attempted here is to compare human translation and MT in as far as they are formal transformations, meaning by the latter any operation which converts one set of symbols into another. It will not prove possible to avoid all philosophic problems. Indeed, MT appears to be a breeding ground for new problems and new viewpoints concerning the philosophy of language. Such philosophic issues will only be touched upon here when they are inescapable.

Any passage in any language can be treated as an ordered set of symbols. By order is meant significantly arranged.

LANGUAGE AND SYMBOLISM

A symbol is anything, or a part or aspect of anything, or several things, or several parts or aspects of a thing or things which, either alone or in conjunction with other symbols, indicates something, usually but not necessarily, other than itself.

Curiously enough, there is no term in general use in English for whatever a symbol indicates. To facilitate discussion, the term indicatum will be used here. This is to be understood extremely widely; it may be one or many, it may be an object, a part, aspect, activity or state of an object, it may be a relation between objects, it may be anything that Wittenstein (1922) would have designated as a 'state of affairs', it may be a mental image or concept, it may be a falsehood, a contradiction or a tautology, it may be a symbol or set of symbols, and so on. The dyadic relation associating a symbol with its indicatum will be designated indifferently either by the verb 'indicate' or by the verb 'yield'.

In discussing what a particular set of symbols indicates, it is useful to establish its domain of symbolization. Thus a set of visual symbols, A, may indicate a corresponding set of concepts; these may indicate further concepts and these finally indicate a woman reading a book. The visual symbols, A, the two sets of concepts and the woman, the book and the act of reading together constitute the domain of symbolization. If the book is a novel it will contain printed symbols of its own but these will be excluded from the domain of symbolization of A. Similarly, A may be indicated by some other set of symbols B, but B is likewise excluded from the domain of symbolization of A.

In any particular domain of symbolization, a symbol which is not an indicatum of some other symbol or symbols will be termed an initial symbol. Similarly, an indicatum which is not itself a symbol yielding some further

indicatum will be designated a terminal indicatum. Symbols which both yield further symbols and are the indicatum of others will be termed mediate symbols.

The symbols of natural languages are of two principal types, auditory and visual, corresponding to spoken and written language respectively. It is convenient to treat the concepts corresponding to spoken or written language, or in fact any concepts, as symbols too since they also yield indicata. It is obviously not possible in a written paper to introduce other than visual symbols. We can however, introduce written symbols to indicate auditory symbols or concepts. The latter are of various categories, many being mediate symbols representing further concepts. It is convenient to distinguish between different categories of symbols by prefixing an indicator in the form of a capital letter followed by a full stop.

The following categories will be used:-

A	auditory symbol category
V	visible symbol category
P	phonemic category
G	graphemic category
U	uttered-word category
W	written-word category
L	lexico-grammatic category
S	syntactic category
N	'naked idea' category

All but the first two of these symbolic categories are concepts. When necessary, the categories can be subdivided by subscripts, e.g. L_1, L_2.

The significance of these categories is as follows. A.u: stands for a particular utterance of a long u: sound. It approximates to the [u:] of descriptive linguists. V.c stands for a particular token mark interpreted as of type c.

The A and V categories are initial symbols. A problem, however is involved in deciding what is to be regarded as an initial symbol. The letters of the roman alphabet seem to be fairly obvious initial symbols but should German V.ü, for instance, be regarded an an intial symbol, or an ordered arrangement of V.u and V.¨? When V.a and V.e are conjoined in the digraph V.æ, there would be some advantage in considering V.a and V.e as the initial symbols in English since variants with the letter disjoined will be frequently encountered. However, in Danish the digraph is regarded as a separate letter and might well be best regarded as the initial symbol. The Chinese ideographs present the problem in an acute form. If each ideograph be regarded as an initial symbol, many thousands would have to be recognized. On the other hand, it is possible to regard the ideographs as composed of ordered arrangements of a very considerably smaller number of units. Pushing the matter to an extreme, all written symbols could be

regarded as composed of a single visible symbol, replicated in two dimensions to correspond to whatever symbol is being considered. It is clear that the decision as to what constitutes an initial symbol must depend on the use to which this category is to be applied. A different decision might well be made for different natural languages.

P.u: stands for the phoneme u:, the universal concept to which pertain all sounds regarded as of the type u:, prescinding thus from pitch, intensity and other minor differences in pronunciation. In descriptive linguistics P.u: is commonly represented as /u:/. There are various interpretations of the phoneme concept among descriptive linguists; the one adopted here takes phonetic similarity as the criterion of phonemic identity, the phoneme being regarded as the referend of a phonetic type. However, it is desirable in the present context to extend the phoneme concept to cover all phonological features to which the concept of auditory type is applicable. Thus the tonic accent in Spanish subsumes a range of minor variations in pitch and intensity and can therefore be treated as a concept of the P category. G.c stands for the universal concept to which pertain all letters regarded as of type c. It thus subsumes handwritten c, typewritten c, and c in standard-roman and modern-face founts.

A question sometimes arises as to which symbols are to be regarded as of the same type. In a book set up entirely in italics, symbols in this fount might well be regarded as of the same type as the corresponding symbols in a book in roman fount. When, however, a few italic words occur in a passage otherwise in roman fount, it might be advisable to regard corresponding italic and roman letters as of different types since the italic words represent different indicata than the same words in roman fount. There are thus no formal criteria by means of which token symbols can be adjudicated to be of the same or different types. The decision must be made in the light of the application to which the type concept is being put.

Each symbol in the A category indicates a corresponding symbol in the P category and similarly for the V and G categories. Moreover, most graphemes in the G category also indicate phonemes in the P category since most written languages attempt to mirror speech. Cases do occur where graphemes indicate concepts but not phonemes, for instance the MT interlinguas of Richens (1956b, 1958), but usually, even in languages like Chinese where the ideographic role of the symbols in manifest, a phoneme sequence is also represented. Punctuation, however, may have no phonemic indication, as, for example the hyphen in the place name Weston-super-Mare. It also happens that a grapheme, say G.M, indicates a phoneme, P.m, and also a conceptual category, L.capital, with no corresponding phoneme.

Neither P.u nor G.c yields any particular terminal indicatum, or rather each potentially, that is, in combinations with other symbols, yields almost any indicatum. When combined with other symbols, P and G symbols yield relatively precise terminal indicata. Rarely, a single P or G symbol

does yield a relatively precise terminal indicatum as the Italian plural definite article P.i or the Polish preposition G.w.

An ordered combination of phonemes may constitute an uniflected uttered-word or a segment of an inflected uttered-word. Thus, the English phonemic sequence

P.d P.o P.g P.z

represents the uttered-word symbols

U.dog U.z

the former representing the animal and the latter plurality. Similarly, the graphemes

G.u G.n G.v G.e G.i G.l G.e G.d

represent the written-word symbols

W.un W.veil W.ed

It is probable that everybody has a concept of a word or word segment at a higher level of generality than the foregoing, that is prescinding from either the phonemic or graphemic categories. We may call this general category, the lexico-grammatic category. In descriptive linguistic studies, the term lexico-grammatical is sometimes used for words or word segments that partake of both lexical and grammatical characteristics as traditionally conceived; it is thus in the nature of a logical product. As used here, lexico-grammatic is far more equivalent to the logical sum of these two categories; it corresponds in a number of respects to the morpheme concept. The written-word symbols

W.un W.veil W.ed

would then indicate the lexico-grammatic symbols

L.un L.veil L.ed

The lexico-grammatic category requires subdivision, to allow for indication of the various types of grammatic category encountered. Thus French L_1.chien indicates L_2.masculine gender. The Latin affix L_1.itur indicates

L_2.third person L_2.singular L_2.present tense L_2.passive

The syntactic category comprises the indicata of the ordering relationships indicated either by single lexico-grammatic symbols or more usually by L symbols in some particular arrangement. Identical superscripts are used to indicate which symbols are bonded to which, bond signifying any syntactic connexion, however neutral.

Thus Latin

L_1.in L_1.flagrant L_1.e L_1.delict L_1.o

indicates

L_2.ablative adjective L_2.ablative noun

which indicates

S.qualifying adjective^a S.qualified noun^a

English

L_1.he L_1.cough L_1.ed

indicates

L_2.pronoun L_2.nominative L_2.verb ,

of which

S.pronoun subject^a S.verb predicate^a

is an indicatum; and Latin

L_1.carp L_1.e L_1.di L_1.em

indicates

L_2.verb L_2.accusative noun ,

which indicates

S.verb^a S.noun object^a

Lastly, for our penultimate indicatum, we have a category of 'naked ideas'. These made a somewhat provocative entry into MT discussions at the symposium on machine translation held at King's College, Cambridge, in August 1955. No one supposes, of course, that a concept can be discussed except by using language, but it is surely apparent that there are concepts that constitute a unique indicatum of numerous diverse renderings both within a single language and in different languages. Thus, there is a unique indicatum yielded by all the English words L.give, L.gift, L.present. L.donate, L.donation, L.receive, and Latin L.do, L.dono, L.praesto, L.largior, which symbolizes a state of affairs in which somebody A causes another person or perhaps an animal B to enter into a proprietary relation to an object C. Such a generalized 'naked idea' we can express as N.give. It is important to note that whereas any particular terminal indicatum may be symbolized in many ways using L symbols, only one set of N symbols is appropriate.

The N category may be subdivided, if the naked idea is analysable further. On this matter circumspection is necessary. Not only can numerous analytic schemata be devised for any one concept but the extent to which analysis is pursued in any one analytic scheme is frequently unlimited. It is also well to bear in mind Wittgenstein's (1953a) warning that an analysed statement does not necessarily constitute an increase in understanding over the unanalysed statement. The paradox of analysis, that analysis must be either trivial or false, is also a warning against analysis without a particular application in mind. Analysis in the present context is indeed trivial in as far as the final indicatum of any domain of symbolization remains unaffected. It is not trivial in as far as new mediate N symbols

are derived, which by virtue of their relations with one another, play a useful role in translation.

Thus, we have seen that L.give, L.present, L.donate all yield the same naked idea N_1.give. This, however is analysable further to

$$N_2.\text{cause} \quad N_2.\text{pertain}$$

with suitable syntactic bonding, where cause and pertain are catchwords for the corresponding general concepts. It is possible to analyse further, but in the present context it is doubtful if this would serve a useful purpose.

It is not implied that a human being, confronted with a linguistic passage, reacts by producing a domain of symbolization of the sort described or that all or any of the symbols set down above are indispensible in human translation. What does seem reasonable to assert in that each of the categories of symbols described are employed in human translation in some cases. It has been pointed out often that natural languages are over-determined and use several symbols to indicate a single point. Overdetermination, however, is usually a general characteristic and not a particular characteristic of any group of symbols so that, whereas it might be possible to ignore a symbol as redundant in one passage, it might well prove indispensible in another.

To exemplify a domain of symbolization, we will take the short English sentence, 'She came to'. The token symbols of this sentence are

$$\text{V.S} \quad \text{V.h} \quad \text{V.e} \quad \text{V.c} \quad \text{V.a} \quad \text{V.m} \quad \text{V.e} \quad \text{V.t} \quad \text{V.o} \quad \text{V..}$$

and the corresponding graphemes

$$G_1.\text{S} \quad G_1.\text{h} \quad G_1.\text{e} \quad G_1.\text{c} \quad G_1.\text{a} \quad G_1.\text{m} \quad G_1.\text{e} \quad G_1.\text{t} \quad G_1.\text{o} \quad G_1..$$

Of these, the first G_1.S indicates the corresponding lower case letter G_2.s together with a lexico-grammatic symbol L_3.capital.

The gramphemes indicate a series of phonemes which one may set down roughly as

$$\text{P.}\check{\text{s}} \quad \text{P.i:} \quad \text{P.k} \quad \text{P.e:} \quad \text{P.m} \quad \text{P.t} \quad \text{P.u}$$

and these yield the uttered-words

$$\text{U.}\check{\text{s}}\text{i:} \quad \text{U.keim} \quad \text{U.tu}$$

In addition to indicating phonemes, the graphemes also indicate the written-words

$$\text{W.she} \quad \text{W.came} \quad \text{W.to}$$

and the U and the W symbols both yield the lexico-grammatic symbols

$$L_1.\text{she} \quad L_1.\text{came} \quad L_1.\text{to}$$

The L category requires subdivision. Thus L_1.came indicates a verb the root form of which can be expressed as L_2.come; it also yields a grammatic

indicatum which may be designated L_3.past tense. The principal grammatic indicata in the L_3. category are

L_3.capital L_3.pronoun L_3.singular L_3.nominative L_3.verb L_3.past tense
L_3.verbal postposition L_3.stop

The L_3. symbols have indicata in the S category, expressible as

S.pronoun subject^{c} S.verb governing postposition^{abc}
S.verb qualifier^{b} S.verbal postposition^{a}

And lastly, we can proceed to the naked ideas

N_1.3rd person^{cd} N_1.female^{c} N_1.one^{c} N_1.become^{a}
N_1.conscious^{abd} N_1.past time-tense^{b}

which can be analysed one step further to

N_2.x^{d} N_2.female^{cde} N_2.one^{c} N_2.become^{a} N_2.conscious^{able}
N_2.past^{b3} N_2.communication situation^{b2}

The term x^{d} stands for the antecedent of the pronoun; past time-tense indicates past relative to the time of the communication.

The foregoing scheme provides one out of many possible bases for a formal description of translation which will be the topic under discussion in the next section.

TRANSLATION AS A FORMAL TRANSFORMATION

Symbols can be classified in many ways. For present purposes we can regard the corpus of symbols used by a group of people within which communication is easy as a language. Minor differences in usage within such a group can result in local dialects or technical expressions. Most languages are auditory and have a corresponding language in written symbols. Languages can also be classified in respect of subject matter; in some respects scientific Japanese is closer to scientific English than to literary Japanese.

A passage need not consist of a single language. Greek is a regular component of Hisperic Latin, English of medieval Law Latin and English and German of scientific Japanese.

A language need not consist exclusively of symbols. It may contain, for instance, sounds devoid of significance but with a certain quasi-musical role. Derry derry down, for instance, may convey a vague sense of joviality in some contexts, in others it signifies practically nothing.

Transformation of a set of symbols in one language to another set in the same language, with little change in the terminal indicatum, we may regard as constituting paraphrase. If the transformation is to a different language,

we may regard this as translation. It is rare for a translation involving natural languages to yield exactly the same indicatum as the original passage since most languages have particular shades of meaning or particular vaguenesses that are not simply expressible in other languages. The amount of difference in the indicata of the original passage and its translation, and the nature of the differences in the two indicata that are tolerated in translation are a matter of choice. They might well differ according to the particular purpose for which the translation is being made.

Translation, moreover, may involve further considerations. The domain of symbolization of a particular set of initial symbols is usually highly elaborate. It may be required to effect a translation in which, not only the terminal indicatum, but various of the mediate symbols are either kept unchanged or altered to only a slight degree. Thus, whereas the terminal indicatum of a Latin sentence in the passive voice, i.e. with the symbol L_3.passive in its domain of symbolization, can be translated into an English sentence in the active voice without change in the terminal indicatum, it is often required to conserve such a symbol as L_3.passive during translation. Similarly the Italian noun

$$L_1.\text{prod} \quad L_1.\text{uzion} \quad L_1.\text{e}$$

can be translated into English as an abstract noun,

$$L_1.\text{produc} \quad L_1.\text{tion}$$

or as a gerund,

$$L_1.\text{produc} \quad L_1.\text{ing}$$

The terminal indicatum is unaffected but the former translation involves fewer changes in the mediate L_3. symbols in the domains of symbolization.

It is implied in the previous paragraph that a basis for comparison exists between different language in respect of their mediate symbols. Considerable progress has been made towards an exact description of the A, V, P, G, V, W, L and S categories of symbols within a single language; this in fact is the subject matter of descriptive linguistics. The techniques of investigation of descriptive linguistics, however, lean heavily on commutability relations which are practically nonexistent between different languages. Comparative grammar has been investigated but no sufficiently exact system for establishing formal comparisons between grammatic symbols in different languages has yet been developed. Allen (1953) has shown along what lines descriptive linguistics might be expected to contribute to formalized comparative grammar; Lambek (1958) has approached the problem from the mathematical angle.

Thus, there is obviously a close relationship between the English mediate symbol L_3.noun and the Italian L_3.noun. However, the former can be frequently used as an uninflected qualifier preceding another noun while the latter cannot. There are perhaps no exact grammatic equivalences

between different languages and, at the moment, we are hampered by lack of suitable techniques for characterizing what partial equivalences there are with sufficient precision.

Some types of translation aim at conserving the auditory symbols as far as possible, that is the symbols of category P. Verse translation is a related exercise; in this case it is not the phonemes that are conserved but a phonemic pattern, or what would more usually be described as a phonological pattern which, as the indicatum of the P (P_1) symbols, we may designate as category P_2. The preservation of alphabetic patterns in psalm translations and the preservation of acrostics are likewise instances in which P_2 symbols are as far as possible kept invariant.

It is possible to translate according to more complex criteria, for instance onomatopeic proportionality, in which highly onomatopeic words in the original passage are rendered by onomatopeic words in the translation and conversely for words with little onomatopeic suggestiveness. This procedure conserves particular relation between the P and N symbols. It is not necessary here to pursue such refinements further.

From what has been said above about translation, it would appear that the process of analytic understanding is itself a form of translation since it, like translation, involves transformation of symbols with invariance of the terminal indicatum. Each of the categories of symbols enumerated can in fact be regarded as a language on its own, so that, in as far as translation between natural languages involves mediate symbols, it can be regarded as a concatentation of subordinate translations. The notion that analysis is a form of translation has been discussed from the philosophic angle by Langford (1942).

A final topic that requires discussion before human translation and MT can be compared in detail is the range of use of symbols. In any particular passage, each symbol signifies something and is therefore precise in some sense (cf. Wittgenstein, 1953b). In the sentence 'My cat likes rum in her milk', V.cat refers precisely to a particular species of animal. There are however numerous varieties of cats and in respect of varietal distinctions, the sentence is vague. All ambiguous passages are vague at one level of precision, yet if a passage represents either N.x or N.y, it is precise if taken as representing N.x or y.

A great many symbols yield very different indicata in different contexts. In the sentence 'My sister let the cat out of the bag', the indicatum of V.cat is vastly different from what it was in the earlier example.

A distinction is frequently made between normal and metaphorical usage of a symbol. It is not clear whether a formal distinction between these two usages can be established. It is certainly the case, however, that most symbols have a range of use and that the determination as to which use they have in a particular passage is a crucial operation in both human translation and MT.

The decision as to which use a symbol has in a particular passage depends on its relations with other symbols. Two principal types of relevant relationship can be distinguished, (1) relations between lexico-grammatic and syntactic symbols in the L and S categories and (2) relations between naked ideas in the N category. These two types of relation are not of course to be considered as exclusive. To exemplify, the coexistence of the symbols L.let, L.cat, L.out, L.bag affords a strong presumption that the unwitting release of information is being indicated. In the Italian word sequence

$$L_1.la \quad L_1.prod \quad L_1.uzion \quad L_1.e$$

L_1.la indicates L_3.definite article; in other contexts it indicates a preverbal pronoun but this usage is excluded here since

$$L_1.prod \quad L_1.uzion \quad L_1.e$$

indicates L_3.noun and not L_3.verb.

The significance of 'plant' in the English passage 'a pineapple slicing plant' is revealed by a combination of relationships involving both S and N categories. This passage has the following indicatum

$$L_3.indefinite\ article \quad L_3.noun \quad L_3.verb\ stem$$

$$L_3.present\ participle \quad L_3.noun$$

which represents

$$S.noun\ object \quad S.verb \quad S.noun\ agent$$

Then, finally, the decision as to whether L_1.plant represents the botanical entity, N_1.plant, or a machine installation, N_1.machine, can be approached by examining the congruence between the indicatum of S.verb, i.e. N_1.slice, and the two potential indicata of S.noun agent, i.e. N_1.plant and N_1.machine. N_1.machine and N_1.slice are fully congruent but N_1.plant and N_1.machine are less so and therefore N_1.machine will be chosen as the N_1.indicatum of L_1.plant.

This preamble provides one means of examining the translation process as a whole and we can now pass on to a survey of the various operations involved and of the different ways in which these are performed in human translation and in MT.

HUMAN TRANSLATION AND MT

The first translation operation is recognition of the passage to be translated. In human translation, this involves either auditory or visual perception. The allocation of symbols to phonemic or graphemic categories is accomplished with great facility by the human translator, though it is almost invariably the case that recognition is partly based on context. This is shown by the frequency with which printer's errors are overlooked

in proof, the reader recognizing a letter as otherwise than what it is owing to the context in which it occurs. In reading handwriting, a great deal of recognition depends on context.

Till now MT has principally proceeded from written texts and the initial recognition and subsequent encoding has been done by a human operator. Photoelectric scanning is, however, feasible even though it is hardly yet a practical alternative, but is not likely to reduce all the variant token symbols of a single type to their G indicatum anywhere near so readily as the human translator. It may, for instance, be necessary to use separate matching operation to identify, say, roman, italic or gothic founts representing the same letter. Then, by encoding all variants of the same grapheme in the same way, reduction to the same G symbol can be accomplished.

Any photoelectric scanning procedure will probably have to allow for a margin of uncertainty in identification. How much deviation can be tolerated is a matter which the operator must decide. As mentioned earlier, a choice has to be made as to what is to be regarded as an initial symbol. This need not be the same in human translation and MT.

It is probable that the human translator is seldom unaware of the phonemes and uttered-words represented by a written text. In cases when script is vaguer than the spoken language, as in the use of invalid in 'His grandfather left an invalid will', and 'His grandfather had been an invalid for years', the human translator will probably elaborate different P and U symbols in the domains of symbolization of the two passages.

In MT it is possible to establish auditory distinctions of the sort mentioned but there is no point in doing so since the distinction can be treated as a particular instance of range of use. Current MT procedures, therefore, differ from human translation in having no analogues for A, P and U symbols.

Having recognized the symbols of the input, it is possible both in human translation and MT to proceed unanalytically. In the case of the human translator, this could involve no more than matching the entire input against a phrase book and writing out the equivalent provided. No understanding of either input or output is required; the only requirement is ability to derive the graphemes of the input symbols. Given a German-Italian phrase book, an English operator, ignorant of either foreign language, could correctly translate from German to Italian whenever the phrase book included the set of symbols of the input.

An analogous unanlytic procedure is feasible mechanically and once an input passage has been recognized and encoded, the problems involved in matching it against a mechanical phrase book, finding the equivalent, and putting this out, are trivial. However, it is obvious that such a procedure is of practical value in only a very limited number of cases. It may have some relevance in highly stereotyped situations such as those involving business correspondence or ordering meals in foreign restaurants.

Already, however, we are faced with the pervasive problem of dictionary searching. The human translator, in searching for equivalents, may rely on his memory, an extremely efficient operation from the point of view of access speed but less efficient from the point of view of comprehensiveness of entries and freedom from errors. Alternatively, he may consult a dictionary, usually alphabetically arranged; this is a fatiguing procedure, though the storage capacity of a printed dictionary is high.

Computer storage media are being so rapidly developed that it would be premature to state what the practical limits of storage capacity and access time are. To date, three principal dictionary searching procedures have been considered for MT: direct alphabetic searching, reverse alphabetic searching (Richens and Booth, 1955) and the partitioning or bracketing method (Booth *et al.*, 1958). The first and third are suitable for invariant phases or uninflected words since decomposition of the phrase or word into smaller units is not required. Under these circumstances, the method of Booth *et al.* is the speedier. The reverse alphabetic method is probably best for dealing with languages containing inflected words, since this method permits both decomposition of inflected words into their constituent units and location of the units in the mechanical dictionary. This method, however, involves an analytic operation in addition to recognition and matching of the input symbols and is more properly discussed below.

Phrase books normally connect two languages directly but it is possible to conceive of phrase books in which translation is achieved in two stages via some interlingua. One set of phrase books would be of the form, input language - interlingua, and the other, interlingua - output language. Perhaps the only advantage of such a procedure in unanalytic translation would be the reduction in the number of phrase books required for translation in all directions between n languages: $n(n-1)$ in the case of ordinary phrase books and $2n$ in the case of the interlingual method. It must be remembered that neither natural languages nor the current artificial languages can serve as satisfactory interlinguas since it is seldom the case that a translation into any of these has exactly the same indicatum as the passage translated. It would also be possible for unanalytic MT to proceed via an interlingua though this would be of little practical interest either. The sole advantage of unanalytic interlingual translation is the reduction in total dictionary requirements. When, however, analytic MT translation is being considered, several additional advantages of interlingual methods become manifest and under these circumstances they became a practical issue of some importance.

In analytic translation, the input passage is no longer treated as a single unit but as a set of symbols which can be analysed.

In both human translation and MT, the input symbols are analysed into ordered sets, each yielding a word or word segment as its indicatum.

The human translator may utilize uttered-word segments (U symbols), written-word segments (W symbols), or generalized lexico-grammatic symbols of the L category, but for MT only the L category is normally considered. It is possible to treat entire words as the units for memory or dictionary consultation but it is more usual in human translation and of greater general utility in MT for the search unit to be a word segment. Thus L_1.dogs will not be found in a normal dictionary but L_1.dog will. In MT, mechanical dictionaries have been devised that take the entire word as the search unit. Such dictionaries make heavy demands on storage space and lengthen the access time required for a single search through the entire dictionary, particularly in highly inflected or compounding languages such as German, Finnish or Hungarian. A very considerable economy in storage requirements is effected by using a mechanical dictionary of word segments, stems and affixes, each of which has a significance or a range of significance of its own.

We are confronted, then, with the necessity of deciding which symbols, phonemes or graphemes are to be associated together in the words or word segments. Most modern written languages indicate the boundaries of words by spacing. In Arabic script, additional indication of the word boundaries is provided by the special forms of initial and final letters. In Japanese, however, the boundaries between words in hiragana script are not indicated and have to be ascertained by other means.

The boundaries of word segments are more difficult to discover and the human translator employs many criteria, probably discerning the most distinctive segment first and then identifying the rest by elimination. Thus in

G.d G.o G.g G.s

he will first isolate L_1.dog; in

G.j G.o G.k G.i G.n G.g

L_1.jok; and in German

G.g G.e G.t G.r G.a G.g G.e G.n

L_1.trag. The human translator tends, after isolating a word segment such as L_1.jok, to transform this into a conventional form such as L_1.joke which is an actual word representing the segment and is used by dictionary compilers in arranging their entries. This transformation of word segments to a conventional word is redundant in MT, which can operate directly with L_1.jok, which is entered as such in the mechanical dictionary.

The human translator can pick out distinctive word segments from the beginning, interior or end of words. This generalized operation is not easy to mechanize and the usual procedure is to identify word segment from the beginning of the word. One of the most effective methods (Richens and Booth, 1955) is to compare words against a mechanical dictionary arranged

in reverse alphabetic order and to effect a match when the word is first wholly contained within the dictionary entry. This procedure would result in misdivision of the word in some cases, thus discontent is liable to be divided into disc-on-tent. This problem has been discussed by Richens and Halliday (1957); the general solution for the difficulty is to place awkward words as separate undivided entries in the mechanical dictionary.

The purpose of recourse to memory or to a dictionary or grammar is to obtain as many indicata in the domain of symbolization of each L symbol as are requisite for translation, and also to obtain appropriate corresponding symbols in the language of the output. Whenever a human translator utilizes symbols of the N category, he is in fact translating interlingually, since as noted above, only one set of N symbols is appropriate to any particular terminal indicatum. It is probably exceptional for human translation to proceed without any recourse to symbols of this degree of generality.

Most MT procedures have been devised to proceed from a particular input language to a particular output language without recourse to an analogue of the N symbols. In favour of this procedure is the simpler programme. Richens (1958) has urged the advantages of interlingual MT, on the grounds, mentioned above, that fewer mechanical dictionaries are required, and has pointed out that even if an analogue of the N symbols is dispensed with in obtaining equivalent symbols from the mechanical dictionary, no method of solving the semantic problems connected with multiple use without recourse to N-symbol analogues has been devised. The type of interlingua envisaged is a logically formalized standard language fulfilling, as far as possible, Wittgenstein's (1922) criterion that a linguistic passage and the state of affairs symbolized by it should have the same logical multiplicity.

Between direct MT translation, a one-one procedure, and Richens' interlingual method, a many-many procedure, there are the many-one methods of Masterman (1956, 1957), Halliday (1956) and Parker-Rhodes (1956b). In these, translation proceeds from any input language to the output language via a mechanical thesaurus in the output language. This approach is referred to in more detail below.

Returning to the example quoted earlier, this would be encoded mechanically in some such form as

$$L_3.\text{capital G.s G.h G.e G.c G.a G.m G.e G.t G.o G..}$$

and this, utilizing the ordering information provided by the linear sequence of the symbols and by spacing, can be matched against a mechanical dictionary to yield the sequence

$$L_1.\text{she } L_1.\text{came } L_1.\text{to}$$

Since none of these words is inflected, problems attending decomposition into word segments do not arise. In languages where a single word may consist of up to five segments, not uncommon in German or Finnish, these problems may be of some complexity.

The problem of multiple use arises as soon as dictionary matching has been achieved. Thus, if we had the English word L_1. plant, it might indicate the botanical entity, a machine installation, the act of inserting the botanical entity into the soil, the act of founding a community, or the perpetration of a fraud.

The human translator's approach to the problem of deciding the use intended is one of extreme flexibility, and many different criteria may be used for the various cases within a single sentence. The tendency in MT is to solve these problems as early in the programme as possible since the solution of one such problem is often a prerequisite for the solution of another.

Thus, in such a passage as 'He planted a new colony', corresponding to

$$L_1.\text{he}\ L_1.\text{plant}\ L_1.\text{ed}\ L_1.\text{a}\ L_1.\text{new}\ L_1.\text{colony}$$

the fact that L_1.plant is followed by L_1.ed suffices to eliminate the uses, botanical entity or machine installation, but is of no help in deciding between the remaining uses.

In the case of

$$L_1.\text{she}\ L_1.\text{came}\ L_1.\text{to}$$

the problem of range of use will probably be deferred in MT till L_1.came has been collated with the dictionary entry L_2.com. It is possible to deal with the range of use of the various forms of strong verbs independently of each other but this would entail quite unnecessary duplication of dictionary material for each entry. It is probable too that human translators consider L_1.came and L_1.com as members of a single more general category whose range of use can be considered as applying to all its variants.

The MT procedures just described differentiate between inflexion and ablaut. Thus from

$$G.c\ G.a\ G.m\ G.e$$

we proceeded to L_1.came and thence to L_2.com.

$$G.k\ G.i\ G.c\ G.k\ G.e\ G.d$$

however, would yield

$$L_1.\text{kick}\ L_1.\text{ed}$$

with Latin

$$G.r\ G.e\ G.x\ G.i\ G.t$$

would yield

$$L_1.\text{rex}\ L_1.\text{it}$$

and L_1.rex would be referred to L_2.reg.

The distinction should not be pressed. L_1.sang could be made to yield

L_2.s-ng L_2.-i-

and quite possibly this sort of symbolic relationship is utilized by human translators.

In MT, it is easier to handle affixation and ablaut separately and therefore the structure of the domain of symbolization of the two is differentiated. Whether as L_1.came or L_2.com, we are confronted with a wide range of use. L_1.com, considered alone, can indicate motion towards, or happen, and in combination with other L_1. symbols we have the uses indicated by come about, come across, come by, come home to, come in for, come into the world, come of, come on, come to a head, and come to, the latter combination indicating either motion towards or recovery of consciousness.

It is doubtful whether either in human translation or MT, a decision as to the use of L_2.come or L_2.to can be decided without reference to the relations between L_3. symbols which yield the S categories. L_1.she has a narrower range of use and, if a nautical context can be excluded, will seldom be other than the third person singular nominative female pronoun. The L_3.symbol sequence has been given as

L_3.capital L_3.pronoun L_3.singular L_3.nominative L_3.verb

L_3.past tense L_3.verbal postposition L_3.stop

The human translator will have little difficulty in inferring that the L_3.capital ... L_3.stop sequence bounds the syntactic unit termed the sentence. This can also be ascertained mechanically though it is necessary to discover whether the capital preceded a word normally capitalized and whether the final stop is a sentence-conclusion indicator, an abbreviation indicator or a combined abbreviation and sentence-conclusion indicator.

The nominative pronoun in an initial position will suggest to the human translator that it is the grammatic subject of the sentence, the following verb will be taken as the predicate and the final particle with no following noun or noun equivalent will be attached as a postposition to the verb. It is important to note that the syntactic relations are described here as derived from ordered L_3. grammatic categories. This is not invariably so and the flexibility of the human translator is shown again in the way in which other criteria, such as semantic congruence, can be used in establishing syntactic relations.

For instance, in the Japanese sentence 'ruku ka sen shoku tai ni zigzag jo nomonogo mii da sareta', i.e. 'zigzag configurations were found in the hexavalent chromosomes', the postposition L_1.ni is translated an 'in'. It is also used however as the agent of passive verbs. The verb is passive here but L_1.ni is not taken as the agent since chromosomes do not make observations.

MT syntactic analytic methods for deriving the S categories are far more stereotyped than those of the human translator and have tended to be based solely on the relationships between grammatic categories. On the other hand, though the MT methods for syntactical analysis are sterotyped, they have sometimes been based on principles that have no obvious parallel in human translation.

Among the MT syntactic analytic techniques with close analogies to those of the human translator are those of Yngve(1955) and Richens(1956a). In these, syntax relations are determined by detection of sequences of grammatic classes of word segments, the significance of the sequence being obtained by consulting a mechanical dictionary of such sequences, in effect a mechanical grammar. Parker-Rhodes (1956a) has experimented with a technique of syntactic analysis which is algorithmic and aims at performing syntactic analysis with no or little recourse to a dictionary of grammatic forms. The efficiency of this method is not yet known, but should it prove effective, it would provide a good example of an MT procedure based on an algorithm with no obvious analogue in human translation.

In human translation and most types of MT, when the syntactical patterns in the S categories have been determined, an appropriate arrangement of symbols in the output language can be obtained by reordering transformations and the provision of syntactic flexional forms. Thus, from Spanish 'una casa grande' reordering 123 $\rightarrow$ 132 yields English 'a large house'. English 'the eagle does not catch flies', does not require reordering to obtain the Latin 'aquila non capit muscas', but needs appropriate affixes for the grammatic subject and object of the verb.

In the interlingual MT technique of Richens, syntax is expressed solely in terms of a configuation of syntactic bonds at this stage and is left as such in the interlingua without rearrangement. Reordering is only introduced when proceeding from the interlingua to the output language.

Once the syntactic relations within a passage are determined, the range of use of the symbols narrows greatly. Thus, in our example, the facts that there is a syntactic bond between L_1.came and L_1.to and L_1.to is not followed by a noun or noun equivalent are sufficient, both in human translation and MT to indicate that

$$L_1.\text{came}\ L_1.\text{to}$$

is a semantic unit and to exclude the signification 'motion towards'.

Syntactic analysis resolves many points that arise during translation, but there remains a residuum of problems concerning range of use that depend for their solution on semantic analysis. Here, the powers of the human translator are seen at their subtlest, the range of criteria used in semantic analysis being extremely wide. In one of the examples quoted in the previous section, the derivation of the significance of the phrase 'letting the cat out of the bag' would probably be made on the basis of collocation of symbols in the L category. An example where the use was

determined by semantic congruence of N symbols was also given and it is easy to think of cases in which the use of a symbol is decided by some item of knowledge, either scientific or historic. Thus, in the sentence 'Jane Shore was the mistress of Edward IV', the decision as to the significance of 'mistress' might well be made on purely historical grounds.

The mechanization of semantic analysis is perhaps the most recondite of MT problems, especially as it is probably the case that not one but several types of semantic analysis are required for adequate translation.

Booth *et al.* (1958) deal with semantic problems by characterizing each use of a symbol as pertaining to one or other of a limited number of semantic fields. The field is either assumed to be known in advance or is determined during the course of translation by counting unambiguous words.

A far more elaborate development of the semantic field conception has been made by Masterman (1956, 1957 and personal communication), Halliday (1956) and Parker-Rhodes (1956b). In their approach, the thesaurus method, semantic analysis proceeds concurrently with the treatment of syntax. The first operation in this method involves obtaining for each L symbol, a series of thesaurus heads, corresponding roughly to N symbols, covering the entire range of use of each L symbol. Thesaurus heads represented once only in a sentence are eliminated by diallel comparisons of the sets of heads for each L symbol. Next, L symbols with relatively invariant significance are withdrawn from further comparison. Then, finally, a splay of synonyms in the output language is obtained for each remaining thesaurus head, and the synonym or synonyms common to the remaining heads of each L symbol are used to provide the basis for the output.

The semantic analytic technique of Richens (1958) differs from the foregoing in being more tightly associated with syntactic analysis which it follows. Three separate semantic analytic operations are envisaged: (1) semantic congruence between syntactically bonded N symbols, (2) identification of special collocations of L symbols, and (3) a form of the semantic field method.

It is interesting to recall at this point the Japanese sentence considered above. It is doubtful whether syntactic analysis together with semantic-field considerations could correctly determine the use of L_1.ni in this sentence. However, the semantic incongruence between the N indicata of the verb and of the putative agent is sufficient to determine the alternative use of L_1.ni and hence to derive the correct syntactic bonding.

A difficulty arises at this point. Neither human translation nor MT is feasible without some semantic analysis yet passages are frequently encountered where the metaphorical use of words presents serious difficulties. Take Crashaw's

> ..... soft powers
> Whose silken flatteries swell a few fond hours
> Into a false eternity.

A pedestrian semantic analytic programme may splutter at powers being soft, at flattery composed of silk, at hours being either fond or swelling and at eternity being false. There is every possibility that semantic incongruences will be detected in such a passage and it is necessary, when these would otherwise inhibit translation, to relax too tight a form of semantic analysis or even discard it entirely and translate on the basis of syntactic coherence or semantic-field considerations.

In the example we have been following through,

$$L_1.\text{she}\ L_1.\text{came}\ L_1.\text{to}$$

yielded

$$L_3.\text{capital}\ L_3.\text{pronoun}\ L_3.\text{singular}\ L_3.\text{nominative}\ L_3.\text{verb}$$
$$L_3.\text{past tense}\ L_3.\text{verbal postposition}\ L_3.\text{stop}$$

whence we obtained

$$\text{S.pronoun subject}^{c}\ \text{S.verb governing postposition}^{abc}\ \text{S.verb qualifier}^{b}$$
$$\text{S.verbal postposition}$$

$L_1.$come, when bonded to $L_1.$to and in the absence of a noun or noun equivalent governed by $L_1.$to, forms a semantic unit with $L_1.$to which would transform the sentence into the form.

$$\text{S.pronoun subject}^{a}\ \text{S.verb predicate}^{a}$$

If a MT programme for direct translation from English to French is being run, reordering would be accomplished at this point. In this instance the reordering is of the form 1 2 → 1 2 and no change is required. However, assuming that the verb is rendered by the French 'se remettre', the correct reflexive pronoun has to be introduced between the subject and the verb, and assuming that the perfect tense is used, agreement has to be established between the gender of the subject and the past participle. The final output would be 'elle s'est remise'. Should an interlingual MT programme be used, an interlingual rendering of some such form as

$$N_2.x^{d}\ N_2.\text{female}^{cde}\ N_2.\text{one}^{c}\ N_2.\text{become}^{a}\ N_2.\text{conscious}^{able}$$
$$N_2.\text{past}^{b3}\ N_2.\text{communication situation}^{b2}$$

described in the first section, will be derived. It is possible at this point to regulate the type of translation according to the degree of similarity in the mediate symbols required. If the grammatic categories of the input are to be mirrored in the output, we might obtain such renderings as French 'elle s'est remise', Latin 'animum collegit', or Japanese 'sho ki ninatta'.

It is relevant to point out that none of the translations proposed, nor in fact any likely to be offered by a human translator, have the same terminal indicatum as the input passage. The French verbal stem is vaguer

but its tense, with its perfective aspect, more precise, the Latin fails to indicate the sex of the subject and the Japanese, literally 'there was a becoming to consciousness', indicates neither sex of the subject nor differentiates between first or third person or singular or plural number.

In direct MT, the shift in the terminal indicatum may take place at any stage in the programme. In interlingual MT, no shift in the terminal indicatum should occur till after the interlingual rendering has been derived.

A few remarks should perhaps be added on some rather special cases, which though of infrequent occurrence cannot be totally ignored in any formal study of translation.

It has already been noted that semantic analytic procedures encounter difficulties with some types of metaphor. Similar difficulties may arise in passages containing logical contradictions or contradicting definitions of words in the mechanical dictionary. An extremely rigorous semantic analytic programme might even throw out tautologies as vacuous. Further it would be unwise to depend too exclusively on syntax since sentences defective in grammer or syntax are not infrequently encountered.

Sentences such as 'pig has three letters', where the terminal indicatum of 'pig' is a symbol and not the animal also require care, lest, for example a translation such as French 'porc a trois lettres', be obtained. It would be feasible in such cases, though perhaps not worth while, to include special directions in the mechanical dictionaries concerning entries referring to symbols, so as to avoid this trouble.

In such sentences as 'Chic is a foreign word', foreign is relative to the language of discourse and so, in translating into French or any other foreign language, is best rendered as 'foreign to English'. There is also a convention that passages in quotes should be translated, even though their function is often to indicate the original form of a communication.

These examples are not quoted as serious problems, but merely to stress that neither human translation nor MT are procedures to which a single rule of transformation is applicable. Human translation is characterized by its flexibility and the varied methods it employs for dealing with different types of passage. The need for differentiating between situations in which syntactic analysis must be paramount and those in which semantic analysis is of prime necessity has already become evident in MT, and it is possible that the power of discriminating between procedures rather than the intensive development of the procedures themselves will be the major MT research problem of the future. There is no immediate prospect of MT rivalling the all-round performance of the human translator though it is highly probable that individual translation operations might, in some cases, be more efficiently accomplished mechanically, with, on occasion, the production of a translation superior to human.

ACKNOWLEDGEMENTS

The preparation of this paper has been greatly assisted by discussion with Professor R. B. Braithwaite, and Miss M. M. Masterman, Dr. M. A. K. Halliday and the other members of the Cambridge Language Research Group. Thanks are due to the National Science Foundation of the United States for a grant in support of this work.

REFERENCES

1. ALLEN, W. S. Relationship in comparative linguistics. *Trans. Philol. Soc.*, 1953, 52.
2. BOOTH, A. D., BRANDWOOD, L. and CLEAVE, J. P. *Mechanical resolution of linguistic problems*, 1958.
3. HALLIDAY, M. A. K. 1956. Linguistic basis of the thesaurus-type mechanical dictionary and its application to English-preposition classification. *Mechan. Translation*, 2, 36. (Abst.)
4. LAMBEK, J. The mathematics of sentence structure. *Amer. Math. Monthly*, 1958, **65**, 154.
5. LANGFORD, C. H. The nation of analysis in Moore's philosophy. *The Philosophy of G. E. Moore*. edit. P. A. Schilpp. 1942.
6. MASTERMAN, M. Potentialities of a mechanical thesaurus. *Mechan. Translation*, 1956, **3**, 36. (Abst.)
7. MASTERMAN, M. M. The thesaurus in syntax and semantics. *Mechan. Translation*, 1957, **4**, 35.
8. PARKER-RHODES, A. F. An electronic computer programme for translating Chinese into English. *Mechan. Translation*, 1956a, **3**, 14.
9. PARKER-RHODES, A. F. Mechanical translation program utilizing an interlingual thesaurus. *Mechan. Translation*, 1956b, **3**, 36.
10. RICHENS, R. H. Preprogramming for mechanical translation. *Mechan. Translation*, 1956, **3**, 20.
11. RICHENS, R. H. A general programme for mechanical translation between any two languages via an algebraic interlingua. *Mechan. Translation*, 1956b, **3**, 37. (Abst.)
12. RICHENS, R. H. Interlingual machine translation. *Computer J.* 1958, **1**, 144.
13. RICHENS, R. H. and BOOTH, A. D. Some methods of mechanized translation. *Machine Translation of Language*, 1955, edited by W. N. Locke and A. D. Booth, p.24.
14. RICHENS, R. H. and HALLIDAY, M. A. K. Word decomposition for machine translation. *Rep. Eighth Annu. Round Table Meet. Linguist. Language Stud., Georgetown, Washington, 1957*, p.79.
15. WITTGENSTEIN, L. *Tractatus logico-philosophicus*, 1922, p.71.
16. WITTGENSTEIN, L. *Philosophical investigations*, 1953a, pp.21-24.
17. WITTGENSTEIN, L. *Ibid.* 1953b, p.45, p.139.
18. YNGVE, V. H. Syntax and the problem of multiple meaning. *Machine Translation of Language, loc. cit.*, 1955, p.208.

DISCUSSION ON THE PAPER BY MR. R. H. RICHENS

PROF. Y. BAR-HILLEL: Mr. Richens, as most of you probably know, is one of the pioneers of machine translation. A paper he wrote, in collaboration with Booth, in 1948, which was reproduced in 1952 for the *First International Conference on Machine Translation* and finally published in 1955 *(ref. 1)* was the first serious contribution to machine translation altogether, and I am sure that everybody who was working on MT in these early times drew great profit from studying it. Unfortunately, however, during the last years Mr. Richens went off into lines of thinking which I do not believe will be very fruitful. I have no time to discuss here the issue of interlingua which was the major topic of Mr. Richens' oral presentation - I do this, however, in a forthcoming report on "The present state of machine translation in the United States and Great Britain" -. I shall rather go into the theoretical foundations which Mr. Richens provides us with in his paper. I personally am much more interested in these foundations than in the practical applications and was therefore highly disturbed by the way Mr. Richens accomplishes this job.

Mr. Richens' approach relies heavily on two terms, 'indicate' and 'concept'. Both terms are notoriously vague in ordinary usage and one might have expected an attempt of clarification on behalf of one who is going to use them so extensively. However, what is meant by 'concept' is never discussed at all. With regard to 'indicate' and its counterpart 'indicatum', not only is no indication of their prospected usage given, but it turns out that Mr. Richens uses these terms such that a sound indicates a letter, a letter a word, a word a concept, a concept other concepts as well as other entities. Why Mr. Richens should believe that by employing a single term for a whole gamut of what are usually regarded to be rather different relations he will gain something, is beyond me. I am sure that nothing useful can result from such a treatment.

Mr. Richens' pair of terms, 'indicate' and 'indicatum', is obviously coined to replace the more customary terms, 'designate' and 'designatum', 'denote' and 'denotatum', 'signify' and 'significatum'. Mr. Richens' preference might be due to the fact that the other terms were not always used with sufficient rigor. It may also be due to the fact that some of these terms were used, e.g. by Rudolf Carnap in his books *(refs. 2* and *3)* with too high a degree of rigour for his purposes. At any rate, some indication of the relationship would have been helpful.

Refs. 1, 2 and *3* on page 305.

There are quite a few additional passages which must cause a great deal of confusion. Take the definition of 'symbol'. A symbol, according to Mr. Richens, page 282, is anything which either alone or in conjunction with other symbols indicates something. From this definition I gather that a symbol which indicates something *only* when in conjunction with other symbols does not indicate anything by itself. The insight that not every symbol symbolizes is as old as Aristotle. 'Syncategorematic' was the term used for this kind of symbols by the Schoolmen. Unfortunately, Mr. Richens himself looses very soon sight of his own definition and talks as if every symbol had its indicatum.

There is also a lot of confusion in the use of the term 'initial symbol'. You find on page 282 that an initial symbol is a symbol which is not an indicatum of some other symbol or symbols. However, a page or two later the same term is used as a synonym for 'simple symbol' in contrast to 'complex symbol' rather than to 'mediate or terminal symbol'. As a result of this confusion, you will find on page 283 the problem raised whether the German 'ü' should be regarded as an initial symbol or as an ordered arrangement of 'u' and Umlaut. I do not think that this question has anything to do with the question whether 'ü' is an initial symbol in the sense defined on the page before. The issue is, of course, whether 'ü' should be regarded as a simple or as a complex symbol, which is a totally different distinction from the initial-noninitial one.

On page 286, the term 'naked ideas' is introduced and we are told of its provocative entry in 1955. It is well known, of course also to the Cambridge Group itself, that similar conceptions played a big role in medieval speculations as well as in the speculations of such authors as Descates, Leibniz, Locke and Bishop Wilkins in the 17th century, and that they did not do much good. Though this fact does not prove that a similar idea could not meet with better success in the 20th century, no real argument is given why one should expect a change. At any rate, I do not understand what it could possibly mean that 'give', 'gift', 'present', 'donate' and 'donation' have a unique indicatum. Without a more or less complete system telling in detail what difference it makes whether one assumes that these words do or do not have a unique indicatum, I cannot see how anything of consequence can be said about the subject. A couple of sentences are surely not enough to endow it with any scientific import whatsoever.

Let me conclude with a few minor comments.

I was disturbed to find on page 289 a reference to a paper by Lambek to the effect that Lambek has approached the problem of formalising comparative grammar from a mathematical angle. I know this paper pretty well. What Lambek does there is to develop further, especially in the algebraic direction, a notation for syntactic description which I proposed a few years ago *(ref. 4)*, which itself was based on certain ideas brought up by the Polish

Ref.4 on page 305.

logician Ajdukiewicz *(ref. 5)*. I cannot see what this has to do with comparative grammar, though I can see its connection with what has been termed 'universal grammar'.

I am also profoundly disturbed by the fact that Prof. Braithwaite is mentioned as one of the people who assisted Mr. Richens in discussion of his ideas. I would be highly gratified to learn that Prof. Braithwaite is not at all to be held responsible for any of the ideas I have been criticising. I can hardly believe that he would entertain such ideas, knowing Prof. Braithwaite personally.

In the last part of his paper, Mr. Richens makes some very interesting remarks on certain difficulties in machine translation which might have been noticed before but were never discussed before in print, perhaps because there were graver problems that required discussion. However, I would not agree with the way in which Mr. Richens almost dismisses the problems he pointed out himself. He mentions, for instance, the problem of translating into French such a sentence as "Pig has three letters" (where 'pig' is written without quotation marks, as even scientists - especially in England, I am told - are in a habit of doing). Translating this sentence as '*Porc a trois lettres*' is clearly wrong, as already indicated by the fact that the English sentence is true and the French one false. However, "*Jean a quatre lettres*" is a wrong translation of "John has four letters" though both sentences are true. I do not believe at all, in distinction from Mr. Richens, that it would be easy to program a machine to come with the correct translation "'*John*' *a quatre lettres*". Even more generally, I am convinced that the treatment of machine translation of material in quotation marks (or of material that should have preferably been printed in quotation marks) will cause a lot of trouble, to such a degree, indeed, that I regard it as highly unlikely that a fully automatic treatment of such material will at all be possible. By the way, I know that human translators have troubles with such material, too. But this does not mean that the problem is not serious or that its treatment can be postponed for long.

Another problem is the translation of such a sentence as "Chic is a foreign word". Translating this into French as "Chic [or 'chic', for that matter] est une parole étrangère" is plainly wrong since 'chic' is not a foreign word (in French). 'Foreign' is a context-dependent word, and

REFERENCES

1. RICHENS, R. H. Some methods of mechanized translation in Machine Translation of Language ed. by W. N. LOCKE and A. D. BOOTH (1955).
2. CARNAP, R. Introduction to Semantics. *Harvard University Press*, (1942).
3. CARNAP, R. Meaning and Necessity. *University of Chicago Press*, (1947).
4. BAR-HILLEL, PROF. Y. A quasi-arithmetical notation for syntactic description. *Language*, 1953, 29, 47.
5. AJDUKIEWICZ, K. Die syntaktische Konnexitat. *Studia Philosophica*, 1935, I, 1.

translating an English sentence containing this word into another language changes the context decisively. An unsophisticated translation of such a sentence will therefore not be a good translation.

Again, contrary to Mr. Richens' opinion, I believe that the problem involved is serious. There is no simple procedure to find out which, and in what way, the words of the English language are context-dependent. And I don't think that the issue can be belittled for the reason that context-dependent words do not occur in scientific discussions and writings. They might not be too abundant in ordinary scientific papers on matters physical or chemical, but there would surely be plenty of them in discussions of matters linguistic, for instance. This might be one reason why so far hardly anybody has tried to machine translate papers in linguistics. As soon as this is attempted, the seriousness of the problem will become immediately evident.

MR. R. H. RICHENS (in reply): I appreciate Prof. Bar-Hillel's stimulating comments on what he takes to be my philosophical views. I should however point out that my paper was not primarily intended as a contribution to philosophy nor am I a philosopher. The general principles set down in my paper were not developed first and then applied to mechanical translation; rather, the machine translation schedules were devised first, with a great deal of trial and error, and then afterwards it proved possible to discern certain principles of general application.

Prof. Bar-Hillel states that he would be highly gratified to learn that Prof. Braithwaite were not responsible for the views that he has criticized. I can offer Prof. Bar-Hillel little solace here. Prof. Braithwaite is, of course, in no way responsible for any of the views set out in my paper, but he is cognizant of the entire conceptual scheme developed, he was present at the research meetings of the Cambridge Language Research Group at which these views were exhaustively discussed, he made a number of suggestions that were incorporated into the final text, and, I am assured, he is not in strong disagreement over any major issue. The term 'indicatum', to which Prof. Bar-Hillel takes a strong dislike, was adopted after a discussion of Prof. Bar-Hillel's alternatives with Prof. Braithwaite. Each of these alternatives carries implications which I would repudiate in the present context.

Prof. Bar-Hillel fails to understand why the term 'indicate' is used to cover 'a whole gamut' of relations. I think that, on this point, he has confused vagueness with generality. I am well aware that the connections between the various categories of symbols that I have used are diverse. However, I imply nowhere that 'indicate' is a monolithic notion. What I do assert is that there are features common to all the categories of 'indicate' which justify the use of a single term. Moreover, there is a precise

analogue to 'indicate' in the machine-translation programme on which I have been working, namely substitution of a term or terms by the following entries in a mechanical dictionary or inventory of higher-order categories.

The reproof that I talk as if symbols indicate in isolation is rebutted by my definition of a symbol. I did not include examples of indication by pairs or multiple associations of symbols in my paper; they are common in practice and duly allowed for in machine-translation programmes.

Finally, about naked ideas. I know that they have been floating around the world for a very long time. Whether or not they have been misused by others is not my concern. They are adopted here for a precise practical end, the resolution of ambiguities requiring semantic analysis. Possibly Prof. Bar-Hillel can see nothing in common in the words 'give', 'receive' and Latin 'do', 'dono' and 'praesto'. Most people, excluding philosophers and linguists, can. Even if they could not, the naked ideas discernible by semantic analysis are still useful in mechanical translation for the purpose I have mentioned.

After all, the justification for the principles I have enunciated, assuming that they have any, is not that they accord with any particular philosophical outlook but that they are effective in producing specimen translations from and to diverse languages. I assert nothing more.

SESSION 2

PAPER 6

PRONOUN REFERENCE IN GERMAN

by

L. BRANDWOOD

BIOGRAPHICAL NOTE

L. Brandwood was educated at the Priory School, Shrewsbury, and at University College, London, where he obtained a B.A. degree in Classics in 1953. He subsequently did research there on the literary style and chronology of Plato's dialogues before moving in 1955 to Birkbeck College as a Nuffield Research Fellow. At Birkbeck he worked on mechanical translation problems for three years under the direction of Dr. A. D. Booth, being co-author with the latter and with Dr. J. P. Cleave of "Mechanical Resolution of Linguistic Problems". At present he is an Assistant Lecturer in the Classics Department of the University of Manchester.

PRONOUN REFERENCE IN GERMAN

by

L. BRANDWOOD

THE PROBLEM AND ITS CAUSE

THE problem arises simply from the fact that German regularly attributes the masculine and feminine as well as the neuter gender to inanimate objects, whereas English rarely departs in this respect from a uniformly neuter representation. Conversely, an animate being may be denoted as neuter, but since in this case the same is true for the most part of English, the problem is encountered here to a much less degree. Thus a neuter pronoun in German referring to animals or even a human being, if it is a child, may readily be translated into the equivalent neuter pronoun in English; it is only when a somewhat later stage of life is reached (e.g. *das Mädchen*) that English cannot follow the German usage.

In short, therefore, a German masculine or feminine pronoun must sometimes be converted in English to a neuter pronoun, a neuter pronoun to a masculine or feminine pronoun. In order to determine whether any conversion is necessary, the antecedent of the pronoun must be identified, which for reasons that will become apparent through the illustrations presented is not always a simple matter from the point of view of mechanical translation.

THE MATERIAL INVESTIGATED

The present study was undertaken to determine how far purely syntactical considerations would solve the problem of pronoun reference. For this purpose it was decided to examine scientific texts in which persons were but infrequently mentioned. While this had the advantage of excluding the human/non-human criterion and so of making a more purely syntactical solution possible in theory, it was also automatically rendered defective. For since the human/non-human distinction was the only problem, and since a complete solution of this was not possible by consideration of syntax

alone, the means of discovering criteria for promoting the solution further were lost through this exclusion.

The investigation was limited to personal pronouns, the problem being most acute in their case, as they are generally capable of wider separation from their antecedent than the other pronoun forms. Several further restrictions also applied:

1. Since on occasions there is difficulty in determining whether it is functioning as a pronoun proper or merely as an introductory particle, the neuter form *es* was not considered for the sake of simplicity in compiling material.
2. The genitive forms were omitted as being more closely related in their use to the possessive pronouns, with which they might be treated more satisfactorily.
3. By contrast, although no problem exists in translating the plural forms, the nominative and accusative *sie* were included in order to increase the amount of material available from the texts examined. Without them this would have been reduced by about 38%. Needless to say, they are just as useful as the other forms for ascertaining the rules of pronoun reference.

The forms of the personal pronoun examined, therefore, were: er, ihn, ihm (masculine and neuter), sie (feminine singular and plural of all genders, nominative and accusative), ihr.

The texts, which were from books on the various sciences - physics, chemistry, biology, mathematics, geology, zoology, etc., amounted to approximately 200,000 words and yielded the following:

nominative pronouns (er, sie) - 416
accusative pronouns (ihn, sie) - 74
dative pronouns (ihm, ihr) - 14
accusative pronouns preceded by a preposition - 10
dative pronouns preceded by a preposition - 30

As a result of their low occurrence no attempt was made to proceed further with the cases other than the nominative, since whatever system they might have been reduced to would not have been of any great applicability. Even the 400 odd instances of the nominative pronoun are far from being sufficient material for a final analysis. Some facts, however, have become clear from the present study, defective though it be, and these are accordingly given in the following account. It must be left to future investigations to remedy these defects through a larger material, selected more appropriately from texts in which persons as well as things find frequent mention.

PRONOUN CATEGORIES

The pronouns could be divided into three main categories:

A. those occurring in a clause which was not the first in its sentence, in which case the reference was usually, but not invariably, to some substantive in a preceding clause of this sentence;

B. those occurring in the first clause of a sentence, the last clause of the preceding sentence being independent. In this case the reference was usually to a substantive in this last clause of the preceding sentence;

C. those occurring in the first clause of a sentence, the last clause of the preceding sentence being a dependent one, in which case the reference could be to this dependent or a preceding independent clause.

Within these categories it would have been possible to classify the various types of pronoun reference in several ways: the one actually adopted was as follows:

CLASS A. I. INCLUDED CLAUSES

A. I. (a) If the pronoun occurred in a dependent clause (other than a relative clause) included within another clause, it referred to the nearest congruent substantive in the preceding part of the thus divided clause, though excluding any substantive in parenthesis:

e.g. *Wird ein Glied eines solchen Systems durch einen *"Treffer"*,
wie er in der Strahlengenetik vorkommt,
* ionisiert,
so muss sich die Ladung durch Verschiebung der innermolekularen Elektronen und der zwischenmolekularen Protonen über einen Teil des Systems ausbreiten.

This was true of all the 15 instances which occurred in the texts examined.

It should be noted that where the part of the clause preceding the included clause was divided yet again by another included clause, the pronoun was not to be referred to any substantive in this previous included clause;

e.g. **Der Saft* der Bauchspeicheldrüse,
in dem das Enzym Trypsin enthalten ist,
* spaltet,
wenn er nicht zuvor mit dem Darm in Berührung kam,
* keine höheren, eigentlichen Proteine.

---

* In the German examples, where a clause is divided by an included clause, then the two halves of the divided clause will each be preceded by an asterisk.

A.I.(b) If the included clause containing the pronoun was a relative clause, the only method of identifying the antecedent seemed to be to examine all the substantives (except that referred to by the relative pronoun) in the preceding part of the divided clause in the hope that only one would show congruence in gender and number. Any methods of resolving the problem, if more than one should be congruent, did not come to light from the material at hand, inasmuch as in the 5 instances of this type the congruence was always unique:

e.g. *Eine Mischung von *Sauerstoff und Wasserstoff* im Verhaltnis,
in welchem sie im Wasser enthalten sind,
* heist Knallgas.

A.I.(c) If the pronoun occurred in the second of two included clauses, both being of the same type and dependent on co-ordinate elements, it referred to the subject of the first included clause:

e.g. *Mit der Erkenntnis,
dass *die Sonne* den Hauptkörper und den Mittelpunkt unseres Systems bildet,
* und der naheliegenden Vermutung,
dass sie uns, von einer sehr grossen Entfernung aus gesehen, wahrscheinlich als ein Stern erscheinen würde,
* gewann diese Sphäre Fixarum Stellarum plötzlich ein ganz neues Interesse.

This was the only instance in the texts examined.

A.I.(d) In 3 instances of the type A.I.(a) the substantive referred to was itself a pronoun, with the result that its antecedent, either in the same or a preceding sentence, had to be identifiable to make a correct translation of the following pronoun possible:

e.g. Wenn es auch möglich ist,
alle oben zusammengestellten kolloidalen Systemeherzustellen,
so wurden doch *Suspensionen und Emulsionen* am genauesten studiert;
*gerade *diese* werden auch,
wenn sie im oben definierten Dispersionsgrad aufgeteilt sind,
* kolloidale Lösungen genannt.

In the 3 instances concerned this identification would in fact have been made in accordance with the relevant rule in each case.

CLASS A. II. PARENTHETICAL CLAUSES

A.II.(a) If the pronoun occurred in a parenthetical clause, it referred to the subject - provided this was congruent - of the clause in which or after which the parenthesis stood:

e.g. *Während

wie unten noch begründent wird,

* *die klassische Dispersion* praktisch keine wichtige Rolle spielt

(sie tritt nur in einer höhreren Näherung in Erscheinung),

tritt eine solche im Gas mit chemisch reagierenden Molekülen auf.

There were 12 such instances including 5 where the substantive referred to was itself a pronoun, and with each of these the further antecedent was correctly identifiable.

It should be remarked that, apart from clauses enclosed in brackets, those introduced by d.h. were also regarded as parenthetical.

e.g. *Die Linse* weist für alle diese Elektronen das gleiche Brechungsvermögen auf,

d.h. sie ist achromatisch.

A.II.(b) There was a further instance of this type, in which the pronoun could not from lack of congruence refer to the subject of the clause after which the parenthesis stood, and in fact referred ultimately to the subject of the preceding sentence:

e.g. *Die Kristalle* vieler Hydrate mit hohem Wasserdampfdruck haben das Bestreben,

*Wasser zu verlieren,

wenn sie der Luft ausgesetzt sind,

* und zerfallen in Pulver.

Mannennt diesen Vorgang Verwitterung (sie verwittern).

In such cases the pronoun might be treated as falling under Class B or C, whichever was relevant. In the present instance the pronoun would be referred to its correct antecedent under rule B.II.a (q.v.)

CLASS A. III. CO-ORDINATE CLAUSES

A.III.(a) If the pronoun occurred in the second of two co-ordinate clauses, it referred to the subject of the first, provided this was congruent. In the case of independent clauses both those connected specifically by a conjunction (e.g. und, und zwar, oder, oder aber, aber, denn,

jedoch, auch, dann, teils - teils, einerseits - andererseits) and those left unconnected were regarded as co-ordinate:

e.g. (a) *Einteilung* nach der Wertigkeit hat gewisse Aussichten,

aber sie bringt Substanzen von diametral entgegengesetzten Eigenschaften zusammen.

(b) *Manche feste Stoffe und Flussigkeiten* ziehen Wasser an,

ohne sich jedoch notwendigerweise darin aufzulösen,

sie heissen ebenfalls hygroskopisch.

There were 34[ø] instances where the pronoun in the second co-ordinate clause, being congruent with the subject of the first, referred to this, 1 instance where, though congruent, it did not:

e.g. Der Endpunkt der Titration wird durch eine kleine, zu der Lösung zugegebene Menge *eines Indikators* angezeigt,

er ändert die Farbe,

je nachdem die Lösung sauer oder alkalisch ist.

It may be noted that in this case the reference was to the agent in a passive construction, that is to say, what would have been the subject of the corresponding active construction.

A.III.(b) In 1 instance the pronoun was congruent not with the subject of the first co-ordinate clause but with its dependent genitive, to which it referred.

i.e. Die Einfachheit *der Gesetze* der chemischen Vereinigung führt zu der Überzeugung,

dass sie auf Grund einer einfachen Vorstellung über den elementaren Aufbau der Materie erklärt werden können,

und tatsächlich lassen sie sich aus der Atomhypothese leicht ableiten.

A.III.(c) In 2 further instances the pronoun, being congruent neither with the subject nor its dependent genitive, referred to the direct object of the first co-ordinate clause:

i.e. Die Kraftwerke vom Rheinfall bei Schaffhausen bis zur niedersachsischen Tiefebene geben *ihre Energie* in das grosse westdeutsche Hochspannungsnetz,

und dem Bedarf entsprechend, wird sie den Verbrauchern zugeleitet.

ø In 4 of those the substantive referred to was also a pronoun, but in each case further reference to the correct antecedent was practicable under the appropriate rules.

A. III. (d) Finally there was 1 instance of this type which fell under neither (a), (b), or (c), referring instead to a substantive in the preceding sentence:

i.e. Wenn *eine reine Substanz* in zwei oder mehrere Substanzen von niedrigerem Gewicht zerlegt werden kann,

so nennt man sie eine Verbindung;

ergeben sich stets kompliziertere Substanzen bei ihrer Vereinigung mit anderen Stoffen,

und spaltet sie sich nie in einfachere Substanzen,

so heist sie ein Element.

If a pronoun were thus found to refer outside its own sentence, it could be treated in the same way as the exception in A. II. (a) above, that is as coming under Class B or C. The present instance in fact would be referred to its antecedent by rule B. II. (a).

CLASS A. IV. SUBSTANTIVAL CLAUSES

There were comparatively few examples of a nominative personal pronoun in this type of clause, so that little could be learned about the nature of the reference here. Those kinds occurring were as follows:

1. *clause (with or without dass) = direct object of preceding clause*

(a) The pronoun referred to the subject of the preceding clause (1 instance):

i.e. *Der Verfasser* sagt im Vorwort,

er beanspruche nicht,

irgend etwas wesentlich Neues zu bieten.

(b) The pronoun was not congruent with the subject of the preceding clause and referred instead to its dependent genitive (1 instance):

i.e. Oft zeigt schon die Prüfung *eines Körpers* mit dem blossen Auge,

dass er aus mehr als einem Bestandteil zusammengesetzt ist.

(c) The pronoun referred to the substantive in the initial prepositional phrase of the preceding clause: in 1 instance the preposition was *für*,

i.e. Für *diese Bahnen* können wir aber auf die gleiche Weise wie auf Seite 84 zeigenr,

dass sie sich alle in einem anderen Punkte P^1 der Ebene $x = x_0$ schneiden müssen.

in 5 instances *von* (2 of the substantives in this case being a relative pronoun)

e.g. *Man vermischt die fragliche Substanz *mit einer anderen,*

von der man vermutet,

dass sie mit ihr identisch ist,

* und bestimmt den "Mischschmelzpunkt."

(d) The pronoun was not congruent with any substantive in the preceding clause, containing as this did only the impersonal *man*, and referred instead to the direct or prepositional object of the clause (conditional) before this:

e.g. Fragt man am Schluss *nach den therapeutischen Anwendungen* von Chinonen,

so wird man einsehen,

dass sie noch sehr begrenzt sind.

2. *clause (with dass) = subject of preceding clause*

(a) In 3 instances there was no substantive in the preceding clause for the pronoun to refer to, and it referred in fact to a substantive in the clause before this - in 2 of the instances to a substantive in a prepositional phrase,

e.g. Bei der Analyse der Absorptionskurven gehen wir zunächst *von der nuclearen Komponente* aus,

da empirisch bekannt ist,

dass sie den Hauptanteil der Sterne liefert.

in 1 to the direct object,

i.e. Kristallisiert man *diese Formen*,

so zeigt sich,

dass sie zueinander spiegelbildlich sind.

(b) In 1 instance the pronoun was not congruent with the substantive in the preceding clause; it referred to the substantive in a prepositional phrase in the preceding sentence:

i.e. Auf die Frage nach *den Ursachen* des warmen Tertiärklimas soll hier nicht näher eingegangen werden.

Mir scheint jedenfalls,

dass sie trotz der Überlegungen von Kerner und Brooks nicht nur in der veränderten Land- und Meerverteilung und ähnlichen irdischen Faktoren gesucht werden.

A resemblance between this and the example under A.IV 1 (d) above may be observed: in both cases the construction is passive or virtually passive

(man) and impersonal, with the result that the substantive after *nach* acquires the importance of the subject function.

CLASS A. V. WHERE THE PRONOUN OCCURRED (EXCLUDING THOSE TYPES OF CLAUSE ALREADY MENTIONED) IN

1. an independent clause preceded by one dependent clause or by two, if co-ordinate,
2. a dependent clause preceded by an independent clause,
3. a dependent clause preceded by another dependent clause, the main clause following later.

- and provided that in each case the clause preceding that containing the pronoun possessed neither direct object nor predicative substantive.

If this preceding clause consisted of two co-ordinate predicates with a common subject, reference was made to the construction of the second predicate to decide whether the clause contained a direct object or predicative substantive.ϕ

A. V. (a) If the pronoun was congruent with the subject of the preceding clause, it referred to this:

e.g. (a) Obwohl *die zentrische Formel* in mancher Richtung unbefriedigend ist,

erklärt sie das Fehlen von Isomerie in den Ortho-Di-Substitutionsprodukten.

(b) *Zerfällt zum Beispiel *ein Stoff* nur,

wenn er am Katalysator adsorbiert ist,

*und ist bei normalen Bedingungen schon die Sättigungsgrenze der Adsorption erreicht,

so wird eine Erhöhung des Druckes keine Vermehrung der Zahl der zerfallenden Moleküle bedingen.

ϕ The reason for the exclusion in No. 1 above of two preceding dependent clauses, unless co-ordinate, is the difficulty encountered with such instances of determining to which of the two the pronoun in the independent clause refers. Similarly in No. 3 only those instances are taken into account in which the main clause comes after the dependent clause containing the pronoun, since in those instances where the main clause precedes the two dependent clauses the pronoun may refer either to a substantive in the preceding dependent clause or to one in the main clause before this.

This was true of 39 instances$^{\phi}$ not true of 4. These were

e.g. (a) Ein Gang ist meist *durch einen Fehler* in der Messanordnung bedingt,

auch wenn er innerhalb der Fehlergrenzen liegt.

(b) Die Rolle der "Zwischenstrahlung" kann natürlich auch *von der Primärstrahlung* übernommen werden,

indem sie ihre Qualität oder ihre Energie ändert.

(c) **Der Saft* der Bauchspeicheldrüse,

in dem das Enzym Trypsin enthalten ist,

* spaltet,

wenn er nicht zuvor mit dem Darm in Berührung kam,

* keine höheren, eigentlichen Proteine (Eiweissabbauprodukte);

+ erst wenn ein Extrakt der Darmschleimhart,

in dem die Enterokinase enthalten ist,

\+ dazugegeben wird,

spaltet er auch höhere Eiweisskörper.

(d) Die Namen *neu entdeckter Elemente* endigen im allgemeinen auf -um,

wenn sie Metalle sind.

In the first two the pronoun, though congruent, with the subject, referred to the agent of the preceding clause†, in the third to the subject of the preceding sentence, and in the fourth to the subject's dependent genitive (see (b)).

A.V.(b) If the pronoun was congruent not with the subject of the preceding clause, but with the dependent genitive or *von* substantive (= dependent genitive) of the subject, it referred to the latter (3 instances).

e.g. Die biologische Bedeutung *der Klopfsignale* dürte auf dem Gebiet der Fortpflanzung liegen,

indem sie eines der Mittel darstellen,

durch welche das Männchen sich als solches gelten läst.

ϕ 6 of these having a pronoun as the subject of the preceding clause, each with its antecedent determinable under the relevant rule.

$\dagger$ c.f. the exception in A.III (a).

There were 10 instances in which the pronoun did not refer to the subject of the preceding clause or its dependent substantive, having no congruence with either. These were of the following types:

A.V.(c) The pronoun referred to the agent of the preceding clause (2 instances);ø

A.V.(d) The pronoun referred to a dative object after *entsprechen* (1 instance);

A.V.(e) The pronoun referred to a substantive after the impersonal construction es handelt sich um... (1 instance);†

A.V.(f) The pronoun referred to a substantive in a prepositional phrase (2 instances):

e.g. *Bei den wirklichen Gasen* ist diese aber merkbar und zwar um so mehr,
je näher sie ihrem Kondensationszustand sind.

It may be noted that in each of these 6 instances the pronoun could refer for reasons of congruence to only one substantive in the preceding clause. In the other 4 instances the pronoun could not for the same reasons refer to any substantive in the preceding clause. They were as follows:

A.V.(g) The pronoun referred to the subject of the preceding independent clause but one:

i.e. *Diese Partikel* fallen ungefähr entlang den magnetischen Feldlinien auf die Sonne zurück,
und es besteht eine merkliche Wahrscheinlichkeit,
dass sie durch Kerntreffer Neutronen erzeugen.

A.V.(h) The pronoun referred to the subject of a clause included within the preceding clause:

i.e. *Die Leichtigkeit,
mit der sich *diese Stoffe* ineinander umwandeln lassen,
* spricht dafür
dass sie alle nur Modifikationen der einen Substanz, Wasser, sind.

A.V.(i) In the last 2 instances the pronoun referred to the subject (pronominal) of the last clause in the preceding sentence:

e.g. Man besass solche Tafeln schon im Altertum,
und zwar ist uns als erste die Sehnentafel *des Ptolemäus überliefert*,

---

ø c.f. the exceptions mentioned in A.III.(a) and A.V.(a)

† c.f. the remarks on the prepositional phrases in A.IV.2(b)

die er für astronomische Zwecke ums Jahr 150 n.Chr.
zusammengestellt hat.

Sie befindet sich in seinem Werke "Megale syntaxis,"

in dem er auch das nach ihm genannte Weltsystem entwickelt.

CLASS A. VI. *WIE* CLAUSES (NOT INCLUDED)

In the 4 instances where a pronoun occurred in a *wie* clause after a clause containing a direct object, it referred either to this:

e.g. An erster Stelle habe ich *diese Gruppe* zu nennen,

wei sie sich vor Erfindung der Logarithmen entwickelt haben.

or to its dependent substantive:

e.g. Diese Analogie läst uns vermuten,

dass es möglich sein wird,

auch mit Elektronenstrahlen die gleiche Art von *Abbildung* zu erzielen,

wie sie auf dem lichtoptischen Gebiete seit langem bekannt ist.

CLASS A. VII. WHERE THE PRONOUN OCCURRED (EXCLUDING THOSE TYPES OF CLAUSE ALREADY MENTIONED) IN

1. an independent clause preceded by one dependent clause or by two, if co-ordinate,
2. a dependent clause preceded by an independent clause,
3. a dependent clause preceded by another dependent clause, the main clause following later

- and provided that in each case the clause preceding that containing the pronoun possessed either a direct object or a predicative substantive.

A.VII. (a) If the pronoun was congruent with the subject, but not with the direct object or predicative substantive of the preceding clause, it referred to the subject. There were 11 such instances:

e.g. *Diese Potentialdifferenz* kann in einem lokalisierten Teil des Nervs die anliegenden Nerventeile reizen,

wenn sie stark genug ist.

In 2 of the 11 instances the subject referred to was itself a pronoun; both would have been referred to their correct antecedent by the relevant rule in each case.

A.VII. (b) If the pronoun was congruent not with the subject of the preceding clause, but with the direct object or predicative substantive, it referred to the latter:

e.g. Wenn man *Sacchorose* mit verdünnten Säuren erwärmt,

wandelt sie sich in gleiche Teile Glukose and Fructose um.

Though true of 13 instances, this clearly could not apply in the following:

e.g. Die obige Reaktion kann man ebenso gut erklären unter der Annahme,

dass *die Moleküle* von Wasserstoff und Chlor die gleiche gerade Anzahl von Atomen enthalten.

Es gibt jedoch gewichtige physikalische und chemische Beweise dafür,

dass sie zweiatomig sind.

A.VII. (c) If the pronoun was congruent with both the subject and the direct object or predicative substantive of the preceding clause, it referred to the subject:

e.g. Zum Beispiel können NH_4Cl *und* $FeCl_3$ Mischkristalle bilden,

obgleich sie nicht isomorph und bestimmt nicht chemisch ähnlich sind.

This was true of 6 instances, not true of 1:

i.e. Wenn die Schauer des 3. Maximums *zwei geladene Mesonen* enthalten,

sollten sie sich in der Nebelkammer unter geeigneten Bedingungen zeigen.

It should be observed, however, that in this case the pronoun reference was genuinely ambiguous. Only through the subsequent context did it become clear that the reference was to the object and not the subject.

CLASS A. VIII. DEPENDENT CLAUSE AFTER DEPENDENT AND MAIN CLAUSES

The pronoun occurred in a dependent clause following another dependent clause, but the main clause of the sentence came earlier than both instead of later as in A.V. The possibility existed, therefore, that the pronoun would sometimes refer to a substantive in the main clause or an earlier dependent clause rather than to one in the immediately preceding dependent clause.

A.VIII. (a) This in fact happened in all 3 instances where the pronoun stood in a relative clause:

e.g. Vielleicht liegt dies daran,

dass *die Verfasser* nicht auf die Korrektur achteten,

*welche nach der Theorie des Skineffektes den geraden Draht von der Spule,

die sie tatsächlich zu ihren Messungen benutzten,

* unterscheidet.

In each case the pronoun referred to the subject (or its dependent substantive) of the next but one clause preceding.

A.VIII.(b) Of the other 10 instances of this type all but one referred to some substantive in the dependent clause immediately preceding, the exception being:

Die Kristalle vieler Hydrate mit hohem Wasserdampfdruck haben das Bestreben,

Wasser zu verlieren,

wenn sie der Luft ausgesetzt sind.

where congruence indications pointed away from the immediately preceding clause to the one preceding this. The other 9 followed the rules of A.VII.

A.VIII.(c) If the pronoun was congruent either with the subject of the preceding clause or its dependent substantive (but not with both), it referred to this. There were 8 such instances, in 3 of them the subject referred to being also a pronoun. Reference of this to its antecedent would have been difficult in one case without consideration of meaning, or rather the human/non-human distinction:

i.e. Wir danken *den Herren* für die Übersendung von Separaten,

welche denselben Tagesgang aus Koinzidenzregistrierungen statistisch erfassten,

indem sie die Methode der Stichtage auf den Tagesgang anwandten.

Sie wählten....

A.VIII.(d) If the pronoun was congruent with the direct object and not with the subject (or its dependent substantive) of the preceding clause, it referred to this (1 instance).

CLASS B.I.

The pronoun stood in the first clause of a sentence, the last clause of the preceding sentence being independent and containing neither direct object nor predicative substantive.

B.I.(a) If congruent with the subject of the last clause in the preceding sentence, the pronoun referred to this:

e.g. *Sauerstoff* läst sich leicht durch Erhitzen von vielen sauerstoffhaltigen Verbindungen darstellen, so aus Quecksilberoxyd,

aus dem er zuerst von Priestley gewonnen wurde,

* weiter aus Kaliumchlorat und aus Kaliumperamanganat.

Für Handelszwecke wird er durch Wasserelektrolyse oder aus der Luft fabrikmässig hergestellt.

What may have become apparent from previous illustrations may be noted explicitly, namely that when only part of a clause comes at the end of a sentence (as in the above), the whole clause to which this part belongs is nevertheless to be regarded as the final clause for the purposes of pronoun reference.

The above rule applied to 65 instances, 9 of these having a pronoun for the subject referred to. In 1 of the latter this pronoun would not have been referred to its correct antecedent by the relevant rule. Apart from this there were 3 instances to which the above rule did not apply. In one this was due to the reference being to the agent in a passive construction instead of to the subject$^{\phi}$. In another the pronoun, though congruent with the subject of the last clause of the preceding sentence, referred not to this but to a subject in the sentence before:

i.e. *Die Opisthobranchiata* haben infolge Rückdrehung meist keine Chiastoneurie mehr

und die eine erhaltene Kieme liegt meist hinter dem Herzen.

Die Schale ist oft rückgebildet,

die Geschlechtsorgane sind zwittrig.

Sie leben im Meer, einige auch freischwimmend.

In the third the pronoun referred to a substantive in a prepositional phrase, or to its apposition. This was in the nature of a predicative substantive in this case:

i.e. An manchen Stellen sind die Lymphkanäle *zu stecknadelkopf – bis erbsengrossen Knoten* erweitert, den sog. Lymphdrüsen.

Sie bilden die weissen Blutkörperchen,

die hier den Giften und Bakterien entgegentreten.

B.I.(b) If congruent not with the subject, but with its dependent genitive or *von* substantive, the pronoun referred to the latter. In such cases the subject was generally a "mass" or other abstract noun;

e.g. *Anzahl, Gruppe, Menge, Art, Ordnung, Bestimmung, Bedeutung, Prüfung, Einfachheit.*

ϕ c.f. the exceptions under A.III.(a) and A.V.(a)

This was true of 5 instances but not of a further 1, where the reference was again to the agent in a passive construction - or rather to the dependent substantive of the agent:

i.e. Weitgehende Aufklärung dieser Gesetzmässigkeiten und Vorgänge, insbesondere in quantitativer Hinsicht, wurde durch die Untersuchung *der Atomspektren* der Elemente erhalten.

Sie haben den Charakter von diskontinuierlichen Linienspektren.

B.I.(c) In 1 instance the pronoun referred to the dative substantive (= subject) after the impersonal construction *es gelingt...*

B.I.(d) In another instance there was no sentence preceding that in which the pronoun stood, only the chapter title (VORPRÜFUNGEN), to which it in fact referred.

B.I.(e) In 5 instances the pronoun was congruent neither with the subject of the last clause in the preceding sentence nor its dependent substantive. It referred to a substantive in a prepositional phrase, though there was not in each case only one with which it was congruent. Where it was congruent with more than one, however, it referred to that in an "unattached" prepositional phrase.[ø]

e.g. Es läst sich auch hier das Potential in der Nähe des Achsenpunktes z_0 in *eine Reihe* nach y und nach $\Delta z = z - z_0$ entwickeln;

sie lautet V (z, y) = ...

B.I.(f) In 5 instances the pronoun could not, from lack of congruence, refer to any substantive in the preceding sentence. It referred instead to the subject (or its dependent substantive) of the sentence before this. Had there been any substantive in the last clause of the preceding sentence congruent with the pronoun, those instances would have formed exceptions to rule B.I.(a) like the second mentioned there.

B.I.(g) In 1 instance the pronoun referred to a substantive in a prepositional phrase, as in B.I.(e), but with this difference, that the reference was to the plural component of a singular compound noun, thereby creating some difficulty in establishing congruence:

e.g. Danach ist für diese *Nucleonengruppe* die Multiplizität je Stern in Blei höher als in Luft.

Sie sind aber nicht mehr zur Auslösung von Sternen fähig.

CLASS B.II

The pronoun occurred in the first clause of a sentence. The last clause of the preceding sentence was independent and contained a direct object.

ø i.e. unattached to a substantive.

B. II. (a) If the pronoun was congruent with the subject of the last clause in the preceding sentence, it referred to this:

e.g. *Die Schauer* des 3. Maximums enthalten also wahrscheinlich zwei geladene Teilchen;

daneben können sie auch eine gewisse Zahl n_0 von neutralen Teilchen enthalten.

There were 31 such instances, the subject in 4 being itself a pronoun. The antecedent of this in turn would have been identified by the relevant rule in each case.

B. II. (b) If the pronoun was congruent not with the subject, but with the direct object, it referred to the latter.

e.g. Der umgekehrte Vorgang, *die Spaltung* von Estern durch Wasser bezeichnet man mit verseifung:

sie ist eine Hydrolyse.

Though this applied to 11 instances (in one of which the direct object was a pronoun, but likewise causing no difficulty in reference to its antecedent), there was 1 to which it did not:

i.e. Die thixotrope Regelatinierungsgeschwindigkeit kann man oftmals durch *eine geeignete leichte Bewegung* erhöhen;

Sie ruft eine mechanische Koagulation der Teilchen hervor und heist Rheopexie.

This may be compared to the exceptions in A.III.(a), A.V.(a), B.I.(a) and (b), since the reference here is to what is really the agent in the virtually passive *man* construction.

B. II. (c) If the pronoun was congruent with neither the subject nor the direct object, but with the dependent substantive of the subject, it referred to this (1 instance).

B. II. (d) In 2 instances the pronoun referred to the dependent genitive or *von* substantive of the direct object. In one it was congruent with neither the subject, its dependent substantive, nor the direct object; in the other it was also congruent with the direct object, to which it would be referred by rule (b) on the previous page. The direct object in this case, however, was *Art*, and it might be asked whether a pronoun ever refers to such a noun when it has a dependent substantive and not rather to this substantive.

B. II. (e) In 2 instances the pronoun referred to a substantive in apposition to the direct object, not being congruent either with the subject or the direct object:

e.g. Das Element geht eine wohldefinierte Verbindung mit Wasserstoff ein, *den Schwefelwasserstoff.*

Er ist ein unangenehm riechendes Gas.

B.II.(f) In 1 instance the pronoun referred to the indirect object in the preceding sentence, again lacking congruence with both subject and direct object:

i.e. Die OH-Gruppe gibt *den Phenolen* einen saueren Charakter;

Sie können mit Basen unter Salzbildung reagieren.

B.II.(g) In 1 instance the pronoun referred to the substantive governed by the impersonal construction *man hat es zu tun mit...*[ϕ]

B.II.(h) In 1 instance the reference was to a substantive in a prepositional phrase (other than *von*) dependent on the direct object. There was no congruence with the latter or the subject of the sentence:

i.e. Wir mustern nun aus diesen Gesichtspunkten das experimentelle Material *über die Schauer* des 2. Max.

Nach Ziff. 7 und 10 werden sie durch geladene Teilchen ausgelöst, deren.....

B.II.(i) In 1 instance the pronoun could not refer to any substantive in the preceding sentence by reason of lack of congruence; its reference was in fact to the direct object of the sentence before this. There was, however, in the intermediate sentence a possessive pronoun referring to this direct object:

i.e. Für Gebirgshöhen haben Dallaporta, Puppi und Merlin *einen schwachen nuclearen Übergangseffekt* auch für Sterne in Blei errechnet.

Sein Vorhandensein in 2950 m Höhe können wir unseren Messungen nicht mit Sicherheit entnehmen.

Er kann in Kurve d (Abb. 2) enthalten sein.

CLASS B.III

The pronoun stood in the first clause of its sentence, while the last clause of the preceding sentence was independent and contained a predicative substantive.

ϕ c.f. A.IV. 2.(b).

B.III.(a) If the pronoun was congruent with the subject of the last clause in the preceding sentence, it referred to this:

e.g. Der angegebene Wert gilt allerdings nur bei kleinen Geschwindigkeiten,

denn *die Masse des Elektrons* ist keine konstante, sondern eine von dessen Geschwindigkeit abhängige Grösse;

sie wächst mit der Geschwindigkeit und wird bei Annäherung derselben an die Lichtgeschwindigkeit unendlich gross.

There were 32 such instances, 3 of these having a pronoun also for the subject referred to. In each case this would in turn have been referred to its antecedent by the appropriate rule.

B.III.(b) If congruent not with the subject, but with the predicative substantive, the pronoun referred to the latter:

e.g. $$\text{Das Verhältnis} \frac{\text{Lichtgeschwindigkeit im Vakuum}}{\text{Lichtgeschwindigkeit im Medium}} = \text{n heist}$$

der Brechungsindex einer Substanz.

Er ist von der Wellenlänge des Lichtes abhängig.

This was valid for 3 instances; not, however, for the following:

e.g. Die oben erwähnten Reaktionen sind die Grundlage des sogenannten Kontaktverfahrens für die Fabrikation der Schwefelsäure.

Sie wird auch nach dem Bleikammerverfahren technisch hergestellt.

where the reference was to the dependent genitive of the prepositional phrase dependent on the dependent genitive of the predicative substantive.

B.III.(c) In 1 instance the pronoun referred to the dependent genitive of the predicative substantive, being congruent neither with the latter nor with the subject of the sentence.

CLASS C.

The pronoun stood in the first clause of a sentence, and the last clause of the preceding sentence was a dependent clause.

This class, comprising about 10% of all the instances, was the least amenable to syntactical treatment for the solution of pronoun reference. The difficulty was to determine not only to which substantive but also to which clause the pronoun referred, since although the reference was more frequently (73%) to the main than to a dependent clause in the preceding sentence, the ratio was not sufficiently one-sided to permit those instances where reference was to the latter to be ignored. On the basis of analogy from the previous classes, however, it was to be expected that

reference would be primarily to the subject of a clause, and reliance on this for identifying the antecedent's clause as well as the antecedent itself produced the following results.

CLASS C.I.

The pronoun referred to a substantive in the main clause of the preceding sentence.

C.I.(a) Of all the subjects of the clauses in the preceding sentence the pronoun was congruent only with that of the main clause, or if also with a pronoun which was the subject of its clause, then this too referred to the subject of the main clause:

e.g. (a) *Die qualitative Analyse* einer chemischen Verbindung gibt Auskunft darüber,

welche Bestandteile: Ionen, Atome oder Atomgruppen, in ihr enthalten sind.

Sie zerfällt in drei Teile: ....

(b) *Freier Schwefel* findet sich in vulkanischen Gegenden,

wo er oft mittels überhitztem Wasserdampf geschmolzen und von Erde befreit wird.

Er kommt als Stangenschwefel oder in Pulverform als Schwefelblüte in den Handel.

There were 21 such instances. In 5 of them the subject of the main clause was itself a pronoun, but in each case the antecedent of this would already have been identified by the relevant rule.

C.I.(b) The pronoun was congruent either with no subject of a clause or only with a relative pronoun subject. It was, however, congruent with the dependent genitive or *von* substantive of the subject of the main clause, and if also with a relative pronoun subject, this referred to the former (2 instances).

C.I.(c) In 2 instances, where the pronoun was congruent - as in (a) - with the subject of the main clause and a relative pronoun subject, the latter did not refer as there to the subject, but to the direct object of the main clause. In one of the instances the pronoun referred to the subject, in the other to the direct object of the main clause:

i.e. (a) *Die Parabelmethode* gestattet eine Abwandlung,

welche sich von der Annahme "r klein gegen I" freimacht.

Sie wird deshalb überall dort mit Vorteil verwendet,

wo die Elektronenbahnen Umkehrpunkte aufweisen.

(b) Die dazwischenliegenden Elemente habe *restliche Elektronen*,

die nicht zu vollständig abgeschlossen, stabilen Gruppen vereinigt sind.

Sie üben Kraftwirkungen nach aussen aus und sind Einflüssen von aussen zugänglich;

C.I. (d) In 1 instance the pronoun was congruent not with the subject of the main clause, but with those of two following co-ordinated relative clauses, the relative pronoun itself being the subject in the second. It was to this, and through it to the direct object of the main clause, that the pronoun referred.

Neuerdings hat man *Elemente* gefunden,

deren Kernladung und Masse zwar gleich sind,

die aber eine verschiedene Anordnung der Protonen und Neutronen im Kern besitzen.

Sie heissen Kernisomere . . .

It may be noted that here the direct object is really the subject of a virtually passive construction[ø] and that therefore the instance approximates to the C.I. (a) type.

C.I. (e) The pronoun was not congruent with the subject, dependent substantive of the subject, or direct object of any clause in the preceding sentence. It referred to the dependent genitive of the direct object in the main clause (1 instance):

i.e. Im Ausgleich dieser Ladung hinter der Erde vermuten Chapman und Ferraro das Zustandekommen *des Ringstroms*,

dessen Natur und Aufbau keineswegs geklärt sind.
Er scheint so hohe Stabilität anzunehmen, dass...

C.I. (f) Not congruent with the subject of the main clause the pronoun referred to the relative pronoun subject of the intermediate clause, and through this to the predicative substantive of the main clause (1 instance):

i.e. Ke = -eE ist *die Kraft*,

welche das elektrische Feld auf das Elektron ausübt.

Sie ist der elektrischen Feldstärke entgegengerichtet.

C.I. (g) Though not congruent with the subject of the main clause, the pronoun was congruent with the subject(s) of the intervening clause(s). These subjects, being pronouns, referred in one instance to the indirect, in the other to the direct (dative) object of the main clause:

ø c.f. B.II.(b).

i.e. *Die klare Erkenntnis,

dass es sich hier wirklich um ein ganz neues astronomisches Problem handelte,

* und die ersten Ansätze zur Vorbereitung seiner Lösung verdankt die Wissenschaft *William Herschel*,

der zu Ende des 18. Jahrhunderts als erster mit Hilfe des von ihm erbauten grössten Spiegelteleskops seiner Zeit den bewussten Versuch unternahm,

die Gestalt des unsere Sonne umgebenden Sternsystems auf Grund von Beobachtungstatsachen zu ermitteln.

Hierbei bedient er sich einer Reihe von an verschiedenen Stellen ausgeführten Sternabzählungen.

and

Wir danken *den Herren* für die Übersendung von Separaten,

welche denselben Tagesgang aus Koinzidenzregistrierungen statistisch erfassten,

indem sie die Methode der Stichtage auf den Tagesgang anwandten.

Sie wählten.....

C.I. (h) The pronoun was not congruent with the subject of the main clause, but was congruent with a relative pronoun subject, this referring to the agent in the main clause (1 instance):

i.e. Eine andere Methode zur Auffindung der Wertigkeit von Elementen wurde durch *Mitscherlich* eingeführt,

der eine Ähnlichkeit zwischen den Kristallen analoger Verbindungen entdeckte.

Diese Erscheinung ernannte er Isomorphie.

C.I. (i) The pronoun was not congruent with the subject of any clause, the reference being in fact to a substantive in a prepositional phrase in the main clause (1 instance):

i.e. In diesen Arbeiten von *F. Tödt* finden sich auch Angaben,

die geeignet scheinen,

jene Frage nach dem Mechanismus der Korrosion durch einen unmittelbaren Befund aufzuklären.

Er teilt nämlich mit, dass ...

C.I.(j) The pronoun was congruent both with the subject of the main clause and with that of the intervening dependent (relative) clause. It referred, however, to the prepositional relative pronoun, and through this to the direct object of the main clause (1 instance)

i.e. Die photochemische Chlorknallgasreaktion besitzt
eine "Induktionsperiode" zu Beginn der Bestrahlung,
während der keine Reaktion stattfindet.
Sie ist durch die Anwesenheit von Verunreinigungen bedingt.

CLASS C. II

The pronoun referred to a substantive in a dependent clause of the preceding sentence.

C.II.(a) Of all the subjects of the clauses in the preceding sentence the pronoun was congruent only with that of a dependent clause, or if also with a relative pronoun subject, this referred to the former:

e.g. Wer heute in Westdeutschland eine Lampe anknipst,
kann nicht mehr sagen,
woher *die Energie* stammt,
die sie zum Glühen bringt.
Sie kann ebensogut aus dem österreichischen Wasserkraftwerk
Bludenz wie aus den grossen Moorkraftwerken an der
holländischen Grenze kommen.

There were 7 such instances.

C.II.(b) Not congruent with the subject of any clause nor with a dependent substantive of the subject in the main clause, the pronoun referred to the dependent *von* substantive (or its apposition) of the subject in the intermediate dependent clause (1 instance):

i.e. Es ist ein unangenehm riechendes Gas,
von dem sich eine Reihe *von Salzen, die Sulfide,*
ableiten;
sie sind meist in Wasser unlöslich ...

C.II.(c) The pronoun was not congruent with the subject (or its dependent substantive) of any clause, nor with the direct object of the main clause, but referred to that of the dependent clause (1 instance):

i.e. Jedenfalls haben wir ein Interesse daran,
die sphärische Aberration nach Möglichkeit zu verringern;
ganz beheben läst sie sich ... im Gegensatz zur Lichtoptik nicht.

C.II.(d) The pronoun was not congruent with the subject of the main clause, congruent with the relative pronoun subject of the intermediate clause, which in turn referred to the direct object of the main clause. According to the rules of C.I. the pronoun would have been referred in such a case to this relative pronoun - and thereby to the direct object of the main clause, whereas in fact its reference was to the direct object of the relative clause (1 instance):

i.e. Die zweite Gruppe enthält die energieärmeren
Nucleonen $S-107 < E < S.10^8$ eV,

die noch Sterne auszulösen vermögen und in der photographischen
Platte *die sogenannten "grauen" Spuren* bilden.

Soweit sie aus mehrspurigen Sternen hervorgehen,

ist ihre Winkelverteilung weitgehendisotrop.

C.II.(e) Though not congruent with the subject of the main clause, the pronoun was congruent with the subject of the three succeeding dependent clauses, this in one being a relative pronoun referring to the previous subject. It was to this relative and its antecedent that the pronoun referred (1 instance)

i.e. Die Beobachtung lehrt,

*dass *parallele Lichtstrahlen*,

die zur Hauptachse parallel auffallen,

* durch die Brechung in der Konkavlinse divergent
gemacht werden, und zwar so,

dass ihre Rückwärtsverlängerungen durch einen Punkt der
Hauptachse gehen;

sie scheinen also von diesem Punkt aus zu divergieren.

C.II.(f) The pronoun was congruent with the subject of the main clause and with the subjects of the two succeeding dependent clauses, the latter of these being a relative pronoun referring to the former. As in C.II.(e), the pronoun referred to this relative pronoun and its antecedent (1 instance):

Diese Kurven sind dadurch charakterisiert,

dass an ganz bestimmten Stellen des Spektralbereichs *selektive
Absorptionsbanden* mit ausgeprägten Maxima liegen,

die durch ihre spektrale Lage, ihre Höhe und ihre
Halbwertsbreite gekennzeichnet sind.

Sie sind bestimmten Atomen oder Atomgruppen in
Molekül, sogenannten Chromophoren, einzuordnen.

CONCLUSION

The nature of the syntactical items referred to and the number of pronouns out of a total of 392[ø] referring to each was as follows (column A.):

	A	B	C	D	E
Subject	296	283	10	2	1
Direct object	38	33	2		3
Substantive in prepositional phrase	22	18	2		2
Dependent substantive of subject	16	14	1		1
Agent	7	2			5
Dependent substantive of direct object	4	3			1
Predicative substantive	3	3			
Apposition of direct object	2	2			
Indirect object	1	1			
Dependent substantive of agent	1	0			1
Dependent substantive of prepositional phrase	1	0			1
Dependent substantive of predicative substantive	1	1			

Correct reference, however, was not possible in all cases; Under B are given the number of instances in which the pronoun was congruent with and referred to an antecedent in its expected location (i.e. the immediately preceding clause or sentence) according to the relevant rule.

In some cases the relevant rule failed to find a substantive congruent with the pronoun in question in the expected location, but a switch to the next relevant rule (c. f. for example A. II. b) succeeded in identifying the correct antecedent. Such instances are shown under column C and were of the following types:

Subject A. II. b, A. III. d, A. V. g, A. V. h, A. V. i, B. I. f-4.

Direct object A. IV. 2. a, B. II. 1.

Substantive in prepositional phrase A. IV. 2. a, A. IV. 2. b.

Dependent substantive of subject B. I. f.

ø The 24 instances of class A. I. are excluded, the pronoun there referring without exception to the nearest congruent noun before the included clause.

On the other hand there were one or two instances of this nature where the relevant rule did find a congruent substantive in the normal location, with the result that the pronoun was referred to the wrong antecedent, the true antecedent being in an abnormal (i.e. earlier) location. These are given under column D. They were of the following nature:

A.V.a. and B.I.a. - the pronoun was referred to the subject of the immediately preceding clause or sentence, but actually referred to that of the sentence before this.

Finally there were occasions where a particular rule, adapted to the majority of instances, met with an exception and made an error in reference. These are listed under E alongside the syntactical item to which they ought to have been referred. In fact, most of them were referred - as would be expected - to either the subject or the direct object of the clause. The types were as follows:

Incorrectly referred to the subject; actually referred to:-

the direct object: A.VII.c, C.I.j, C.II.d.

the agent: A.III.a, A.V.a.(2), B.I.a.

a substantive in a prepositional phrase: B.I.a, B.I.g.

the dependent genitive of the subject: A.V.a.

Incorrectly referred to the direct object; actually referred to:-

the subject (of the preceding sentence): A.VII.b.

the agent: B.II.b.

the dependent substantive of the direct object: B.II.d.

Incorrectly referred to the predicative substantive; actually referred to:-

the dependent genitive of a prepositional phrase dependent on the dependent genitive of the predicative substantive: B.III.b.

Incorrectly referred to the dependent substantive of the subject; actually referred to:-

the dependent substantive of the agent: B.I.b.

The number of errors, then, produced by the above in connection with the texts concerned would have been 19; that is, those listed under D and E plus two others, the antecedent of which though correctly identified was itself a pronoun, and the antecedent of this would not have been identified correctly. This means something over a 95% correct solution. Naturally, being adapted to a particular set of instances, the rules produce a better result than would be obtained, if they were applied to an entirely fresh set.

Nevertheless, syntactical methods may always be expected to solve a fairly high percentage of pronoun references - or at least so one would assume from an experiment carried out in the course of compiling the present data. For after half of the instances had been collected, the rules were formulated on the same lines as the above to cover as many of them as possible. Then, when the other half had been collected, these rules were applied to them to see in how many cases they would solve the pronoun reference. The result was 83%.

It would seem reasonable, therefore, to assume that with a more detailed and reliable set of rules, such as might be expected from further investigation of the problem, a solution in the region of 90% would be likely.

A complete solution of the problem by syntactical considerations alone is not possible, since examples exist which, though identical in syntactical structure, nevertheless differ in their pronoun reference. In such cases the reference rule can only be chosen on a basis of attempting to cover the greatest possible number of instances, thereby at the same time necessitating that in the "exceptions" the pronoun will be referred to the wrong antecedent. That the translation also will be wrong does not automatically follow. This, as was remarked at the beginning, will arise only from confusion of the human/non-human distinction in a context mentioning persons as well as other animate and inanimate things.

The possibility of a more complete solution to the problem lies in a combination of syntactical methods with a system of word classification. The latter has been found in the course of past investigations to be essential for solving the residue of most problems after application of syntactical methods. Often an extremely detailed classification appears to be required, but in the present instance a very broad one would suffice, such as a division of the words in the dictionary, especially verbs and adjectives, according to whether they represented or were characteristic of a human animal, a non-human animal, an animate object, or an inanimate object. Of course, some would be characteristic of more than one, and then perhaps more so of one than another; such things would have to be taken into account.

In some instances it would then be possible to ascertain whether a pronoun represented a human or non-human being or thing simply from the immediate context (e.g. the nature of the verb, a predicative adjective or substantive, or a prepositional phrase). For the majority, however, the classification would need to be used in conjunction with syntactical rules, although these would have a somewhat different function than before. Instead of making a final decision about the identity of the pronoun's antecedent they would serve merely to indicate the various possible antecedents from the point of view of congruence and usage. The appropriateness of each antecedent in the pronoun's context would then be tested. If this

scheme were used, the only instance where an error could occur would be where there were two (or more) possible antecedents, one human, one non-human, and each was found to fit equally well into the context of the pronoun. How often such instances occur - if at all - without also being ambiguous in the German is a matter for further investigation. They would obviously be least likely when the other antecedent besides the human one belonged to the class of inanimate object, most likely when it belonged to that of non-human animal. If instances did occur where the context provided no criterion for distinguishing between the possible antecedents, the syntactical rules could be used as a final resort in the original manner described above with a 9:1 chance in general of selecting the correct antecedent.

A much simpler and better solution from the point of view of economy in the translation program would be to re-adopt in principle the idea of the pre-editor, though putting the responsibilities on the writer of the article or book rather than on an editor in the usual sense. In short, prescriptive grammar. With pronoun reference, as with most problems in mechanical translation, 90% of the instances follow a regular pattern. Why should not the other 10% do so too? In most cases a small alteration would have sufficed, as for example the conversion of a passive to the corresponding active construction (c.f. Class A.III.a.):

Der Endpunkt der Titration wird durch eine kleine, zu der Lösung zugegebene Menge *eines Indikators* engezeigt,

er ändert die Farbe,

je nachdem die Lösung sauer oder alkalisch ist.

becoming

Eine kleine, zu der Lösung zugegebene Menge *eines Indikators* zeigt den Endpunkt der Titration an,

er ändert die Farbe, ...

which would have been correctly solved under A.III.b.

Once aware of the existence of such difficulties for mechanical translation an author could quite easily avoid raising them. The only question mark against such prescriptive grammar concerns the number of rules an author might be expected to memorise and apply to his compositions. If they were to grow so numerous as to cause considerable inconvenience in writing, a professional pre-editor would become a necessity, and if - as some believe - such extraneous assistance to mechanical translation is to be avoided, then it would be advisable to resort to the prescriptive grammar solution only with the most intractable problems.

ACKNOWLEDGEMENT

This work was supported in part by the U.S. Army (Signal Corps), the U.S. Air Force (Office of Scientific Research, Air Research and Development Command), and the U.S. Navy (Office of Naval Research); and in part by the National Science Foundation.

DISCUSSION ON THE PAPER BY DR. L. BRANDWOOD

MR. R. H. RICHENS: Dr. Brandwood's descriptive syntactical analysis of selected German sentences will be of value, not only to those working on the mechanical translation of German, but also to those working on other languages where analogous problems arise. However, the results we have been given raise in an acute form the problem of the relative importance of syntactic and semantic analysis in machine translation. Dr. Brandwood has shown that the antecedents of the German pronouns that he has studied cannot be located by any simple syntactic procedure; it was necessary to devise a considerable number of probability rules, most of which have exceptions with appreciable frequency. The human translator does not trouble with such rules since the antecedent is located by semantic congruence and it is important therefore to integrate both syntactic and semantic analytic procedures into any machine translation programme with pretensions to generality. The methods being investigated by the Cambridge Language Research Group for semantic analysis are the thesaurus method and a technique involving three types of analysis of interlingual formulae. Dr. Brandwood tentatively resurrects the notion of the pre-edited text. I cannot follow him here, since, leaving various impracticabilities aside, the interest of machine translation resides largely in attacking rather than minimizing the problems inherent in existing natural languages.

MR. E. A. NEWMAN: I thought it worth while to mention what seemed to me an example of the very great importance of context in translation. We saw an example of translation this morning which might have been called English to English, English to Russian, Russian to Russian, Russian to English, English to English, and we noted, many of us, I think, that the English to Russian and Russian to English was very easy, but the English to English and the Russian to Russian showed signs of having a lot of difficulty. This does rather illustrate that context is of very great importance.

PROF. Y. BAR-HILLEL: We heard this afternoon two examples of investigations into machine translation which in a certain sense lay at opposite ends of this research. None of these investigations, I would say, is representative of much of what is going on in machine translation. But it is, of course, not my task to fill in the gap.

Though the investigations we listened to are then not typical of MT research in general, they are indicative of a certain disappointment in the

achievements of MT in the last years. This disappointment seems to me a necessary reaction to the original sin of having staked the aims too high. Occasionally, however, the despair has yielded strange fruit. From the fact that we still don't have a working programme for translation between one pair of languages that would be both fully-automatic and high-quality, some drew the conclusion that one must now aim for fully-automatic high-quality translation between any pair of languages. Others drew the conclusion that one has to go into extremely minute investigations of minute occurrences. This is indeed necessary so long as fully-automatic high-quality translation is still the aim. Nothing less than what Mr. Brandwood has given us will suffice for solving one of the million problems which one could cite in this connection. But clearly what Dr. Brandwood did is not even enough to solve the problem he set himself. He went with extreme care over certain samples in order to derive from their analysis rules that would fit these samples hoping that they would fit in general. I don't think there is any good reason for this hope. Nobody who knows German will find any difficulty in concocting counter-examples to any of the rules which Dr. Brandwood derived from his analysis. Dr. Brandwood himself apparently made no such attempt, misled by a rather frequently found "empirical approach" which distrusts the rules found in the grammar books - probably because they were given by normatively-minded grammarians - and hopes to arrive at better rules from analysis of a large enough corpus.

I regard this approach as enormously wasteful. To throw away insights of the best grammarians (only because you were unable to derive from these insights a working method of MT and had therefore to find a scapegoat on whom to put the blame, instead of concluding that your original aim was too high!) and to start from scratch seems to me extremely bad policy. But, worse than that. The refusal to put the "derived" rules to test by cooking up counter-examples, and the concomitant decision to wait instead for a counter-example to come up by itself in some future sample, is a caricature of the empirical approach and definitely contrary to the standard scientific method which, if only possible, dispenses with observation in favour of experimentation.

The decisive question whether fully-automatic, high-quality translation is a reasonable aim is not one whose answer should be dependent on one's temperament. I am convinced that this question must and can be decided on a purely scientific level. I think I can prove that this aim is totally unrealistic on the basis of machinery available or under construction today, and of programming techniques in use or in prospect today. I would not dare to say anything about machines of which we do not even have blueprints, and of programming techniques of which nobody knows even the principles today.

I do not think that Chomsky's recently obtained deep insights into the structure of languages (*ref. 1*) are at all helpful to machine translation.

Their major impact on MT is rather to show that mechanical syntatical analysis is incomparably more difficult than linguists had believed before. (For an elaboration of this evaluation, see my papers, *refs. 2* and *3*).

Let me finish my comments with a challenge. I defy anybody to give me any *non-ad-hoc* rules which will provide for the correct machine translation from English into French of both the following two sentences: "The pen is in the box" and "The box is in the pen". Notice that within a suitable longer context - and I leave it to you to imagine the two contexts - the translation will be unique for anybody who knows sufficiently well English and French. Now, would someone volunteer to translate the sentences into French?

DR. L. MEHL (CHAIRMAN): *"La plume est dans la boite"* and *"La boite est dans la plume"*. (Laughter).

You see the inconvenience of having a chairman who does not speak English perfectly! He translates like a machine.

PROF. BAR-HILLEL: Now this is exactly what I expect a machine to come up with, if it must come up with a unique translation. Dr. Mehl gave us these translations because he worked for a minute like a machine and because he might not know the other meanings of 'pen' not being a native English speaker.

DR. L. KOWARSKI: I think your example is probably even worse than you expect, because a possible French word for 'pen' in the second sense is *'box'*. *"La boite est dans le box."*

PROF. BAR-HILLEL: Well, I am surprised. I indeed did not know that the correct French rendering of 'pen' in the second sentence is *'box'*. Now, why is it the case that you, Dr. Kowarski, who knows English and French well had no difficulty in finding the correct renderings? Knowledge of the two languages is not enough! You had to use, in addition, your knowledge of the world acquired in many years of learning, schooling, and observation. You knew from experience and not only from your knowledge of the languages, that the entities denoted by *'plume'* are generally smaller than the entities denoted by *'boite'* whereas the entities denoted by 'box' (in French) are usually larger. It was this empirical knowledge of the relative sizes of certain objects, a knowledge which it is not very difficult for a

REFERENCES

1. CHOMSKY, N. Syntatical Structures, *s'Gravenhage* (1957).
2. BAR-HILLEL, Prof. Y. Decision method for structure in ordinary language. *Logique et Anlyse*, 1959, 2.
3. BAR-HILLEL, Prof. Y. Some linguistic obstacles to machine translation. *Proc. of the Second International Congress on Cybernetics, Namur* (1958).

human to acquire, I admit, which was a necessary condition for the uniqueness of the translation. I know of no method, surely not the Thesaurus method advocated by the Cambridge group, which would enable a machine to come up with the correct translation and not be completely *ad hoc*. It is pointless to believe that a machine will have built into it the knowledge of all size relationships of all the objects in the world. A machine, however, that will be able to *acquire* this knowledge will surely be of a degree of complexity of a human being, and the point of having machine translation would be lost altogether.

If my challenge is left unanswered, then the aim of high-quality, fully-automatic translation must be considerably lowered. I admit that thereby the great attractiveness of the problem will be reduced, for me and for other people. However, those people who are interested in MT not as a purely intellectual problem but as something of great practical value will do best if they reduce their aim to either fully-automatic but low-quality or else to high-quality but only partially-automatic translation.

DR. A. M. UTTLEY: I have never done anything on machine translation, but seeing this problem I would like to write something more on the board. "The spring is in". Now I might ask the mechanical translators to translate that. I think their reply would be that that is not fair.

PROF. Y. BAR-HILLEL: Is that a full-stop?

DR. UTTLEY: No. They would say: "That is not fair, let me have a little more context. I do not know whether this is a spring in a lock, or a spring in a garden". I would say the same thing of the first two sentences and that given a little more context, a translation could be made mechanically by non ad hoc methods.

MR. P. E. TRIER: I believe the author has set himself a very difficult problem in the codification of these pronoun references for machine translation, since even he himself has given several ambiguous or wrong references. On page 320 at the top: "Ein Gang ist meist durch einen Fehler in der Messanordnung bedingt". That would refer at least as probably to the word "Gang" as to the word "Fehler."

In the middle of page 322: "Diese Analogie läst uns vermuten"- "sie" refers to the "Art" and not "Abbildung".
Page 323- "..... when he talks about the 'maximum' he does refer himself to the ambiguity which is quite unresolvable within the sentence that is given.

Lastly on page 324, about the middle "Wir danken den Herren für die Übersendung von Separaten, welche denselben Tagesgang aus Koinzidenzregistrierungen statistisch erfassten." Here "welche" must refer to

"Separaten" and not to "Herren. With these genuine difficulties, I believe the machine is going to have quite a problem.

PROF. Y. BAR-HILLEL: No, it is not the missing context that does not allow the machine to translate correctly my two sample sentences. Even had I embedded the two sentences in two suitable paragraphs, the machine would still have been unable to decide that in the one case 'pen' has to be rendered by *'plume'* and in the other by *'box'*, except by an *ad hoc* procedure. However, as soon as someone will come up with a procedure, according to which the machine will be able to resolve the ambiguity, I commit myself to concoct in fifteen seconds a simple, nothing special, counter-example for which this procedure will fail. (This is the reason why I use the term *'ad hoc'.)*

By the way, there exists to this day no procedure which takes into account larger-than-sentence contexts. But though such contexts mean a new order of magnitude of complexity, I am quite ready to agree that this extension will have to be tackled if one wishes to increase the machine part in the whole translation process.

MR. R. H. RICHENS: Professor Bar Hillel is misinformed in supposing that no procedures exist for dealing with ambiguities of the type he mentions. Furthermore, in the procedure in which I am most interested, the paragraph and not the sentence is the field for semantic comparison. Briefly, this procedure involves translation of all word segments into the range of interlingual semantic formulae which each potentially represents. Two principal operations are then performed on these expressions: (1) testing the semantic congruence between the terms of dyadic relations and the relations themselves; and (2) production of a semantic spectrum for the entire paragraph and using this for making decisions for such ambiguities as remain to be resolved. This procedure is fallible but gives a correct solution in an as yet unknown number of cases; it is mathematically definite.
Prof. Bar Hillel quoted two sentences devoid of context. They would not be expected to occur in the same paragraph except in a paper dealing with difficulties in machine translation. If they were embedded in likely contexts, they would not appear to offer insuperable difficulties to existing semantic analytic procedures.

PROF. Y. BAR-HILLEL: As a matter of fact, I discussed my sample sentences with other members of the Cambridge Group, Miss Masterman, Mr. G. Parker-Rhodes and Mr. Needham, less than two weeks ago in Washington, and they were unable to present a procedure that would provide for a correct translation of both sentences. By the way, I did not claim at all that both sentences have to appear together in some far-fetched story. I am quite

satisfied with dealing with either sentence at a time. I still claim that a machine will have no way of deciding from the given sentence, in many if not all contexts of paragraph length, that 'pen' means *'plume'* rather than *'box'*. I presented you with two sentences only for pedagogical reasons. The members of the Cambridge Group in Washington sat for some time to cook up a Thesaurus which would resolve the ambiguity, without success. And I think that Miss Masterman at least was quite ready to admit at the end that a Thesaurus would not resolve all semantical ambiguities. I do not know what the other members of the Group believe in this respect.

DR. L. MEHL: If we consider legal questions, it is possible that in a certain passage of such a text, the same word can be taken to have a different meaning. Your example is quite obvious here, but in juridical text, for example, ambiguous words are relatively frequent, and the automatic translation of juridical texts would be very difficult. On the question concerning the translation of "pen" you say "box" in French. I could have said "enclos" or "parc" just as well as "box" because the French "box" means "garage" or "parking". But for horses "box" is correct.

DR. KOWARSKI: By the way, there is another translation of "The box is in the pen" in French. It would be "Le combat a lieu au pènitentiaire."

MR. J. F. COALES: I would like to make one point that seems to me relevant, that both these sentences are essentially ambiguous, ambiguous in English. You can take "pen" and "box" in either way and the sentence makes sense in a certain context; which to choose is really a question of probability. You cannot translate an imprecise sentence into a precise sentence, and I suspect such a case should be treated on a probability basis, which I think the Thesaurus method does in fact do. It must take into account the size of the pen versus the box, and the question of whether it can be small or large then has some effect on how you translate it.

DR. L. MEHL: It is clear in some of these examples that the probability of the appearance of these two sentences in the same passage is highly unlikely, but in fact I have experience of juridical texts. Having written a book concerning tax questions, I made an alphabetic index for the book, and I noticed that certain words appearing in the same place in the alphabetic index because they were identical in form were totally different in meaning in the same book. I think it confirms your own observations concerning the difficulty of automatic translation.

DR. L. BRANDWOOD (in reply): Mr. Trier has kindly pointed out several "errors". He is correct in saying that on page 320 the pronoun reference could be to *Gang*

just as easily as to *Fehler* - in which case the example would not be an exception, but would comply with rule A.V.a.

He is not correct in saying that on page 322 *sie* refers to *Art* instead of *Abbildung*. This may be most easily seen by subtracting each of the two words in turn from the sentence, thus

(1) ..... auch mit Elektronenstrahlen die gleiche Abbildung zu erzielen, wie sie auf dem lichtoptischen Gebiete seit langem bekannt ist.

conveys practically the same information as the original sentence, whereas

(2) ..... auch mit Elektronenstrahlen die gleiche Art zu erzielen, wie sie auf dem lichtoptischen Gebiete seit langem bekannt ist.

has an entirely different sense.

Mr. Trier is also incorrect when he declares that on page 324 the pronoun must refer to *separaten* not to *Herren*. That the *welche* does in fact refer to *Herren* is indicated by the following *indem* clause, which must have a common subject with *erfassten*, which it explains, and one could hardly say of *Separaten* that they *die Methode ..... anwandten*. Mr. Trier is apparently not acquainted with the fact that in German a relative pronoun may refer to a syntactical unit other than that which it immediately follows.

Prof. Har-Hillel's discourse is at first sight a plausible attempt - for reasons one can but guess at - to discredit all the work done on mechanical translation in this country. On further consideration it can be seen that there is little logic and even less fact to substantiate his remarks. These may be taken in the order in which they were made.

First there is the statement about my investigating samples in order to find rules that would fit these samples in the hope that they would also fit in general. "I don't think", Prof. Har-hillel says, "there is any good reason for this hope". Apparently he has not read or chooses to ignore the fact mentioned on page 337 of the paper, that the set of rules derived in this way covered over 80% of the instances in a fresh sample. Prof. Bar-Hillel's *belief* to the contrary is of no great value. As for the subsequent "Nobody who knows German will find any difficulty in concocting counter-examples to any of the rules which Dr. Brandwood derived from his analysis", we might ask who would wish to deny this seeing that the investigation was described solely for the purpose of illustrating work in progress and no claim was made to have arrived at a complete solution to the problem. Even so, any such "concocted" counter-examples would fall under the 10-20% not accounted for by those rules with the widest application, and would therefore represent what would be - on the existing evidence - the less common grammatical constructions.

Prof. Bar-Hillel is quite right in assuming that I myself made no attempt to "concoct" counter-examples. It is disputable, however, whether he is also correct in saying that I was "misled by a rather frequently found 'empirical approach' which distrusts the rules found in the grammar books -". If a complete system of a language's grammar is to be obtained, then it is

essential to use the method of examining texts for the simple reason that no individual is acquainted with the whole, merely with a part of his native tongue; he uses not all the available vocabulary and grammatical constructions but only a personal selection. Consequently he cannot "concoct" all the possible exceptions to a grammatical rule derived from the observation of a number of instances in a sample of text. It will probably be generally agreed that in reading one not infrequently comes across a grammatical construction which strikes one as completely different from one's own usage. A rough-illustration might be provided by reference to one of Prof. Bar-Hillel's statements, though the unusualness in this instance is due to incorrect rather than just variant English constructions: "I defy anybody to give me any *non-ad-hoc* rules which will provide for the correct machine translation from English into French of both the following two sentences: 'The pen is in the box' and 'The box is in the pen'. Notice that within a suitable longer context ..... the translation will be unique for anybody who knows sufficiently well English and French." Although the pleonastic "both" and "two" in the first sentence and the position of the adverb between the verb and its object in the second are new to me - I should never have been able to "concoct" them - the sentences exist and it is perhaps conceivable that someone might wish to translate them, together with the rest of the discussion, into German. In this event allowance would have to be made for such constructions in the programme. One would not want to find such phases in the output as *beiden der beiden Sätze.*

Furthermore, to say - as Prof. Har-Hillel does - that the empirical method "distrusts" the rules found in grammar books is ridiculous. The truth is that, while the mechanical translator would dearly like to use the rules formulated by traditional grammars, he cannot, either because they are not sufficiently precise for his purpose or because they are non-existent for the problem with which he is dealing. Let Prof. Bar-Hillel try to find rules or "insights" bearing on the problem of pronoun reference even in the larger German grammars; he will discover that they are conspicuous by their almost complete absence. Why else would an eminent professor of German philology Professor B. Ulvestad of Bergen University, be investigating this problem at the present time?

Prof. Bar-Hillel's emphasis on the sin of proceeding beyond *traditional* grammar or of employing the empirical approach, "contrary to the *standard* scientific method", and his insistence, based on a consideration of *available* electronic computers, that "fully automatic high-quality translation is an unreasonable aim, make it abundantly clear that with his inability to see beyond the present state of progress Prof. Bar-Hillel is the eternal pessimistic conservative, the scientist who before Galileo's observations thought it impossible that the earth should travel round the sun, the military expert who, with the Maginot Line in mind, asserted

that France was impregnable - until Hitler's paratroopers overran it in a matter of weeks, the man who before the Wright brothers gave their demonstration was incredulous of the possibility of flight by machines heavier than air and who even afterwards could see no further than his nose and scoffed at the notion that a later development of this same machine might cross the Atlantic, not to mention reach the moon.

As regards Prof. Bar-Hillel's challenge he is not taking must of a risk in "defying" anyone to produce at such short notice non-ad-hoc rules to translate his two sentences correctly. The problem of multimeaning is one if not *the* most difficult problem of mechanical translation, and it would be foolish to deny that the translation of the type of sentence quote by Prof. Bar-Hillel causes some difficulty. On the other hand it would be equally foolish to deny, as he does, that there is any possible *non-ad-hoc* solution. He is mistaken, moreover, in thinking that an "empirical knowledge of the relative sizes of certain objects" is a "necessary condition for the uniqueness of the translation". This knowledge is in fact derivable from the sentences themselves, since the subject of the verb "to be in" must be smaller in size than the object governed by the verb, at least when both represent material things and can be spoken of as having size. Knowledge of relative size alone is not sufficient to permit a unique translation here, since - as Mr. Coales observes - both sentences are completely ambiguous as they stand. Thus "the pen is in the box" might be translated as

(a) la plume est dans la boite.

(b) le box est dans la boite.

Since in (b) *le box* could refer to this particular piece in a child's toy model of a farm. Similarly "the box is in the pen" could be either of

(a) la boite est dans le box.

(b) la boite est dans la plume.

where (b) would refer to a box small enough to be concealed in a hollow pen, such a box in fact as could have held the poison which the famous Greek orator Demosthenes was said to have carried around with him in his pen, and by which he took his life on the occasion of his impending arrest by the soldiers of Alexander the Great.

The uniqueness of meaning, then, comes only from the context in which the particular sentence is found, and this being so I see no reason why such sentences should not be capable of being dealt with by a Thesaurus method - or in many cases even by the much simpler method outlined in the book "Mechanical Resolution of Linguistic Problems (1958)" pp.94-95 by Booth, Brandwood and Cleave especially in view of the fact that in scientific texts, the material for which mechanical translation was primarily intended, ambiguity is by no means as common as Bar-Hillel's preoccupation with it would suggest, and can be reduced still further by the use of idioglossaries.

SESSION 2

PAPER 7

AUTOMATIC TRANSLATION IN THE USSR

by

DR. A. P. ERSHOV

AUTOMATIC TRANSLATION IN THE USSR

by

DR. A. P. ERSHOV*

I have been asked to say a few words about the present state of automatic translation in the Soviet Union. I am in rather a difficult position. I must admit that I am only a dilettante in this field, so it is rather difficult for me to talk about it. I can just talk about some of our organizations that are working on automatic translation, about some of the people involved and about some of the main outstanding features that have happened recently in automatic translation in the Soviet Union. I apologise if I say anything that is technically wrong in this field.

The following organisations are actively working on mechanical translation at the moment in the Soviet Union: the Mathematics Institute of the Academy of Sciences of USSR, the Institute of Precision Mechanics and Computation Technology, the Linguistic Institute - these all belong to the Academy of Sciences. There is also work being done in the Laboratory on electro-modelling. Also the Leningrad University has special work being carried out in the Experimental Laboratory on machine translation. In other organisations, also, there are other people who are interested in this particular problem. The Moscow Pedogogic Institute of Foreign Languages is interested in this, and they are doing experiments in this field. A special society for machine translation has been formed: it has regular sessions and publishes proceedings now.

There are two different lines along which automatic translation is being pursued. The Institute of Precision Mechanics was the first in the field, and is following the first line. They have a purely experimental approach to the translation problem, and are more interested in the actual process of translation than in the language itself: so that they are really more interested in getting practical results quickly. In actual fact this particular Institute was the first to show experimental results in translation from English. The Mathematics Institute was also experimenting in the same way. There they were actually doing French

---

* This was an unprepared talk given to the Symposium at the request of many delegates.

translations, finding formal methods of getting French translations into Russian. This work has been done by Dr. Olga Kulagina.

So far as the other institutes that are concerned who are studying this problem of translation, they are following a different approach, which is more theoretical. They are studying the purely formal problems of the language itself, the language structure; then, by doing these formal studies of a language, one can formulate definite rules for translation.

Attention is being paid to the problem of an intermediate language. I am sorry I cannot really say anything more about the detailed side of the problem. I have been more concerned with programming the language translation machine, i.e., the machinery for programming the programmes to put into the machine which automates the construction of the input to the translator.

The Mathematics Institute does work on the large compiler, which automatically constructs the programme for translation. This compiler also has been constructed by Dr. Olga Kulagina. In order to use a compiler like this, you have to get the algorithm of the translation; you have to represent the algorithm as a sequence of specific statements that will characterise the translation. The compiler can make use of 12 to 15 of these different statements. This is already being used for making a programme. Automation of Programming is also done in the Institute of Precise Mechanics. Dr. Razumovskij has done some work in this field.

In Moscow there are quite a lot of seminaries now on machine translation. In Moscow University there is a regular one on mathematical linguistics. Dr. Uspenskij is in charge of the seminary.

The Pedogogic Institute of Foreign Languages also organised a conference which was held in May, 1958 in Moscow. There were about 300 people at this conference and 80 papers were read about machine translation. In the recent international conference of Slavic languages there were several lectures about machine translation. I would like to mention specially a general lecture by Dr. Rozencveig. He made a general survey of the linguistic problems in mechanical translation. The value of this is that it has a very good list of references.

Now there are already very specific articles appearing in the Soviet Press about machine translation. I have two special books I would like to show you about machine translation. The first is - "Problems of Cybernetics" (*ref. 1*). It contains three articles on machine translation. There is an article about a mathematical model of language, another about a translation from Hungarian into Russian, the third about translation from English into Russian. The other book is the second edition of Panov's book on translation (*ref. 2*). There are quite a lot of modifications from the first issue. It is bigger and has been altered to a large extent.

Next year there is going to be a large book by Mrs. Belskaja, which will describe the various work which has been done on automatic translation at

the Institute of Precision Mechanics and Computation Technology. Two next issues of "Problems of Cybernetics" are being prepared for this year. I can say that every copy of this contains three or four articles on linguistic translation.

I will just say a few words about the present tendencies in automatic translation. There had not been anything published in the Soviet Union recently about new experimental translations. That is because they are mainly occupied at the moment with theoretical problems.

REFERENCES

1. Problems of Cybernetics, Part I, edited by A. A. Lyapjnov. *State Publishing House of Physical-Mathematical Literature, Moscow* (1958).
2. PANOV, D. Yu. Automatic Translation. *Academy of Sciences of the USSR, Moscow* (1958).

SESSION 3

SPEECH RECOGNITION

Chairman: PROF. COLIN CHERRY,
Imperial College, London

		PAGE
1	Sensory mechanisms and sensation DR. I. C. WHITFIELD, Saint Elizabeth's Hospital Washington, USA	357
	Discussion on paper 1	369
2	An analogue of the speech recognition process PROF. D. B. FRY and MR. P. DENES, University College, London	375
	Discussion on paper 2	385
3	The perception of speech MR. P. LADEFOGED, Edinburgh University	397
	Discussion on paper 3	411

LEARNING IN MACHINES

Chairman: MR. J. F. COALES,
Cambridge University

4	Two theorems of statistical separability in the Perceptron DR. F. ROSENBLATT, Cornell Aeronautical Lab., USA	419
	Discussion on paper 4	457
5	Learning machines DR. A. M. ANDREW, CME Division, NPL	473
	Discussion on paper 5	507
6	Pandemonium: A paradigm for learning DR. O. G. SELFRIDGE, MIT, USA	511
	Discussion on paper 6	527

SESSION 3

PAPER 1

SENSORY MECHANISMS AND SENSATION

by

DR. I. C. WHITFIELD

BIOGRAPHICAL NOTE

DR. I. C. WHITFIELD graduated from University College London (1937); studied physics, chemistry and physiology. After four years in the research laboratories of Messrs May & Baker he went to the Physiology Department of Aberdeen University and thence to Birmingham where he is Reader in Physiology. He has worked mainly in the field of sensory physiology and especially that of hearing and is the author of numerous articles on this subject as well as in other branches of physiology and physiological instrumentation. He is the author of a book on electronic methods in physiology. Dr. Whitfield is at present with the National Institute of Mental Health in Washington, D.C. where he is in charge of the Section on Behavioural Sciences at the Clinical Neuropharmacology Research Centre.

SENSORY MECHANISMS AND SENSATION

by

DR. I. C. WHITFIELD

SUMMARY

THE very considerable similarity in the transformation effected in the patterns of neural activity at corresponding levels in different sensory pathways is discussed, and the special roles of convergence and inhibition emphasized. These experimentally observed systems would seem to have a smaller informational capacity than is possible on the basis of classical ideas of specific-one-fibre channels and the significance of this for the sensory process is discussed. It is suggested that quantitative studies of sensation must pay as much attention to stored data as to sensory input in assessing channel capacity.

The problem of sensation is one which has caused a great deal of difficulty to the physiologist (and indeed the psychologist). The difficulty of relating the sensory experience to the physical changes in the body, has led to numberless versions of a mind/body duality, and to that rejection of the study of sensation at all which led to Behaviourism. This difficulty has not become any less with the accumulation of data either about the sensations themselves or the physical behaviour of the nervous system.

Primitive ideas of coding regarded the sensorium as examining the outside world by means of a limited number of qualities of sensation. This idea was started experimentally by Bell (1811) and developed by Muller in the 1830's into the famous law of specific nerve energies. The code here was one in which the variables were (a) one of a limited number of "modalities" of sensation, (b) a particular spatial position, and a given fibre was regarded as uniquely activated for a given space/modality. The intensity variable was linked with the degree of activation of the fibre. Although this idea received some crude support from the sensory dissociation found in spinal cord lesions, its application to the special senses failed to stand up to experiment. Perhaps because it was earlier possible to control the stimuli, the subdivision of sight and hearing into a number of sub-modalities took place early. Thus, even before Muller, Young proposed a further subdivision of visual receptors in his theory of colour vision. In spite of the great

increase in, and refinement of, experimental techniques, there has been no clear cut demonstration of three types of retinal receptors corresponding to the requirements of his theory. Helmholtz, who was a great proponent of the doctrine of specific nerve energies applied to the hearing system, got equally into difficulties in trying to allocate one fibre to one discriminable tone, thus requiring some thousands of specific energies. He was early forced to abandon the theory (albeit tacitly) and accept that a given stimulus generally involved a number of receptors which were not uniquely excited by that stimulus. Attempts to assign particular sense modalities in the skin to particular receptor structures has never been conspicuously successful and lately the work of Weddell *(ref. 16)* and his colleagues has demonstrated that more than one of the so-called modalities of skin sensation can be elicited in regions containing only free nerve endings. In any event, the number of types of sensation which can be elicited from the skin, such as itch, tickle, pinch and many more unnamed, far exceeds the number of even proposed skin modalities. It may be argued that these represent different combinations of basic modalities, but this in itself involves the idea of response pattern, rather than unique involvement of specific fibres.

We are left with the conclusion that the doctrine of specific nerve energies is applicable only to the limited statement that e.g. stimulation of visual fibres gives rise to a sensation of light, of auditory fibres to one of hearing and cutaneous fibres to one of skin sensation. But how far is even this valid? The interaction of taste and smell in the production of unitary sensations of "taste" is well recognized. Recent experiments in the field of "sensory deprivation" suggest that the structuring of a sensation habitually referred to one particular modality may normally depend on the interaction of several to achieve normality.

If we are to accept excitation pattern as the key to sensory transmission, as I think we must, the question arises as to the mechanism involved, and the amount of information which it can transmit.

Primitive nervous systems tend to consist of largely unstructured networks of neurones interconnected in a far-reaching way. Thus, Coghill *(ref. 3)* found amblystoma gave generalized responses of the whole body at first and that specialization of action gradually developed. Such an interconnected system seems to be retained in the central nervous system in more highly developed forms. Thus, Lorente de No *(ref. 9)* showed that there was a convergence/divergence organisation of the cells of the cochlear nucleus involving hundreds of adjacent channels. Similar patterns are found in the early cell stations of other modalities. A similar convergent system at the extreme periphery is seen on hair follicles where each follicle is supplied by 2 to 6 fibres and one axon may send branches to 40-120 hairs *(ref. 16)*. Although such patterns of convergence may be random in detail the over-all spatial organization seems to be preserved from cell station to cell station. Thus, although Hilali and Whitfield *(ref. 7)* could find no evidence

of relationship between fibre position and anatomical location in the efferent pathway from the cochlear nucleus, such a relationship is found both in the cochlear nucleus itself and at the cortex, suggesting that the fibres are reordered in some way at each cell station. Similar observations have been made on the optic nerve, and Sperry *(ref. 11)* has described the reorientation of the deliberately surgically-disorganized optic nerve in the newt during regeneration. The mechanism of this maintenance of pattern is quite obscure.

Although the orderly convergence/divergence organisation appears to be the basis of sensory patterning, a purely excitatory mechanism based on the known properties of synaptic activity is inadequate to account for the observed transformation carried out by a sensory cell station as Allanson and Whitfield *(ref. 1)* showed for the cochlear nucleus. In order to account for the transformation, it is necessary to postulate an additional inhibitory process, and it seems that it is, in fact, this aspect of the neural activity which is the key to the transformations effected. The nature of sensory receptors seems to be such as to produce a graded response to stimulation, such that the rate of firing increases with the strength of the stimulus. Adjacent receptors respond with slightly different discharge behaviour to a given stimulus, though a stimulus (a) can probably be found which will cause receptor A to respond in the fashion that receptor B responds to stimulus (b). In the ear this receptivity gradient appears to be frequency. There may be a similar continuous graduation of colour sensitivity among retinal receptors *(ref. 5)*. Provided such a differential exists it is at least possible for the neural network to elaborate a pattern corresponding to a given situation. Such a pattern may or may not exhibit a one-to-one correspondence between stimulus and sensation.

The pattern of graded responses which the receptors produce does not appear to be suitable for utilization by the more elaborate processing networks of the higher centres. In all systems we find a tendency for the responses at higher levels of the system to be more nearly on/off and for the distinction between active and inactive regions to be more clearly defined by the inhibitory process operating in the boundary regions. Such effects are seen clearly in the auditory system and in the somatic pathways. In both these cases stimulation of an adjacent receptor may result in suppression of the activity of units near the periphery of the stimulus region. Whitfield, *(ref. 17)* and Allanson and Whitfield *(ref. 1)* have described this mechanism in the auditory pathway *(fig. 1)*, while Powell and Mountcastle *(ref. 11)* have shown an analogous effect at the somatic sensory cortex *(fig. 2)*. The response of such a cortical unit to peripheral stimulation can be inhibited by stimulation of an adjacent skin area and gradations of inhibiting activity exist between the centre and edge of the field; pairs of closely adjacent neurones may be reciprocally related with respect to excitatory and inhibitory stimuli.

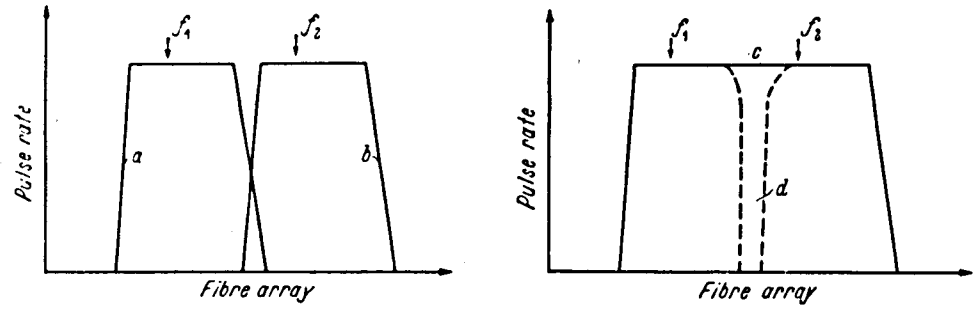
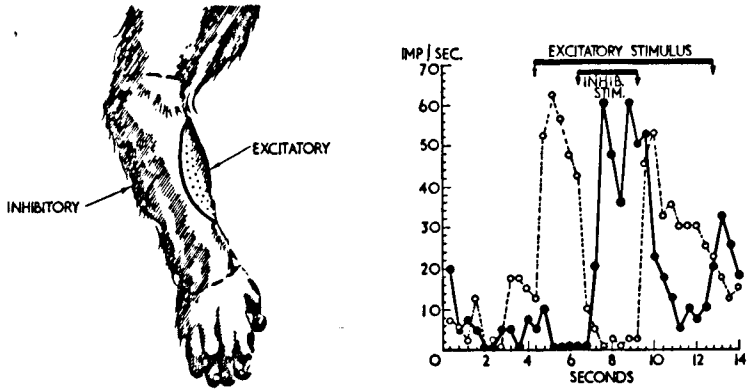


Fig. 1. Curves a and b are the responses of units at higher levels in the auditory pathway to frequencies f1 and f2 presented singly. When the stimuli are presented together the result is not curve c, but rather curve d, owing to the inhibitory interaction between the two responses - (Allanson & Whitfield, *ref. 1*)



RECIPROCALLY RELATED CORTICAL NEURONS 2·5 AND 2·5a

EXCITATION — INHIBITION

Fig. 2. The reciprocal relation between two cortical neurones activated by stimulation of adjacent skin areas. (Powell & Mountcastle, *ref. 11*)

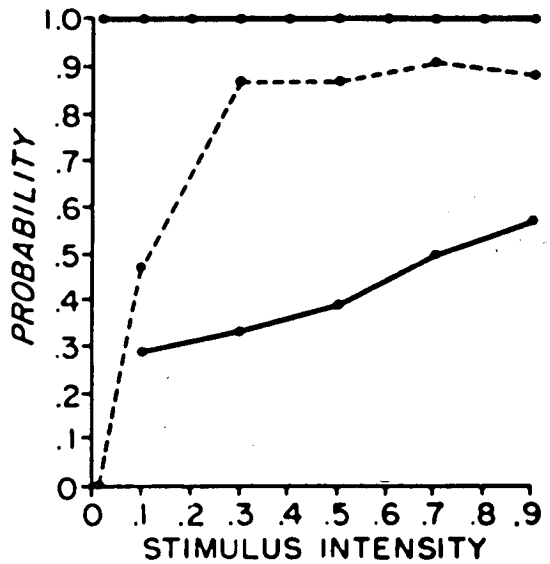


Fig.3(a). Relation of response probability to stimulus intensity in somatic sensory cortex. Solid lines: change of P in one unit fired by adjacent digits, upper from centre of receptive field and lower for near edge of field. Dashed line: most common form of relation found, *viz.*, rapid initial rise of P followed by plateau at some high value of P. (Towe & Amassian, *ref. 15*)

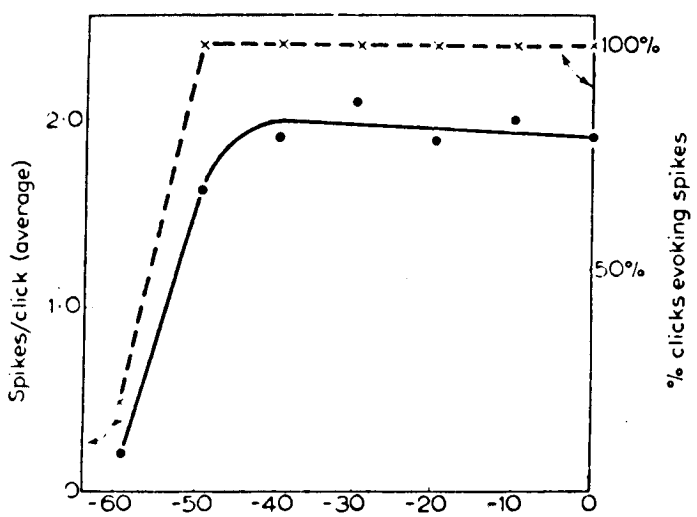


Fig.3(b). Full line: Average numbers of spike responses evoked from an auditory neurone at the thalamic level, by clicks of different intensity. Dashed line: percentage of clicks which evoke any spike, as a function of intensity. From data of Galambos, Rose, Bromely and Hughes (Whitfield, *ref. 17*)

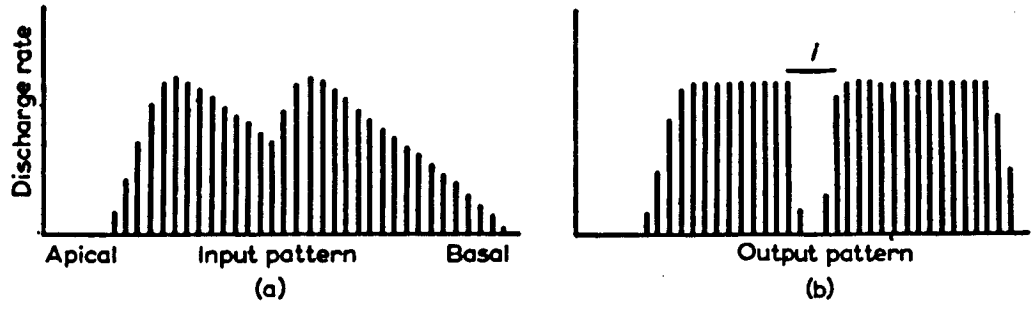


Fig. 4a. Each vertical line represents 10-20 adjacent channels and its height the mean discharge rate in those channels. The afferent array is stimulated by two tones whose frequencies correspond to the peaks of input pattern (a). Inhibition in the region of 'overlap' of the responses results in a gap in the output pattern (b) so that the active regions remain distinct. (Whitfield, *ref. 17*)

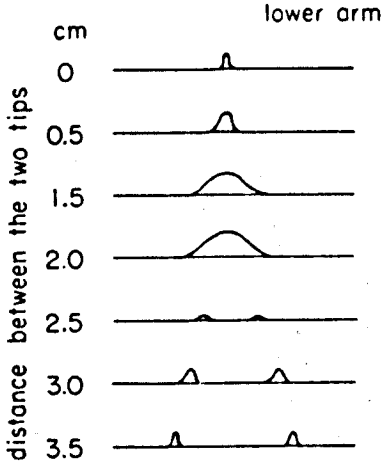


Fig. 4b. Summation and inhibition on the skin surface. Sensation produced by two compass tips pressed against the surface of the skin as the distance between the tips is increased. When the distance is 2.5 cm the magnitude of the sensations is smaller than for one point along. (Bekesy, *ref. 2*).

At thalamic and cortical levels only one or two spikes may be discharged by a stimulus and the latency of this discharge tends to decrease rapidly near threshold and then more slowly at higher intensities. The probability of discharge usually, but not invariably, shows a similar relationship also to stimulus intensity, the curve rising more rapidly near threshold, flattening out, and then even occasionally tending to fall again at very high intensities, as has been found to happen sometimes at lower neural levels *(refs. 7,13)*. Since the events of a cortical unit response are not constant for repetitions of the same stimulus, unit activity at this level is most usefully treated in terms of probability of discharge. Such probability is a function of both the stimulus position and intensity, the receptive field being larger at greater intensity *(ref. 15)*. The behaviour is very similar in both somatic and auditory systems *(fig. 3)*.

Similar reciprocal inhibitory effects are found by Kuffler *(ref. 8)* in the retina, where they are presumably likewise concerned with the isolation of individual stimulus areas in the formation of discrete patterns. An example of a possible subjective result of such mechanisms is seen in the Mach effect in which a gradual intensity change gives rise to a sensation of bright and dark lines near the limits of change.

It is of interest to note that whereas there is good evidence for spatial localization of activity in relation to skin and visual senses and some for frequency in the auditory system, there is no corresponding localization of qualities, such as colour, tone quality, or e.g. tickle. Although many attempts have been made to isolate units responding only to say red or blue at the thalamic or cortical levels, no clear demonstration of such units has been made. There would seem strong circumstantial evidence that these attributes depend on particular relations of pattern, and not on the activities of particular neurones. Colour contrasts which may depend in a similar manner to brightness contrasts on inhibitory processes have been shown to be most readily demonstrated when brightness contrasts, and especially contours, are eliminated *(ref. 14)*. It seems possible therefore that one inhibitory mechanism may disrupt the activity patterns involving another.

As pointed out by Whitfield, *(ref. 17)* inhibitory processes in the sensory pathway tend to prevent the coalescence of the response areas to adjacent stimuli in the auditory pathway *(fig. 4a)*. A similar situation has been demonstrated by Bekesy *(ref. 2)* on the skin *(fig. 4b)*. Two-point stimuli applied to the skin of the thumb summate when the points are closer than 2 cm., and the apparent magnitude of the sensation increases as this separation is approached. At 2.5 cm. the stimulus appears double and the sensation magnitude drops abruptly as the intermediate region response is presumably inhibited. The corresponding animal experiment with cortical recording has been described by Mountcastle, *(ref. 10)* who found a

sharpening of discharge zones and inhibition in the intermediate zone as two stimuli were approximated.

To summarize then, we appear to have a network in each sensory system whose overall effect is to convert a whole array of graded responses to a situation into a pattern consisting of a number of on/off channels, considerable numbers of such channels being activated by any given stimulus and no channel in general being uniquely activated by any given situation. In the case of the ear, for example, a comparatively few bands of activity, each involving some hundreds of channels, appear to convey all the data concerning intensity and pitch available in the pathway at any instant. If such a scheme is a valid interpretation of the observed data, there would appear to be an enormous discrepancy between the actual informational capacity of the system and that implicit in the 'monitoring sensorium' hypothesis. Instead of, for example in the auditory system, some 30,000 independent channels, each capable of operating in many states, we have large blocks of fibres in an approximately two-state system. There is at present no means of assessing what percentage change in the number of fibres involved in a block is discriminable, but it would seem that there must inevitably be a smaller number of functional channels than in the ideal system called for by the specific nerve energy hypothesis. The question therefore arises how far such a system would be capable of accounting for the known capabilities of sensory systems in handling the operations which they have to carry out.

Considering particularly again the auditory system, it is possible by determining the successive difference limens for frequency and intensity to build up an area of unit cells, each constituting a unique state. Such an area contains some 350,000 cells. However, such a measure is rather meaningless, since no subject can in fact identify any one of these cells in isolation. Experiment shows that the approximate specification of a comparatively few parameters is sufficient to determine a recognizable sound quality. Thus, two or three formant frequencies will provide the appropriate sensation for the recognition of a vowel. The recognition of an oboe tone likewise depends on the emphasis of a few formant regions.

The interaction of 'modalities' may restructure sensation for similar situations. Thus, in the example quoted by Russell Brain, with the situation "hand in cold air" the usual statement about the percept is "my hand is cold". Given the same thermal situation by 'hand grasping stone' the percept may be rather stated as "the stone is cold".

In the same way that sensation differs according to the interaction of modalities, the stored data of the system can likewise change the sensation.

Thus, in the tachistoscopic exposure of drawings containing an incongruity, repeated exposure results in abrupt changes in the percept as repeated attempts are made to match the incongruity against previously experienced possibilities. A two-state example of this is found in the numerous reversible-perspective figures, such as Necker's cube and Schroder's

stairs, which do not correspond exactly to the experienced form of either version, but can be 'seen' as either at different times.

The distinction between the sensation of a violin note or a trumpet note depends very largely on the transients of the notes' initiation, which 'structures' the sensation of the subsequent sound. In the absence of these transients the sounds are practically indistinguishable, yet after the transients have provided a clear orientation, the sensations subsequently experienced are quite different.

Thus, it seems that sensory input organizes the internal system to produce a probable inference about the environment, which is the sensation. This is much less easily structured if only one modality is functional. That the internal system is continually attempting to structure any input is suggested by the observation that in "sensory isolation" experiments hallucinations are more readily produced with a zero information visual stimulus than in absence of any stimulus *(ref. 4)*.

A physical stimulus below threshold can give rise to a more than chance proportion of correct answers if a large number of guesses by the subject is recorded, thereby providing an artificial memory store. Bartlett has stated 'even the most elementary perceptions have the character of inferential constructions'. It seems that this is as true of the sensation itself as of the more elaborate percept, and suggests that the distinction is one of degree rather than kind. Like many other aspects of neural function conscious sensation appears to have a probability basis.

It is suggested therefore that the apparently low informational capacity of the highly interconnected interdependent arrays of channels of which sensory pathways appear to consist, as compared with the large capacity of a system where each channel is entirely independent of any others as postulated in the classical Mullerian system, is not incompatible with subjective observation, and that in attempting to measure the behaviour of such pathways account must be taken of the total system. This will involve not only the 'modality' being studied but quite remote 'clues' provided by other systems.

In all this the crying need is for more physiological data. Collection of such data has brought to light an unsuspected homogeneity of function in sensory systems and at the same time radically reoriented the psychosensory problems. Further progress is hampered by our lack of knowledge of the fundamental principles on which the nervous system operates. Just as the pulse-code theory of fibre transmission, which appeared an attractive possibility, has been found to be invalid, so it is by no means established that the central nervous system in general works on a digital basis. There is a considerable body of evidence that some cells can show a graded response to stimulation *(ref. 6)* and this would accord with the non-importance of particular pulses, and the necessity for statistical treatment at the higher levels. We know too next to nothing about the storage

of information in a physiological system - whether such storage is an activity pattern, a property of interneuronal junctions or some microstructural property of the internal or external environment of the neurones; in fact, we do not even know enough to say if neurones are the only structures we need consider. Yet it seems evident that such knowledge is as essential to an understanding of sensation, as is knowledge of the sensory pathways themselves.

REFERENCES

1. ALLANSON, J. T. and WHITFIELD, I. C., 3rd London Symposium on Information Theory. *Butterworth* (1955) p. 269.
2. BEKESY, G. von, *J. Acoust. Soc. Amer.* (1957) **29**, 489.
3. COGHILL, G. E., *J. Comp. Neurol.* (1930) **51**, 311.
4. GOLDBERGER, L. and HOLT, R. R., *J. Nerv. Ment. Dis.* (1958) **27**, 99.
5. GRANIT, R., Sensory Mechanisms of the Retina. *Oxford University Press* (1947).
6. GRUNDFEST, H., *Physiol. Rev.* (1957) **37**, 337.
7. HILALI, S. and WHITFIELD, I. C., *J. Physiol.* (1953) **122**, 158.
8. KUFFLER, S. W., Cold Spring Harb. Symp. (1952) **17**, 281.
9. LORENTE DE No, R., *Laryngoscope* (1933) **43**, 1.
10. MOUNTCASTLE, V. B., *J. Neurophysiol.* (1957) **20**, 408.
11. POWELL, T. P. and MOUNTCASTLE, V. B., *Fed. Proc.* (1958) **17**, 126.
12. SPERRY, R. W., *J. Comp. Neurol.* (1943) **79**, 33.
13. THURLOW, W. R., GROSS, N. B., KEMP, E. H. and LOWY, K., *J. Neuro-physiol.* (1957) **14**, 289.
14. TICHENER, E. B., *Experimental Psychology* (1907).
15. TOWE, A. L. and AMASSIAN, V. E., *J. Neurophysiol* (1958) **21**, 292.
16. WEDDELL, G., TAYLOR, D. A. and WILLIAMS, C. M., *J. Anat.* (1955) **89**, 317.
17. WHITFIELD, I. C., *J. Physiol.* (1955) **128**, 15P; Progress in Biophysics and Biophysical Chemistry. *Pergamon Press* (1957) **8**, 1.

DISCUSSION ON THE PAPER BY DR. I. C. WHITFIELD

MR. J. T. ALLANSON: How far is the similarity between different sensory modalities one which is partly due to the comparatively simple ways in which experimental work is done on each particular sensory modality? When doing an experiment, one picks out means of testing which are generally simple in themselves; for instance, in the auditory system the great bulk of the work has been done either with pure tones, or with clicks, because it would be very difficult to try to analyse results obtained from such a signal as is contained in a speech wave form. With skin, again (Bekesy's work), essentially very simple pressure sensations were used, concentrated on one or two points. It seems to me that we may be in danger of drawing conclusions from these simple situations which, in fact, are only valid for stimuli of a comparatively simple kind and would not be valid for much more complex stimuli varying rapidly with time. I would like to hear Dr. Whitfield's opinions on this.

The second point is connected with the first. When we examine the results obtained by testing in simple situations and then examine the histology, it is possible to make predictions such as would enable us to carry out further experiments. Almost invariably one then finds that the actual biological material is far more complicated than was assumed. For instance, it is possible to predict the incidence of two-tone inhibition in the auditory nervous system from an examination of such results as shown by Dr. Whitfield. It is possible to develop a general theory of the system, but then one finds that the number of times that the inhibition can be found is a comparatively small proportion of the total number of times units may be found to show a simple response to a sinusoidal signal. In the majority of these cases, it is not possible to find any second tone to inhibit activity at all. Or perhaps there will be a unit which can be inhibited, but it is not possible to find any signal to which it gives a positive response. I wonder how far we are being concerned about phenomena which have been noticed largely because of their peculiar character (because they are responses to stimuli which are in reality rather exceptional) and are not, in fact, typical of the great bulk of the responding units.

MR. LAWRENCE: Taking up the point made by the last speaker, I would say that, as a worker in speech, I feel very strongly that the sine wave is a bad stimulus to use. It is one that does not occur in nature. To make a sine wave one has to do very unnatural things with electronics. My own feeling is that speech wave forms provide a very wide range of stimulus and

a very important one, because we can get introspective information about the perceptions - things that are perceived when speech waves are used as the stimulus. Synthetic speech waves are perhaps the ideal stimulus to use, because we can vary them in very special ways. We know just what we are doing, and we can ask the subject questions of a very sophisticated nature about what he hears.

I would like also to take up with Dr. Whitfield the question of whether he has any physical models that will represent these phenomena of inhibition that he has shown us, which interest me very much. Can he, for instance, show us how it would be possible to set up a model of the basilar membrane, which would exhibit just these phenomena of inhibition, showing the cross-connections between two adjacent spectral maxima?

DR. H. B. BARLOW: There are three points I would like to raise about Dr. Whitfield's paper. The first is to question whether one must abandon, as completely as Dr. Whitfield suggests, the idea that sensory nerve fibres belong to a limited number of different types, each type having a rather specific range of sensitivity. I agree that the ranges of sensitivity of each type do not seem to correspond to the named modalities or qualities of sensation in such a simple manner as nineteenth century physiologists may have thought; this, however, need not imply that the sensory fibres cannot be classified into distinct types. Certainly in the case of colour vision the evidence still seems consistent with the original trichromatic theory *(ref. 1)*.

With regard to the mutual inhibition between two different stimuli, this is, as he points out, rather similar to lateral inhibition as found in the retina. In the eye an additional fact is known, namely that the degree of mutual inhibition can vary with the past history of the retina *(ref. 2)*. I would like to ask if the pattern of mutual inhibition varies in the auditory pathway according to the past history of stimulation.

Finally, if I understand Dr. Whitfield, he interprets lateral inhibition as an example of converting an analogue quantity into a digital quantity; mutual inhibition converts the trains of pulses into a standard form, or inhibits them altogether. I would like to point out the alternative explanation (see Session 4 Paper 1) as an example of "economy of impulses".

MR. R. L. GREGORY: I am afraid that, like Dr. Barlow, I also missed the talk, for which I apologise; but I would like, if I may, to have a little more elaboration of paragraph 4; page 367, where the problem of sub-threshold

REFERENCES

1. BRINDLEY, G. S. Human Colour Vision. *Progress in Biophysics*, 1957, 8, 49.
2. BARLOW, H. B., FITZHUGH, R. & KUFFLER, S. W. Change of organisation in the receptive fields of the cat's retina during dark adaptation. *Jour. Physiol.*, 1957, **137**: 338.

or subliminal stimulation is considered, and the idea of a sort of memory store is brought in. It does not seem necessary to postulate this if we regard sensory discrimination as a signal to noise situation. The question is what we mean by a "threshold". For the subject to be able to report that he has seen a flash of light is one criterion of threshold; to response with a key, or perhaps to get a psycho-galvanic reflex or some other response, may be others. There are, surely, a number of criteria which could be adopted, and what we mean by a "threshold" depends upon which criterion is adopted, and thus what appears as "subliminal".

MR. E. A. NEWMAN: It seems to me that study of the nervous system is greatly complicated by the difficulty of sorting out what processing occurs at the periphery and what is dealt with in the central system.

For example we can hear a note of 1000 cycles, but this does not mean that the central system can handle a frequency of 1000 cycles. It is possible that the peripheral system converts the 1000 cycles into a spatio-temperal coded analogue containing no high frequencies, and the central system handles this. Information-wise it would be adequate to pass on a spatial analogue of the frequency, this analogue changing no faster than the envelope of the 1000 cycle note.

The work Dr. Whitfield does sheds a lot of light on the way in which processing is divided. Could he say anything about what certainty exists in the knowledge of division of processing between various parts of the peripheral system, and the central system.

PROF. CHERRY (Chairman): Perhaps I could add a remark, Dr. Whitfield. In our own Department we have had work going on for some time on binaural fusion; consideration has been limited entirely to the behavioural level, not enquiring into physiology, but regarding the brain as a black box with mechanism inside, adaptable and adjustable.

It appears to us that the behavioural evidence is quite overwhelming, that the process involves a certain kind of short-term correlation analysis. More recently we have been considering the possibility that the basilar membrane may act as a differential delay line, spreading out the information along its length as a preparation for carrying out such correlation processes. We are now considering experiments by which we can test such an hypothesis. I wonder whether it seems reasonable to a physiologist that this could be a primary cochlea function.

DR. I. C. WHITFIELD (in reply): In reply to Mr. Lawrence's point, I agree that obviously one needs to be able to deal with complex stimuli because this is what the body is usually concerned with. On the other hand they are rather intractable from the analytic point of view, and we have tried to start off by using simpler, if perhaps physiologically unusual, stimuli.

I think Mr. Lawrence himself would agree that in fact the things which you have to specify in a speech wave in order to get at least a recognisable perception, are comparatively few compared with the complexity of the sound wave, and we have tried to carry this down to the sensation level. We have in addition been trying the use of frequency-modulated stimuli as an alternative to speech waves.

There is no suggestion that sensory receptors do not belong to different types; it is obvious that some respond to light, some to mechanical vibration, etc. Further, within these categories there must be variations in the degree of responsiveness to different stimuli, otherwise no sort of analysis could take place at all; the point is precisely that these differences do not correspond one-to-one to the sensation experienced.

The number of different regions of responsiveness proposed for colour receptors varies from time to time, but I would point out that the trichromatic theory is in fact a pattern hypothesis, since it is postulated that the sensations produced by stimulation of the receptors in combination are quite distinct from those produced by individual stimulation. The generality of the pattern hypothesis is essentially unaffected by whether such receptors are divided into three, five or n categories with respect to the stimulus parameter (here frequency) provided there is some graduation in the response among the members of the receptor set.

Mr. Allanson has pointed out that simple stimuli may lead us to conclusions which are valid only for those stimuli, and this is of course a danger in all extrapolation. However the pattern revealed by such stimuli appears to be common to several modalities and may perhaps provide ground on which to build a study of more complex situations as techniques improve. Circumstantial evidence, as I have tried to indicate, suggests that the extrapolation may be valid.

I would not regard lateral inhibition as participating in the 'squaring' process. The lateral inhibition appears rather to assist in maintaining the identity of stimuli and is a necessary adjunct to the former process. Inhibition is apparently also involved in the transformation of the pattern due to a single stimulus, but this is inhibition within the single array. I doubt whether the type of transformation we have been considering could fairly be called an analogue/digital conversion (although of course the borderline between these is anyway indistinct). It is more usefully thought of as a transformation from one spatio-temporal analogue to another. This may certainly involve economy of impulses as a by-product.

I agree that there are a number of possible levels of threshold, from that of the receptor onward, and at least as many criteria for assessing these; the one I have been concerned with here is of course that of conscious sensation.

The question whether cells behave differently at different times is a most important one and may tie up with what Allanson was saying concerning

the fact that one finds units which behave in ways which are very varied - units which will not inhibit, units which will respond only to stimuli which are changing and not to those which are constant in time, and so on. This variability is found at every site in the nervous system and we still have little data on how far the variations are meaningful and how far they are statistical "accidents". Such variation does not invalidate the argument, however, since it is by no means essential for it that *every* unit should behave uniformly; the hypothesis does in fact remove the necessity for any great precision of behaviour.

This problem may be tied up with the fallacy of multiplying difference limens i.e. that one can make a very accurate judgement of the difference between two stimuli, but a much less accurate judgement of an absolute value. I think we have to consider the probability that, at quite a low anatomical level, what any particular cell is doing may change from time to time according to the discrimination being made. It would certainly appear that there are descending pathways which can affect the secondary sensory neurone and possibly in some cases even the primary neurone and which can presumably change their properties. It may be that the whole range of these simple patterns carried by a particular neuronal configuration can be changed from time to time according to the discrimination being made. So I would agree that one does find these changes in properties, but at present we have simply no data on their significance, and I believe we are waiting largely on more information about the centrifugal pathways.

Now to the question of the relative roles of basilar-membrane and neural pathway: The broad facts of this 'division of labour' are now fairly well established and are treated for example in reference 17. As far as we can tell, the inhibition phenomena of which I have spoken are the result of neural interaction and are not a mechanical phenomenon situated at the basilar-membrane level.

With regard to whether the basilar-membrane could behave as a delay-line, it certainly exhibits travelling-wave properties with phase differences along its length *(ref. 1)*. Many theories have been proposed involving these time differences and their derivatives, but no one has come up with any solid physiological evidence which would support a detailed hypothesis. I am therefore unable to give an answer as to whether the membrane could usefully function in this way, but it seems improbable. Owing to the uncertainty of firing of neurones it would appear that at fairly high frequencies, after you have gone through two or three synapses the timing in any particular pathway is not very good. On the other hand if you take a statistical picture of the activity over a number of channels, more especially with transient stimuli, then quite reliable time estimates emerge.

We still have no idea, also, of the relation between the low frequency stimulus situation, where there is a clear time relation between the

response and the stimulus waveform, and the high frequency situation where there is not. There appear to be two quite different mechanisms at work which are combined neurally in some way. However I have not wished to consider here such phenomena as these, which appear peculiar to the hearing process, so much as to make comparisons between the patterns in different sensory pathways. It may be that these vary in detail between different modalities - it is clear that they must do so - and I think it is probably quite true that even at a low anatomical level the transforming functions can be modified by the higher levels, since pathways descending straight from the cortex to the sensory nuclei of the spinal cord have been clearly demonstrated both anatomically and physiologically. Nevertheless there do appear to be, aside from such variations, general basic similarities in the behaviour of the various sensory systems and in the sorts of patterns to be found in their pathways.

REFERENCE

1. TASAKI, I., DAVIS, H., and LEGOUIX, J. P., *J. acoust. Soc. Amer.*, 1952, **24**, 502-519.

SESSION 3

PAPER 2

AN ANALOGUE OF THE SPEECH RECOGNITION PROCESS

by

PROF. D. B. FRY AND MR. P. DENES

BIOGRAPHICAL NOTES

Prof. D. B. Fry obtained his Ph.D. (London) in 1934 as a student of Prof. Daniel Jones. He has been a member of the staff of the Department of Phonetics, University College, since 1934, taking charge of the Phonetics Laboratory from 1937 to 1949 and being head of the Department since 1949. He did war service with the R.A.F. from 1939 to 1945, being occupied with experimental work on hearing tests and speech transmission. His research interests are: speech and hearing (including deafness and the education of the deaf), and particularly the perception and recognition of speech, speech synthesis and mechanical speech recognition.

He is now Professor of Experimental Phonetics in the University of London and head of the Department of Phonetics, University College, London. He is also Consultant in Phonetics to the Institute of Otology and the Royal National Throat, Nose and Ear Hospital, London.

Mr. P. Denes, M.Sc., A.M.I.E.E., obtained his B.Sc. and M.Sc. degrees at Manchester University in 1941 and 1943 respectively. From 1941 till 1944 he worked as demonstrator at the Electrotechnics Department of Manchester University and from 1944 until 1946 as research engineer in an industrial organisation. In 1946 he became lecturer in experimental phonetics at University College London. His work includes research on speech communication, speech recognition and synthesis, deafness, hearing tests and hearing aids.

He was visiting fellow of Columbia University New York in 1943. He is a member of the Medical and Scientific Committee of the National Institute for the Deaf, member of the technical committee ACM/4 of the British Standards Institution and Physicist at the Audiology Unit, Royal National Throat, Nose and Ear Hospital, Grays Inn Road, London.

AN ANALOGUE OF THE SPEECH RECOGNITION PROCESS

by

PROF. D. B. FRY AND MR. P. DENES

SUMMARY

THE human speech recognition process both at the acoustic and the linguistic level is discussed. An analogue of some parts of this process is described. Plans for using computers to obtain a more comprehensive analogue of human speech recognition are outlined.

EVERY message that passes between speaker and listener is first formulated and finally decoded as a sequence of linguistic units. In the process of transmission the message takes on a variety of forms. Of these, the acoustic form has been investigated most thoroughly and it is clear that no one acoustic feature nor indeed one set of features will uniquely identify a linguistic unit. The criteria by which the incoming sound waves are classified into linguistic units are not constant and are determined by the conditions of communication, such as the speech habits of the speaker, the conditions of transmission and reception, etc. If we attempt to identify linguistic units, such as phonemes, we find that in any physical dimension a range of values rather than a single value has to be assigned to a phoneme. The ranges for different phonemes overlap and this gives rise to ambiguities which cannot be resolved on the basis of a constant criterion. For example in the case of vowel formants, the ranges of formant frequencies overlap. The data given by Petersen and Barney *(ref. 5)* show that for the vowels /I/ and /ɛ/ the formant ranges are:

	/I/	/ɛ/
F_1	300 c/s - 570 c/s	390 c/s - 800 c/s
F_2	1800 c/s - 2900 c/s	1800 c/s - 2800 c/s

Such a process of acoustic recognition gives rise to a proportion of errors. It is worth noting that such errors occur also in the case of the human listener if he is forced to depend on acoustic recognition alone. For example, when he is asked to distinguish between /m/ and /n/ in nonsense syllables he will make a certain number of errors. When he listens to meaningful speech the errors and ambiguities are cleared up by the use of linguistic information. Let us now consider how this factor operates in the listener and then how it may be used in a mechanical analogue. When the listener receives a message he orders the information in accordance with a system of linguistic units already familiar to him from past experience. The smallest of these units is the phoneme and there are about 40 of them in English. They are roughly equivalent to single letters in a printed text and they are the smallest linguistic units that distinguish meaning. A sequence of phonemes makes a larger linguistic unit, a morpheme. Morphemes are the units that have grammatical function. The next order of linguistic unit is the word; many words consist of a single morpheme but a word may also be a sequence of morphemes. The number of words - and hence of morphemes - in English is tens of thousands. The largest unit that has a function in spoken language is the sentence. The number of possible sentences is very large even though not all combinations of words make sentences. The linguistic units then are the phoneme, the morpheme, the word and the sentence.

It is well known that speech is a stochastic process; the units that have already occurred in the sequence determine the sequential probabilities of succeeding units. This is true for all levels of linguistic units. The sequential probabilities at any one linguistic level are determined not only by the preceding sequence at that level but also by the preceding sequences of larger units. An example of sequential probabilities at the phonemic level is the fact that /h/ does not occur finally in English, or again that /ŋ/ is very unlikely at the beginning of a sequence. As a result, whenever a nasal consonant is heard initially, the listener automatically rules out the possibility of /ŋ/ and considers /m/ or /n/ only, even though all three consonants are acoustically similar. At any point in the phonemic sequence the probabilities governing the occurrence of the next phoneme are determined not only by the preceding phonemes but also by the words and sentences formed by these phonemes. In the sequence *The cat sat on the....* followed by a nasal consonant, for example, /m/ has a much higher probability than /n/ or /ŋ/ because the probability that the next word will be *mat* is high. Thus sequential probabilities at any one level are affected by sequences at all other levels. This reinforces the effectiveness of the system of sequential probabilities in resolving the ambiguities and correcting the errors of purely acoustic recognition. Sequential probabilities are effective in correcting errors even retrospectively. For instance in the sequence *it's meat but not...* there is

nothing as yet in the phonemic, morphemic or word sequence to indicate an error, but if the sequence is continued with the word *gaudy* it is obvious that the second word should be *neat* and not *meat*. There are some occasions when the human listener is aware of this kind of process but it probably happens quite frequently without our being aware of it.

Speech recognition then depends very largely on linguistic, that is, statistical information and this is despite the fact that the listener has a much better instrument for acoustic discrimination than is practicable in an analogue. It is unlikely then that a mechanical speech recognizer will be successful without the use of some form of linguistic information. Our experiments with analogues of the speech recognition process are therefore concerned more with exploring the practicability and effectiveness of combining acoustic and linguistic information in the analogue than with perfecting methods of acoustic recognition.

In considering the application of linguistic information in the analogue the first question to be answered concerns the size of linguistic unit which is to form the basis of the statistics. In the existing analogue the phoneme has been chosen for this purpose. This choice was dictated mainly by practical requirements. Theoretical considerations concerning the length of unit and the possibility of compromise between these considerations and what is practicable in the machine, for example because of limitations of storage, will be discussed later. We shall now go on to describe the use of phoneme statistics in the present analogue.

This machine deals with 13 English phonemes, four vowels /i, a, u, ɜ/ and nine consonants /t, k, f, s, z, ʃ, m, n, l/ and space. The language of the recognizer consists of all English words made up of these sounds which do not contain a sequence of two consonants or two vowels. There are about 200 of these and they were recorded withaa space between them for processing with the recognizer. In the first stage of the experiments the linguistic information was restricted to digram frequencies and these were counted for the "language" just described. This information is stored in the machine in the form of 14 sets of 14 voltages that are proportional to the digram frequencies. At any point in the phoneme sequence the appropriate set of voltages is combined with the output of the acoustic recognizer by a further process of multiplication and maximum detection. The acoustic recognizer provides an estimate of acoustic probabilities, the linguistic section an estimate of linguistic probabilities and the machine makes its final decision by selecting the phoneme for which the combined probabilities are greatest. The proportion of sounds recognized correctly by the analogue is about 70% and of words 45%. The present analogue uses only linguistic information of the most rudimentary kind and the next problem is to find the most profitable way of extending it. As we have already said, this involves theoretical and practical considerations, since the linguistic information can be extended either by extending the

span of phoneme probabilities to three or more places or by increasing the size of the linguistic units that form the basis of recognition. In theory any of the four linguistic units, phoneme, morpheme, word or sentence could be used. The longer the unit, the stronger the constraints because sequential probabilities are always operating, and the stronger the constraints the better the chances of eliminating errors. On the other hand, the larger the unit chosen the greater the storage capacity required both because each item is larger and because in language the larger the unit the greater the number of items. There are only 40 phonemes in English, but tens of thousands of words and an even greater number of sentences. The use of the sentence as the recognition unit would demand such a large storage capacity that it can be regarded, at least at present, as impracticable. It seems worthwhile however to consider the relative merits of the word and the phoneme as units of recognition. If we are to consider words as the basic units we have to be able to segment the input into words. Our previous experiments have shown that we can segment the input into phoneme sequences with some success and to achieve this segmentation with word units would probably require the words to be pronounced with a space between them. If segmentation in this way is practicable, the machine repertory could be stored in the form of word patterns. This storage could be either in the form of complete acoustic patterns each of which corresponds to a word, or in the form of phoneme sequences each of which makes a word. In the latter case the analogue deals with two linguistic units of different size and builds the larger units from the smaller ones. Our experience up to the present has been with phoneme recognition in the analogue and we shall therefore discuss first the merits of this method.

Let us compare two analogues, one in which the phoneme is the only recognition unit and in which a knowledge of phoneme sequential probabilities up to n places is available; the other in which words are stored as phoneme sequences of any number up to n. On a purely statistical basis the first of these will put out some sequences which do not form words but this type of error is not possible in the second machine which will therefore be in this respect superior. This can be expressed in another way by saying that in the first system a final decision is taken at each phoneme position. This decision is made in the light of information about preceding phonemes but once made cannot be corrected. In the second system the final decision is not made until a sequence of phonemes constituting a word is complete. Up to this point, phoneme decisions are provisional and when the sequence is finally compared with the word store, corrections can be made at any phoneme position in the light of the phonemes that follow as well as those that precede this position in the word. It should be mentioned that in this system the provisional phoneme selection can be made with or without the use of phoneme sequential probabilities. Only experiment rather than theorising can decide whether the use of sequential probabilities in the provisional phoneme selection would be an advantage.

Another possibility that we have mentioned is that of storing words as complete acoustic patterns and not as phoneme sequences. In this case the incoming acoustic pattern is compared with a store of patterns corresponding to the words of the repertory. Can we expect an advantage from such a system compared with those already discussed? We have already seen that, on the level of linguistic constraints, lengthening of the unit should give an advantage and reduce errors. Speech however contains articulatory, that is, acoustic constraints as well as linguistic ones, and by dealing with acoustic patterns for words we should gain some benefit from the operation of these acoustic constraints. In order to compare such a system with those using linguistic units we should have to know the relative strengths of the acoustic and linguistic constraints in speech. As this information is not at present available the only way to decide on this point is by experiment.

Once a system using the word as a unit of recognition is available the success of recognition could be improved still further by the use of word transition probabilities.

A number of the methods suggested in this paper raise the central problem of storage of and access to information. Even to extend the span of the phoneme statistics from digrim to trigram frequencies requires such an increase of storage capacity that it seems worthwhile to consider whether the experiments could not be carried out more economically by using a computer rather than a specially constructed analogue. The problem of using computers for this purpose is at present being studied. It is already clear that even the storage capacity required for dealing with trigram frequencies might make demands that some of the available computers cannot meet. On the other hand the flexibility offered by computer techniques is so great that many of the problems discussed in this paper can be explored in principle without necessarily dealing with the full repertory of phonemes or words. A computer can also be used for collecting much of the statistical data needed for these experiments. These data are not at present available and to collect them by any other method would be a long and laborious task.

The principal problems that we intend to investigate by the use of computers are the following:

Collecting trigram frequencies;

Collecting n-gram frequencies;

Comparison of such data obtained from the dictionary and from current speech;

Collection of word transition probabilities for different types of material;

Collection of acoustic pattern data for words;

Comparison of the effect of using trigram and digram frequencies;

Comparison of methods using the phoneme and the word respectively as the recognition unit when the word is defined as:

a. a phoneme sequence;

b. an acoustic pattern.

The use of word transition probabilities.

This work is supported in part by the Air Research and Development Command, U.S. Air Force, under Contract AF 61(514)-1176.

REFERENCES

1. DENES, P. Effect of duration on the perception of voicing. *J. acoust. Soc. Amer.*, 1955, **27**, 761.
2. FRY, D. B. and DENES, P. On presenting the output of a mechanical speech recognizer. *J. acoust. Soc. Amer.*, 1957, **29**, 364.
3. FRY, D. B. and DENES, P. The solution of some fundamental problems in mechanical speech recognition. *Language and Speech*, 1958, **1**, 35.
4. LIBERMAN, A. M. Some results of research on speech perception. *J. acoust. Soc. Amer.*, 1957, **29**, 117.
5. PETERSON, G. E. and BARNEY, H. L. Control methods used in a study of the vowels. *J. acoust. Soc. Amer.*, 1952, **24**, 175.

DISCUSSION ON THE PAPER BY PROF. D. B. FRY AND MR. P. DENES

MR. S. COMET: One of the most difficult problems in vocal handling of speech is the problem of the fundamental and the changes in the fundamental. Many speakers are changing their pitch continuously and often in very wide ranges. Do you think it will be possible to make the machine self-adjusting, in order to follow the changes? The machine should be adjusted to each speaker, but if the speaker changes, himself, the machine will follow him. I think this is one of the worst problems.

DR. R. L. SHUEY: I also have a simple question. I think that to some degree the processes you are using are sequential processes which might destroy information. I would like to know how much information might be destroyed.

Given a waveform, you are in effect breaking it up into sections and getting probability distributions over a finite discrete set, based on those sections. This very likely is not in principle as good as if you kept the whole waveform. For example, if I plotted what might be called the performance figure, as a function of sample length, I could normalize it to 1 for infinite length samples. If I shortened the sample length down till it corresponded to a phoneme, I would expect a performance figure x where x is less than 1. Now, if I increased the sample length to T and used sequential decision processes based on phonemes and the conditional probability of phonemes, I would expect a performance figure greater than x. However, I would expect that that performance figure would be less than what might be obtained if the entire waveform of length T were used. What might be obtainable with the waveform of length T is difficult to define. I would suggest, however, that if Professor Fry were given the waveform, all the time he wanted, and allowed to use any methods he chose, this would be a good indication how much better might be done.

DR. I. C. WHITFIELD: I take it that your statement that the machine, in fact, types out English even if you play music into it, is not any fundamental difficulty. One could provide it with some system by which, if the probability is sufficiently outside what one would expect, it would reject a thing and say that it is not English at all, so it does not seem to be a fundamental difficulty. I want to just make a comment on this point up here, that I think that possibly, although one can do a lot with the wave

form, this is in fact what the human organism is doing in listening to speech. In fact, it does delete the amount of information in the wave very seriously, and this, in fact, is why one has to have a lot of previous statistical information about the nature of the stimulus in order, as it were, to recreate it. In fact, one does not successfully recreate the speech at all, if there is one unknown sound about which one has no previous information.

DR. R. L. SHUEY: I agree to that, but what I am really asking is how much information do you lose by keeping a discrete popularity distribution, instead of continuous ones.

DR. I. C. WHITFIELD: Presumably, you lose no more than in the biological case, if as much. Prof. Fry is, in fact, doing the same sort of thing as the auditory system does, it seems to me.

MR. P. E. TRIER: It is interesting that the authors have abandoned the attempt at pure sound pattern analysis, and have called in statistical aids to improve the results. This strikes me as constituting both a strength and a weakness. The strength is that the human operator in understanding speech makes use of precisely these processes. In listening to a telephone conversation over a noisy channel, the listener says to himself "What is the speaker likely to have said"? But it is also a weakness, because in principle it abandons the attempt at achieving complete accuracy.

I would like to ask two questions arising out of this. First the problem of voice reproduction appears simpler than the reproduction of correctly spelt typescript, and I wonder if, for voice reproduction, the use of a language such as Italian, which consists of pure sounds, would enable one to go away from the statistical methods, and to revert to strict sound pattern analysis. Secondly, with English, I wonder whether Professor Fry and Mr. Denes anticipate that, with a proper extension of this statistical apparatus, in the end, very high accuracy even in English spelling will be achieved.

DR. A. M. UTTLEY: Mr. Trier has suggested there are two possible methods; the statistical with uncertainty and the deductive with certainty. This latter seems to be based on a wrong idea that there is such a thing as a perfect "A", that there are behind it all certain perfect sounds, and that one just has to decide which of these is being spoken. But surely this is not so. There are all the different A's which different people speak. It is easier, perhaps, to make the point with regard to hand writing. Very small children learn the M's, say, of their teacher, their mother and

their father, and then they cannot read Granny's writing. This is learnt later. There is no such thing as a perfect M to recognise. There can be no 100 per cent perfect system. The problem must involve statistics.

DR. J. McCARTHY: What I have to say was suggested by the last two comments. The two alternatives which are suggested seem to be not entirely exclusive - that is the one of using statistics, and the one of attempting to recognise the individual sounds. The third possibility is that we may have to consider the pattern of sounds, but need not consider it statistically. That is, when we examine an English sentence to see whether it is grammatical; we need not make a statistical analysis to see how probable it is as an English sentence, but an analysis to see whether it fits the rules which determine whether a sequence of words is an English sentence. I think that, in taking into account the larger scale pattern in speech recognition, one might also want to do this by means of rules, rather than by transition probabilities or something like that. It seems to me that to carry statistical analysis beyond perhaps second order transition probabilities is not feasible in practice. I wonder if the speakers would care to comment on this.

PROF. C. CHERRY, CHAIRMAN: I think it was John Locke who said that any deductive analysis, which is based upon rules, carries with it concealed inductions involved in the formulation, not only of these rules, but of the initial premisses.

DR. D. M. MACKAY: I wonder if it is useful to say that the ear 'corrupts information'. What the ear and the rest of the system does in hearing speech is, presumably, to make a succession of selective operations from a relatively small set of possibilities. If the ear does this according to the intention of the speaker, then from the standpoint of communication theory there is no corruption or loss of information. The channel is functioning perfectly. In other words, to talk about corruption of information here is to adopt an irrelevant standpoint. I suggest that it makes sense to accuse the nervous system of corrupting information only when it is unable to perform correct selective operations on some ensemble from which its proper functioning requires selections to be made.

In defence of what Mr. Trier has said, it is necessary to distinguish two senses in which a system may "use statistics". In the first place, if different speakers make different noises to represent the same linguistic elements, one must 'use statistics' to determine how many of the thousands of physically different noises are significantly distinguishable, so as to decide how many different phonemes the system has to recognise. This structural analysis cannot be avoided: we have to 'use statistics' here if the system is to be designed at all economically.

Once set up, however, our system could find itself in trouble with 'boderline' signals which would require a statistical mechanism to decide between two or more possible responses. It was this second process of 'using statistics' which I understood Mr. Trier to suggest we could eliminate, if we spoke a language not necessarily uniform in pronunciation, but with sufficiently wide gaps between the ranges representing neighbouring phonemes. Surely he is right?

DR. GREY WALTER: I think most physiologists would agree that all the information received by the brain is corrupted, and statistical information is statistically corrupted. It looks as if we are poised delicately on the razor's edge of corruption. Very minute doses of certain drugs produce a change in the significance level of all signals, resulting in hallucinosis, in which the statistical stores in the brain attain more importance than the moment to moment information. People taking these drugs in very small quantities see lights and hear sounds which do not exist in the physical world outside them, and which they mistake for reality or confuse with it. It seems as though the matching of exogenous and endogenous information is a very delicate process, very liable to second order distortion. To what extent this is true without the action of the drugs is a matter for further experiment; we may be perpetually on the verge of some sort of hallucinatory state, in the sense that all we see or hear is compounded very largely of statistical stores, rather than direct information. As Dr. Mackay says, this mixing and matching depends on the actual ensemble or intention of a nervous system at a particular time.

Another point arose out of Dr. Uttley's suggestion about song recognition by birds. I am not in any sense attacking the notion that the recognition of sound depends on the accumulation of a statistical store, but I think that experiments with very young non-mammalian animals might be misleading. In the first three or four days of the post hatching life of the bird, quite dramatic changes take place in its nervous system, histologically and physiologically, and it is perfectly possible that there may be at birth and for the first few days, a total lack of certain components of the nervous system which mature later.

We can see this even in ourselves; a baby cannot learn to walk at all until the motor control mechanisms of its nervous system have matured toward the end of the first year of life. After that, even if a child has been prevented from walking it learns very rapidly, though it may have very little statistically significant information about the process. I think it may confuse the issue to ignore the importance of physical maturation as it affects different species; in fact this is one of the crucial points in the study of animal learning - whether a new pattern of behaviour is due to the accumulation of statistically significant information about associated

events or to the abrupt physical maturation of a neurohumoral mechanism. At one extreme is the conditioning type of learning in adult mammals, at the other behaviour attributed to Innate Releaser Mechanisms in fish and birds. Consideration of processes analogous to maturation in machines might help to clear up some of these difficulties and suggest better experiments with living organisms.

MR. E. A. NEWMAN: When we listen to someone speaking we might try to judge the meaning of what he is trying to tell us, we might be attempting to recognise him by his voice, or we might attempt to judge his emotional state.

In the first case we are essentially attempting to discover the meaning of the impulses flowing from his cerebral cortex which control muscles involved in speech; these probably transmit less than 40 bits of information per second. We have to measure this information in a very second-hand form.

The control impulses from the cortex control the movements of a complicated interacting muscle set which we could examine. To describe them as they occurred we would need much more than 40 bits of information per second, yet from them can be deduced no more than 40 bits of information/sec about the control impulses. The rest might tell us about the size, shape and tonus of the muscles, but is just noise from the point of view of knowing what was said.

The muscle movements create air movements which emerge from the speakers mouth as sound. Even more information is needed to describe this sound precisely. But, once again, from the point of view of deducing what is being said, most of this is just noise. Consequentially it is nonsense to suggest that every wave in spoken sound contains information about the speakers meaning. Agreed there might be some correlation between apparently different parts of the speech wave, and this might provide useful redundancy in the complicated communication channel from brain to brain, but useful correlation would not appear to be very great.

If we are attempting to recognise the speaker or his emotional state - then of course some of the extra information - about muscle tones etc might be useful. If one remembers that the listener can only absorb about 40 bits/sec however, one questions how much of the extra information is in fact used.

What Fry and Denes have said in effect is that they do not want most of the information that seems to be in speech, since it is mostly not useful. They therefore attempt to look at the sound waves of speech in a crude but relevant way. In doing this surely they are right. Those who say speech waves contain frequencies of thousand of cycles per second - therefore they contain thousands of bits per second of useful information are surely wrong.

DR. R. L. SHUEY: I think my question was misunderstood. I did not say that I thought the ear degraded speech. My question was really aimed at trying

to see how good a model one has in a system like yours.

I should like to take a parallel from English speech. If you look back at Shannon's work, one of the comparisons he made was to compare the Entropy of English based upon the known probabilities of English with the Entropy of English based upon a human being's ability to predict a letter given a previous sentence, etc *(ref.1)*. As you know, the human being was pretty good and much better than the formal statistical process. I am proposing that you do roughly the same thing here. Give a person like Professor Fry all of the waveform up to a certain point and let him do everything he possibly can and see how good an analysis he can make when he has it all. Then see how well he does with one chunk at a time. Compare this with the machine process. In the English text parallel, the human being with his knowledge of possibly even the first half of a book, is very good.

I do not wish to imply that people's brains degrade speech. On the other hand, I do not know that the proposed model is perfect.

Actually, because a machine can in principal record the original waveform, and readjust its processing in light of that waveform, it might in some respects be superior to man. I take it man remembers only what has gone through initial processing. Although that processing is subject to adjustment, it cannot be retroactive.

DR. I. C. WHITFIELD: Arising out of Dr. Trier's remarks about a noisy line, in which one infers subsequently what the speech might be; I think that the situation here is not quite the same, in that it is not a question of hearing a peculiar noise and then thinking to oneself afterwards "it must have meant so - and - so". It is rather that the sensation itself is changed as a result of the stored information; in other words, what you actually *hear* at the time will be affected by your previous knowledge. This, I think, has been demonstrated fairly clearly in the studies of what people hear when certain syllables are spoken, and one can indeed take it right out of the speech area involving semantics into that of, say, musical sounds, where, if you remove the initial transient of a note, the actual quality of the succeeding part of the note is changed or made unclear, and when you put the initial transient back the sensation which you experience is restored. So it is not a question of thinking "What does that peculiar noise mean?", but rather that one's actual experience at the time is changed.

MR. W. LAWRENCE: I think it was a very pregnant idea of Dr. Shuey that this perception is a question of knowing what to throw away, and how to be sure

REFERENCE

1. SHANNON, C. E. Prediction and Entropy of Printed English. *BSTJ*, **50**, Jan 1951.

that you have not thrown anything away. I think we can draw a useful comparison here. Dr. Fry is throwing a lot away, because he says that the sounds he wishes to perceive must fall into 40 categories. If we take another type of analysis-synthesis, telephony, in which we seek to extract from the waves only the possible states of the articulators, and say that the sound we are hearing at a particular moment must correspond to one particular state of the articulators, the vocal organs in general, and if we throw away the possibility that the sound we are hearing was not produced by human articulators at all, we are doing another sort of throwing away information, and it is very interesting to see what we have lost by those two processes. If you insist that sounds have to be categorised as one of 40 English sounds, you throw away the possibility that the human perceiver may reserve judgment on which sound he has heard until he has got the whole sentence of information in his brain at one time. I think you are all familiar with the experience of hearing a sentence, hearing it wrong, hearing the next sentence and suddenly realising that you have heard the first sentence wrong, and then rehearsing it in your mind and realising what the first sentence really was. Of course, any sort of categorising that you do as you go along, makes it vastly more difficult to do that sort of process. If we take the alternative type of analysis, in which you aim at reproducing something that is the equivalent of the actual sounds, you may preserve this possibility of going over the sounds again in your own mind afterwards.

MR. G. L. SHAW (written contribution to papers 2 and 3): As an electronics engineer, my professional sympathies are with Professor Fry and Mr. Denes in their attempts to put more conversations into a given bandwidth by their technique of analysis, coded transmission and synthesis.

Despite my sympathy with their aims, however, I have very serious doubts as to their chances of ever fully realising their intentions because I believe their approach is based on false assumptions and that Mr. Ladefoged is much nearer to a true understanding of the factors involved.

If we go right back to first principles and examine what actually happens when we talk to another person we find that the underlying fact - which is the key to the entire problem - is that we are trying to express the thoughts that are in our mind in such a way as to cause similar thoughts to arise in the mind of our hearer.

In order to achieve this result, we use not only sequences of words in a language familiar to both of us but also emphasis, intonation, facial expressions and gestures. That is, the thought - content is not carried wholly by the actual words, phrases, etc., of normal speech.

There are a number of ways of demonstrating the truth of this last statement. One way is to read a written verbatim report of a conversation one has had and to compare it with one's recollection of the actual conversation. Unless the occasion was a very formal one, the written record (however accurate it may be) is but a pale shadow of the original transaction. A second method is to consider the manner in which the parents of a small child manage to discuss certain matters in the presence of their offspring without the child's being able to understand what is being said. Of the many other possible examples I will give but one more: consider what an accomplished actor can do with a simple sentence.

It is important to bear in mind that both the speaker and the hearer subconsciously know that in normal speech, the thought-content is conveyed only in part by the actual words used and that both rely on this to a large extent. Now, in our present-day technical civilization we have learned that, when our conversation is carried out via a telephone system, we must impose a discipline on ourselves such that the loss of facial expression and gesture is largely offset by more careful choice of words, emphasis and intonation. This is another way of saying that we have to make a deliberate increase in the proportion of the total thought-content that is conveyed by the actual words, phrases and so on, compared to what this would have been had both parties been able to speak face to face.

When even the telephone is denied us and we have to communicate in writing it becomes necessary for us to impose an even stricter self-discipline. We are now forced to rely entirely upon words and word-sequences in order to make our meaning clear to the other person. That is, the words and word-groups must now carry the entire thought-content which we wish to convey if we are to stimulate the desired response in the mind of the reader.

Now, Fry and Denes, by assuming that all the thought-content in a telephone conversation is contained in the words that are spoken and by therefore ignoring emphasis and intonation, are, in effect, converting what the distant speaker is saying into a verbatim report. As we saw above, a written verbatim report has lost a large amount of its significance. A verbatim report, spoken by an expressionless synthesized voice, will have lost so much of its meaning that the listener (who is inevitably seeking human qualities in the sounds and words he hears) will derive little of value from it. By destroying the original sounds the thought-content is also destroyed.

In case it might be thought that I am making too much of this point let us consider a simple sentence, such as "John's book is on the table". This apparently straightforward statement, containing but six words, is capable of conveying many different meanings by variations of stress and intonation yet all of these would be reduced to the one common form if the "verbatim

report" approach were followed because the thoughts of the speaker are not fully expressed by the words alone.

There is so much written material for us all to deal with these days that it is understandable that the written word should have come to be considered as an equivalent of the spoken word. I hope that the foregoing comments have shown that the two are not altogether interchangeable.

Finally, with regard to the actual "mechanism" by means of which we express ourselves in speech (using this term in its broadest sense) and the corresponding arrangements which enable us to perceive what is said to us and to derive meaning from it, I am not qualified to comment. I am somewhat surprised, however, that neither Mr. Ladefoged nor Professor Fry and Mr. Denes seem to be aware of the brilliant work in this field by Dr. K. Koenig. Dr. Koenig's method of teaching the deaf to speak, for example, which arose from many years of experience with handicapped children in Yorkshire and elsewhere, is, I understand, considered to be most advanced and is spreading throughout this country.

MR. P. DENES (in reply) The first question was from Dr. Comet on how the recognizer deals with intonation - that is the variations in the fundamental frequency of the speech wave - and the answer is that it does not. It is a phoneme recognizer. What we are trying to do at the moment is to recognize the sequence of speech sounds as they are said, and we are throwing away all other information. One of the items under this heading of discarded information is the fundamental frequency change, so that when somebody speaks the word "No" into the recognizer, first as a statement, i.e. with falling fundamental frequency, and then the same word "No" as a question, i.e. with rising fundamental frequency, the recognizer will type the phoneme sequence /no/ on both occasions. Obviously the output will not contain all the speech information transmitted by the speaker; for this, other devices dealing with factors such as intonation, stress, etc. would be required. We are investigating these problems but the question does not come under the heading of phoneme recognition.

Before I answer Dr. Shuey's question, I would like to say something about how the machine segments. The acoustic recognition part of the machine on the whole tries to recognize spectral patterns. It has a number of spectral patterns built into it, and it compares the incoming pattern with the stored patterns, it selects the best match, and it keeps on indicating that output as long as the incoming spectrum is best matched by that particular stored pattern that it has selected. As soon as a big enough change takes place in the acoustic input, so that another stored pattern now provides the best match for the incoming pattern, then the recognizer decides that a new sound has arrived. Quite a lot of information is discarded in this way, but it is in the field of acoustic recognition,

with which we are not very concerned. Our aim was to construct a model in which we could influence acoustic recognition by linguistic information, and we are therefore not very much concerned with the amount of information we throw away in the acoustic recognizer. Work is going on elsewhere to explore the question of making better use of the information content of the acoustic wave by taking segments of varying duration and also the question of better utilization of the sequential properties of the sound wave itself.

The next question was whether the recognizer could be made to distinguish speech-like sounds from other sounds and English from non-English speech. Although it might perhaps be possible to achieve these distinctions to a limited extent by making use of statistical and other knowledge about the acoustic wave, these questions are outside the scope of our work. The machine we have described has to recognize the phonemes of English and therefore has to sort all incoming sound waves into one of the 14 categories corresponding to the 14 English phonemes it can deal with.

On the question of the statistical nature of the recognition process, I entirely agree with Dr. Uttley. There is no such thing as a perfect sound. The phonemes exist only on the linguistic level as "perfect" entities and they have only a statistical relation with their acoustic form. The task of the machine recognizer is to recognize the linguistic units and it does this, in the same way as the human listener, by utilizing knowledge of the scatter and overlap of the acoustic characteristics and of linguistic and semantic probabilities.

Mr. Trier also asked if the non-phonetic spelling of English made our task more difficult and if a machine capable of spelling correctly could be achieved. The fact that English is not a phonetically spelt language does not make our task more difficult because we aimed at producing an output which is closely correlated solely with the phoneme sequence; in this way we are not concerned with spelling at all. If one uses a language with a small number of sounds, the task will be easier owing to the small number of categories into which the incoming sound patterns have to be sorted. As to the second part of the question, the phonemic output of the recognizer could be automatically converted into ordinary spelling by extension of the statistical knowledge in the machine. The accuracy of the spelling would be limited by the extent of the statistical knowledge at the disposal of the machine.

I entirely agree with Mr. Shaw's comments: the phoneme sequence is not the only linguistic form in which the speaker embodies information. Intonation, stress, facial expression, etc. are examples of others. In support of this we have stated in other publications that even if a perfect analysis-synthesis telephone system transmitting information about the phoneme sequence only is ever achieved, it will still transmit only part of

the information that is sent from speaker to listener in ordinary conversation, just because these other factors are missing. Inability of an automatic speech recognizer to deal with linguistic elements such as intonation and rhythm will reduce the efficiency of the recognizer for two reasons: 1) it is incapable of recognizing some of the information transmitted by the speaker, as defined in the preceding discussion; 2) the efficiency of phoneme recognition is reduced because the statistical constraints available about the phoneme sequence are weaker in the absence of the information embodied in the other linguistic factors. We are doing research on some aspects of this problem.

There are a number of questions concerning the size of the linguistic unit used as the basis, or as one of the bases, of operation of the machine; the prepared paper contains some discussion of the effect of the size of the unit on the probable success of operation of the machine. If a word is used instead of a sound, or a sentence instead of a word, or several sentences instead of a single sentence, the chance of success is increased for the very reason that Mr. Lawrence has mentioned: that longer units are stored and that it is possible to go back and correct previous provisional decisions on the basis of information coming in over a larger span of time.

PROF. D. B. FRY (in reply) The answer to the question whether it is possible to make the mechanical recognizer reject speech sounds of a language other than the one for which it is intended is that while it would be possible to do so, it would never be economically worth while. The variability of sounds within one language is very great and hence it would probably not be possible to reject "foreign" sounds because they afford a poor match with the stored acoustic patterns. This means that rejection would have to be based on information about the sequence of linguistic units and would require the inspection by the machine of quite long sequences. It would not be an economic proposition to provide facilities for doing this simply in order to reject "foreign" sequences.

SESSION 3

PAPER 3

THE PERCEPTION OF SPEECH

by

P. LADEFOGED

BIOGRAPHICAL NOTE

Peter Ladefoged (M.A. Edinburgh, 1951) is a lecturer in the Phonetics Department of the University of Edinburgh. As such he is concerned with teaching students of languages, speech therapists, and communication engineers. His main interest is in the experimental investigation of speech; he has published a number of general articles on the techniques employed in instrumental phonetics. He has also collaborated with D. E. Broadbent in studies of the acoustic structure of speech sounds; and with members of the Edinburgh University Physiology Department in investigations of muscular activity during speech.

[The author would like to acknowledge his indebtedness to David Abercrombie, Reader in Phonetics in the University of Edinburgh. Many of the ideas discussed in this paper were originally suggested by him in personal communications]

THE PERCEPTION OF SPEECH

by

P. LADEFOGED

SUMMARY

THIS paper reviews recent work on the perception of speech. In contrast to the conclusion of most other writers, it is suggested that speech is not, in any sense, a sequence of small discrete elements. Experiments are described which indicate that listeners find it very difficult to hear a recorded utterance as being composed of a sequence of segments, even when asked beforehand to do so. Further experiments on the perception of speech indicate that vowel quality, nasality and pitch are most simply related to an ordering of the acoustic properties of the stimulus; whereas stress (loudness) and most consonant qualities are ordered more simply in terms of articulatory activity. It is also suggested that the perception of speech involves identifying relative rather than absolute values of both articulatory and acoustic properties.

UNDERSTANDING speech is, in essence, a process of obtaining information from an auditory stimulus. This process involves discriminating between some sounds, and considering other sounds to be similar. But before we can discuss the properties on which a listener bases his judgements, we must consider the influence of the context in which each item of speech is heard.

The context in which an event occurs always affects the perception of that event. In the case of the perception of the sounds of speech, the context influences the listener in two special ways. In the first place there is the effect of the "context of situation" (Firth 1957, *ref. 6*). Every speech event occurs in a particular situation made up of a large number of factors, such as the identity of the speaker and the listener, the time and the place, and all the preceding events, including the relevant speech events. All these factors influence the listener, so that he will probably identify the sounds that he hears not simply as a result of considering them as sounds, but also as a result of expecting them to be some of the

particular set of noises that are likely to make sense in that particular context. Thus if a listener hears the words "Two and two make..." he need not hear any of the sounds of the last word in order to make a reasonable guess as to what it might be. Of course, the listener may be wrong in his guess about the meaning of the utterance; but he nevertheless always has a certain expectancy which will affect his perception of what he hears.

The other way in which the context influences the listener arises from the restrictions imposed not by the necessity of the speech event being meaningful, but by the limits on the combinations of speech sounds which occur in a given language. Thus in most forms of English initial /t/ is never followed by /b, d, f, g, k, l, m, n, p, s.v/ or many other sounds; therefore anyone listening to a speaker of English will have a zero expectancy of hearing any of these combinations. This aspect of the influence of context has been discussed by Fry and Denes (1952, *ref.7)*. Other writers (Miller 1951, *refs.21 and 22*, and Cherry 1957, *ref.2*) have discussed the effect of various forms of contextual influences in slightly different terms. From the point of view of information theory, Shannon and Weaver (1949, *ref.27*) (and, less technically, Weaver 1951 *ref.29*) have pointed out that speech is a Markoff process, i.e. a series of items in which the probability of occurrence of any given item is partly determined by all the preceding items.

It has often been assumed that the listener interprets an auditory stimulus in terms of a sequence of separate items. Halle (1956, *ref.9*) has formulated the resulting problem, as follows:

> "What are the rules that would make it possible to go from the continuous acoustic signal that impinges on the ear to the symbolization of the utterance in terms of discrete units, e.g., phonemes or the letters of our alphabet? There can be no doubt that speech is a sequence of discrete entities, since in writing we perform the kind of symbolization just mentioned, while in reading aloud we execute the reverse of this operation; that is, we go from a discrete symbolization to a continuous acoustic signal."

Similarly Licklider (1952, *ref.20*) in discussing the process of speech perception says that:

> "There is so much evidence that speech is basically a sequence of discrete elements that it seems reasonable to limit considerations to mechanisms that break the stream of speech down into elements and identify the elements as a member, or probably a member, of one or another of a finite number of sets."

Later he makes it clear that he is talking about "discrete elements that correspond to phonemes, syllables, words, and phrases."

Nevertheless there is, in fact, very little evidence for the assumption of discreteness. There are certainly no natural, self-evident, segments of speech; and no one has as yet succeeded in describing a set of operations

which can be used for reducing speech to a set of either articulatory or acoustic segments. Techniques for segmenting the flow of articulatory movements have been suggested by Pike (1947, *ref.25*). But these techniques are concerned with theoretical accounts of a speaker's activity; and they are difficult to apply to an actual speech event. Experimental phoneticians can seldom use them to segment laboratory records such as the cinematographic data of Bogue and Fry (1944, *ref.1*). Nor has anyone devised an adequate technique for segmenting the acoustic form of speech. Many sounds have clear cut boundaries; but some sequences (e.g. that in the word *trying*) can be segmented only by ad hoc judgments which cannot be stated in the form of rigorous procedures that are always applicable. Consequently it is difficult to know why, or in what sense, "speech is a sequence of discrete entities" (Halle 1956, *ref.9*), unless the discrete entities are meant to have some kind of psychological reality.

Segments of speech have been considered to have a psychological reality from two points of view. In the first place the speaker might organise his speech movements in terms of neural patterns corresponding to small segments. Thus Joos (1948, *ref.13*) considerstthat there are three "neuremes" and three partially overlapping innervation waves, corresponding to three segments in a word such as *dad*. On this view the process of speaking consists of ordering an appropriate sequence of neuremes from the static neural patterns which are stored in the speech center. Most American and many European linguists implicitly believe in this form of psychological reality. There is,,however, no experimental evidence which would substantiate it; and it seems equally probable that the movements of speech, like other complex muscular activities, are organised in larger groups which are triggered off as a whole (Craik 1947, *ref.3*).

The second form of psychological reality envisages alphabetic segments as perceptual units. According to this view listeners *hear* a sequence of segments (which may, or may not, be a product of segmentally organised speech movements). It is in this sense that Sapir (1951, *ref.26*) speaks of "The psychological reality of phonemes." He quotes examples from his experience in teaching informants to write their native language which show that "It is exceedingly difficult, if not impossible, to teach a native to take account of... variations which have no phonemic reality for him." But his evidence is not in fact concerned with how a subject hears his native language, but only with how he can learn to describe it in alphabetic terms.

The ultimate basis for the belief that speech is a sequence of discrete units is the existence of alphabetic writing. This system of analysing speech and reducing it to a convenient visual form has had a considerable influence on western thought about the nature of speech. But it is not the only possible, nor necessarily the most natural, form of segmentation. At many times in history methods for reducing spoken language to a written form have been independently devised by different groups of pepple; but only

one of these methods developed into an alphabetic system (Gelb, 1952, *ref.8*). For the practical purpose of providing a simple permanent record of the most important features of an utterance, alphabetic writing represents an immense technical advance over all other systems of writing. But it is not necessarily the best form of analysis to use in discussions on the perception of speech. Because listeners can describe differences between speech events in terms of symbols such as letters of the alphabet, it does not follow that each symbol corresponds to a perceptual event. One word may be symbolised as *cap* and another as *cab;* but the perceptual difference between the corresponding speech events may not be in the last of three segments. (In many pronunciations of English, the difference may be in the length of the vowels; there may be no difference at all in the final consonant articulation.)

There is as yet very little data on the perceptual units involved in listening to speech. But an experiment in progress indicates that listeners are not only aware of large parts of utterances at one moment, but also may find it difficult to consider speech as being composed of a sequence of small segments, even when they have been warned that they are going to be asked to do so. In one version of this experiment, subjects are told that they are going to hear a recording of a sentence such as "There was once a young rat who everybody used to call Arthur." and at the same time a recording of the word *dot*. They are asked to listen carefully so that they will be able to note the precise moment in the sentence when the superimposed word commences. Quantitative results will be reported as an addendum to this paper; at the moment it is possible to say only that subjects are very inaccurate in their judgments. It is not at all uncommon for a subject to consider that the superimposed item occurs two or three words earlier in the sentence than is in fact the case. When asked to verbalise their difficulty in locating the time of occurrence of the superimposed word, subjects typically report that "It all seems to happen at once." It seems that for many listeners a large part of the utterance is in the psychological present.

A possible deduction from these results is that items of speech form a temporal pattern which may be perceived as a whole, much in the same way as other kinds of items (e.g. letters) may form a unified spatial pattern. The formation of temporal patterns has not been adequately described; but it has often been shown experimentally that stimuli may be spread over a period of time and yet all be perceived as present. In a review of the literature Woodrow (1951, *ref.30*) notes that the maximum duration of the psychological present lies between 2.3 and 12.0 seconds. The perceptual span which is involved in listening to many utterances may be below these limits. But it seems probable that items of speech such as syllables, words, phrases and even some sentences, have a perceptual unity; and any theory which attempts to explain the perception of speech in terms of the sequential identification of smaller segments is likely to be unsatisfactory. A corrolary of this view is that it is probably unprofitable to try to build machines which

recognise phonemes, or in any other way to consider the mechanisation of speech processes simply in phonemic terms.

In recent years many linguists (Firth 1957, *ref.6*, Phil. Soc., 1957, *ref.24*) have been developing descriptive techniques which involve the unity of syllables, words, phrases and sentences. (Other linguists have also been developing linguistic techniques which appear to be non-alphabetic in that they describe speech in terms of "simultaneous components" (Harris, 1951, *ref.10*) or "distinctive features" (Jakobson & Halle, 1956, *ref.11*); but the components or features are always considered to be characteristics of phonemes, and both these forms of analysis are actually based on an alphabetic segmental approach.) In addition, phoneticians (Strevens 1958, *ref.28*) have been collaborating with communication engineers (Lawrence 1952, *ref.18*) in the use of parametric specifications of speech which avoid alphabetic segmental procedures. Analyses of this kind may be found to be useful in future descriptions of the perceptual units involved in listening to speech.

We can discuss the sounds of speech in terms of a number of perceptual qualities. Some of these qualities are the same as those we use in discussions of many other auditory stimuli. Thus both speech sounds and musical notes can often be ranked according to their pitch; and the properties of the stimulus which can be correlated with the sensation of pitch are the same for both. But other auditory qualities attributed to speech sounds are not the same as those attributed to other sounds, although in some cases they are referred to by the same term. Thus we can rank both speech sounds and other sounds according to their loudness; but the properties of the stimulus which can be correlated with the sensation of loudness are not the same for speech sounds as they are for most other sounds. In a recent experiment (Draper *et al.*, 1958, *ref.4*) subjects were asked to count from one to twenty, making each word equally loud. Both the subjects and a group of listeners considered that they had followed the instructions not when the sounds had similar acoustic intensities, but when each word was pronounced with equivalent respiratory activity. Some of the words (e.g. *nine*) had an acoustic intensity 10-15 dbs greater than others (e.g. *three*); but subjects were not aware of these differences, although they would have been easily perceived as differences in loudness in anything other than speech sounds. It appears that when listening to speech subjects assess the quality loudness in terms of the respiratory effort required to produce the necessary pressure below the vocal cords.

Speech sounds differ from all other auditory stimuli in that nearly every listener is a potential speaker who is very experienced in making noises similar to those which he is hearing. Consequently the listener is nearly always able to order the auditory qualities of speech sounds in terms of the motor activity which probably produced them as well as in terms of the acoustic properties of the stimulus. Recently there has been a tendency to neglect the articulatory activity of the speaker and to

concentrate on specifying the acoustic correlates of the auditory sensation. This is rather like trying to discuss the visual identification of the shapes of tables and chairs exclusively in terms of the photic stimulation of the retina, and without recognising that the observer might organise his perceptions in terms of objects with which he is familiar. Speech consists of a set of familiar motor activities with equally familiar acoustic attributes. Although we may speak in order to be heard in order to be understood, (Jakobson and Halle, 1956, *ref. 11*), it is also true that we often understand what we hear because we can speak. As any teacher of pronunciation knows, many students have difficulty in hearing the difference between two words in a foreign language until they have been taught to make the difference by articulatory prescriptions. It would seem that in any discussion of the auditory qualities which a listener uses in his identification of speech sounds we must consider both the articulatory and the acoustic properties of speech.

There is very little data on the relative fields of these two sets of properties as bases for the identification of items of speech. But it is at least a plausible hypothesis that the properties which a listener assesses are very similar to the properties which a speaker uses in monitoring and controlling his vocal output. From the speaker's point of view, speech is a skilled bodily action, like touch typing or piano playing, involving an intricate and complicated pattern of muscular activity; and like all other forms of skilled movement, it can be controlled only when the operator has adequate information about the progress of the activity. The speaker has three kinds of feedback about the sounds he is producing: auditory feedback by means of both bone and air conduction; tactile feedback about the contacts between the lips, tongue, velum and other parts of the vocal tract; and proprioceptive feedback about the stretch of the muscles and the movements of the joints. Many aspects of speech may be controlled as a result of information available through more than one of these feedback channels. But it seems that certain aspects are monitored typically via one channel rather than another.

The effect of removing or severely limiting two of the feedback channels was investigated in an experiment conducted within the author's post-graduate class. Five subjects read a test passage and made spontaneous remarks under four conditions: (a) normal; (b) without auditory feedback; (c) with considerably lessened tactile feedback; and (d) without auditory and with reduced tactile feedback. The auditory feedback was eliminated by reproducing a masking noise over headphones; the intensity of the noise was adjusted so that the subject could not hear his own voice even by bone conduction. Tactile feedback was reduced by anaesthetization of the surfaces of the lips, tongue and roof of the mouth by amethocaine hydrochloride lozenges; in most cases the anaesthetization did not extend further back than the soft palate.

The results of this experiment showed that although the speech of most subjects was very disorganised by the removal of both auditory and tactile feedback, it was nevertheless still intelligible. Many possible mistakes were almost entirely avoided. Thus it was seldom that one consonant was replaced by another different enough to change the meaning, although this is a very common kind of speech defect. It would seem, therefore, that the subjects could rely on proprioceptive monitoring to a great extent. Of course, it does not follow that they necessarily do use proprioceptive monitoring in normal circumstances. But is is apparent that this source of information is readily available.

Some features of speech, however, seem to depend more on auditory or tactile monitoring. When auditory feedback was absent many vowel sounds were considerably affected both in length and in quality. There was also inadequate control of the soft palate, resulting in the nasalization of some sounds, and the loss of nasal sounds in other parts where they might have been expected. A third aspect of speech affected by the loss of auditory feedback was the pitch of the voice; there was usually a general increase in pitch, a much narrower range of pitch variation, and a marked tendency to use a simple falling intonation pattern (which might be proprioceptively controlled through awareness of the tension in the laryngeal musculature). On the other hand, all the subjects could sing, fairly well in tune, any song with which they were very familiar. Finally, in the presence of loud masking noise, most subjects tightened the muscles of the pharynx so that there was a striking alteration in voice quality. Of course, both this and the other effects may be due not to the loss of auditory feedback, but to the difficulty of talking in a stress situation created by the unpleasantness of the loud masking noise. But it is at least possible that vowel quality, nasality, intonation and voice quality are subject principally to auditory monitoring. (It is interesting that it is these features which are apt to degenerate most severely in speakers who become deaf. But, of course, what these speakers can or cannot do, is not strictly relevant to a consideration of what non-pathological speakers normally do. The hypothesis suggested here is that normal speakers typically use one feedback channel rather than another for monitoring certain aspects of speech. But when one of the feedback channels becomes damaged, speakers can quickly adapt their behaviour and learn to control their vocal output in a different way.)

Quite different features were affected by the partial elimination of tactile feedback. Subjects varied to some extent, but most of them experienced some difficulty in controlling their lip movements, so that there were often mistakes in the articulation of /p, b, m, f, v/. There were also some variations in articulations involving the tongue. Most subjects experienced difficulty in forming satisfactory /s, z/ sounds; and some of the other fricatives (i.e. the sounds at the ends of the words *bath*, *bathe*, *rush* and

rouge) were often mispronounced. Some of the subjects also used unusual /t, d/ articulations; and there were striking variations in the articulation of /l/.

These results need to be verified by tests involving larger numbers of subjects. But they suggest thatthe control (and perhaps the identification) of vowels and consonants are based on different properties of the sound. The hypothesis indicated on the basis of all our evidence so far is that vowel quality, nasality and pitch are most simply related to an ordering of the acoustic properties of the stimulus, whereas most consonant qualities and features such as stress (Ladefoged 1958, *ref.16*) are ordered more simply in terms of articulatory activity.

Future work will probably show that this hypothesis is subject to several qualifications. But in any case it should be noted that, strictly-speaking, neither vowel nor consonant quality is a perceivable sensation like pitch, which is independent of language and meaning. A listener can judge aspects of phonetic quality only in relation to a particular situation and a particular language or languages. When a listener says that two sounds are the same he seldom means that he can hear no difference between them. He probably means that he can hear no difference of the kind that would distinguish between words; or (if he is a more sophisticated listener) of the kind that would show that the words had been spoken by speakers with different accents. But even when these differences have been taken into account, the sounds may still differ not only in pitch and loudness, but also in personal quality (Ladefoged, 1956, *ref.14*). A listener can easily consider two sounds as being similar in pitch and in loudness and in phonetic quality, but nevertheless as being different in that they were spoken by two different speakers.

It is a basic premise of all phonetic descriptions that sounds can be judged to be equivalent in phonetic quality, although different in personal quality. But the differences which are considered to be merely personal or voice quality differences will vary in different languages and in different situations. Consequently even a phonetician assessing what he may call absolute phonetic quality is not really assessing an absolute perceptual feature. He is merely judging those aspects of the sound which in all the linguistic contexts of situation known to him will convey information other than personal information about the speaker. Thus judgements of phonetic quality can never be considered in the same way as judgements of auditory properties which do not necessarily convey linguistic information such as pitch.

There is a certain amount of additional supporting evidence for the hypothesis that some aspects of speech are identified by reference to the acoustic properties of the stimulus, and others by reference to the articulatory activity which produced the sounds. This evidence is afforded by recent work on the acoustic analysis and synthesis of speech. For many years

differences in vowel quality have been correlated with variations in the formants (i.e. the parts of the spectrum in which there are relatively large amounts of energy). Now both Joos (1948, *ref.13*) and Peterson (1951, *ref.23*) have shown that there is a simple relationship between these acoustic factors and auditorily preceived vowel qualities. Furthermore, in an analysis (at present unfinished) of a number of sets of cardinal vowels (Jones 1956, *ref.12*) spoken by several experienced phoneticians, I have also found that many of the vowels which are considered to be auditorily equidistant are approximately acoustically equidistant when displayed in terms of their formant frequencies. On the other hand, although vowels are often described in articulatory terms, there has been no convincing demonstration that auditory equidistance has any simple articulatory equivalent. So it would seem fair to say that vowel qualities are assessed mainly by reference to an ordering of the acoustic properties of the syllabus.

In the case of consonant qualities, however, the articulatory equivalents of the auditory qualities are of great importance. The identification of consonant quality has been much investigated by many American workers, including notably members of the Haskins Laboratories, a summary of whose work has been given by Liberman (1957, *ref.19*). The view-point of all the American investigators is slightly different from that put forward here, in that the results are stated as specifying the stimuli which can be correlated with the perception of alphabetic segments. But the basic data of their experiments are usually responses of subjects concerning mono-syllabic synthetic utterances, which may have been perceived as unitary wholes. Consequently although the data may provide evidence about the *identification* of segments, they do not demonstrate that these segments are *perceived*. In fact, a great deal of the Haskins work can be interpreted as indicating that it is unlikely that subjects perceive in segmental terms.

The Haskins work involves correlating subjects' responses with systematic variations in electrically synthesised speech-like sounds. During the course of this work, they have found that the perception (identification, in my terms) of a given consonant quality often requires a considerably different acoustic stimulus in slightly different contexts. They suggest that this is because in the different contexts the articulation may remain more or less the same, but "the correspondence between articulation and sound is not always one to one. Small differences in articulation often cause very large differences at the acoustic level." (Liberman 1957, *ref.19*). They come to the general conclusion that "speech is perceived by reference to articulation - that is, that the articulatory movements and their sensory effects mediate between the acoustic stimulus and the event we call perception" (ibid.) Since the Haskins investigations are mainly concerned with consonant quality, this view is in general agreement with the hypothesis put forward in this paper.

Most phonetic qualities probably depend on relative rather than on absolute properties of speech. In some cases this is obvious; thus the identification of a particular kind of intonation pattern (e.g. a tune conveying slight doubt, used when repeating what another speaker has just said) does not depend on the absolute pitch of the voice (which might be that of a man or a woman), but on the relative pitch changes. A recent experiment (Ladefoged & Broadbent, 1957, *ref.17*) has also shown that subjects identify a given sound as, e.g. *bit* as opposed to *bet* or *bat*, when the acoustic structure of that sound is in an appropriate relation to that of neighbouring sounds. It is quite possible for the acoustic stimulus which is identified as *bit* on one occasion to be heard as *bet* on another occasion when it is presented in another context. Similarly there are indications (Ladefoged 1957, *ref.15*) that consonant articulations are ordered not in terms of any absolute features of the shape of the vocal cavities, but in terms of relative positions. It is apparent, therefore, that listening to speech involves identifying patterns of properties. The mechanisation of the underlying thought processes is a considerable undertaking. Learning to talk and understand what people say is a long complex process. It will probably be many years before we can expect a machine to do it for us.

BIBLIOGRAPHY

1. BOGUE, Y. J. and FRY, D. B. Movements of the tongue in speech *Endeavour*, 1944, **3.2**, 104.
2. CHERRY, C. On human communication. *Mass. Inst. Tech. Press*, (1957).
3. CRAIK, K. J. W. Theory of the human operator in control systems *Brit. J. Psychol.*, 1947, **38.2**, 56.
4. DRAPER, M. H., LADEFOGED, P. and WHITTERIDGE, D. Respiratory muscles in speech. *J. Speech & Hearing Research*, in press, 1958.
5. FANT, G. Modern instruments and methods for acoustic studies of speech *Acta Polytechnica Scandinavica* (Physics, 1 1958).
6. FIRTH, J. R. Papers in linguistics, 1934-1951. *Oxford Univ. Press.* (1957).
7. FRY, D. B. and DENES, P. Mechanical speech recognition in Communication Theory, Ed. by W. Jackson. *Butterworth, London*, (1952).
8. GELB, I. J. A study of writing. *Kegan Paul, London.* (1952)
9. HALLE, M. [Review of Hockett's Manual of Phonology] *J. Acoust. Soc. Amer.*, 1956, **28**, 509.
10. HARRIS, Z. S. Methods in structural linguistics. *Univ. of Chicago Press.* (1951).
11. JAKOBSON, R. and HALLE, M. Fundamentals of language. *Mouton & Co, 'S-Gravenhage.* (1956).
12. JONES, D. An outline of English phonetics. *Heffer, Cambridge.* (1956).

13. JOOS, M. Acoustic phonetics *Language*, 1948, **24**, supplement.
14. LADEFOGED, P. The classification of vowels *Lingua*, 1956, **5**, 113.
15. LADEFOGED, P. The use of palatography *J. Speech & Hearing Disorders*, 1957, **22**, 764.
16. LADEFOGED, P. Syllables and stress *Miscellanea Phonetica*, in press. (1958).
17. LADEFOGED, P. and BROADBENT, D. E. Information conveyed by vowels *J. Acoust. Soc. Amer.*, 1957, **29**, 98.
18. LAWRENCE, W. The synthesis of speech from signals with a low information rate in Communication Theory ed. by W. Jackson. *Butterworth, London.* (1952).
19. LIBERMAN, A. Some results of research on speech perception *J. Acoust. Soc. Amer.*, 1957, **29**, 117.
20. LICKLIDER, J. C. R. On the process of speech perception *J. Acoust. Soc. Amer.*, 1952, **24**, 590.
21. MILLER, G. A. Speech and Language in A handbook of experimental psychology ed. by S. Stevens. *Wiley, New York.* (1951).
22. MILLER, G. A. Language and communication. *McGraw-Hill, New York.* (1951).
23. PETERSON, G. E. The phonetic value of vowels *Language*, 1951, **27**, 541.
24. Philological Society, [Special volume] Studies in Linguistic analysis. *Blackwell, Oxford.* (1957).
25. PIKE, K. Phonetics. *Univ. of Michigan Press.* (1947).
26. SAPIR, E. Selected writings ed. by D. Mandelbaum. *Berkley, Los Angeles.* (1951).
27. SHANNON, C. E. and WEAVER, W. The mathematical theory of communication. *Univ. of Illinois Press.* (1949).
28. STREVENS, P. The performance of Lawrence's six-parameter speech synthesiser. *Revista do Laboratorió Fonética Experimental, Coimbra* (in press). (1958).
29. WEAVER, W. Information theory to 1951 - a non-technical review. *J. Speech and Hearing Disorders*, 1951, **16**, 166.
30. WOODROW, Time perception in A handbook of experimental psychology. ed. by S. Stevens. *Wiley, New York.* (1951).

DISCUSSION ON THE PAPER BY MR. P. LADEFOGED[ø]

MR. LAWRENCE: We have said a lot about the impossibility of perceiving a speech communication, piece by piece, as we go along, and the general tone of the meeting has been that we cannot do it even as well as we thought we could. There is, however, one important step in the process of perception that can be taken strictly as a reversible transformation. Without taking into account any probabilistic data whatsoever, and without having a backward look at what has already come in some time ago and judging it in the light of what we now know, we can still arrive at a description of speech which preserves all perceptually valuable features, but is vastly simpler than the acoustic wave form. I call this the parametric description of speech. It specifies such details as the frequencies of resonance of the vocal tract, whether the vocal chords are vibrating and if so at what frequency. I say with some confidence that such a transformation is possible since at S.R.D.E. we have a machine that does just this, and we are able to use the output of this machine to recreate the perceptual equivalent of the original speech. This transformation, from an information theory point of view, is a tremendous reduction in the bit rate - it's a reduction of the order of thirty to one. The point I wish to make now is that, since nothing has been lost in this reversible transformation, you are still just as well qualified to make your probabilistic judgments as you were before. It may well be that speech is held in the short term memory in a form like this.

DR. F. S. COOPER: I find some difficulty in commenting on Mr. Ladefoged's paper as a whole since I am in agreement with what he has to say in almost every second paragraph. In some of the odd paragraphs, however, he questions the existence of the phoneme as a perceptual entity. I rise in defense of a tried and true friend.

If I have understood Mr. Ladefoged correctly, about half of his arguments against the realistic existence of perceptual elements of phonemic size -- "alphabetic segment" is the term he uses — are based on difficulties in finding discrete segments of that order of length in the acoustic stream or the articulatory gesture.

The difficulties are real enough, but are they relevant to the argument?* Let me assume, as is often done, that the act of speaking begins with the

---

ø Mr. Ladefoged demonstrated the experiment he described on page 402.

* The argument assumes, as I have assumed also in these comments, that perception and production are intimately linked. This point is discussed at length in *ref.1*.

selection of a sequence of discrete elements from a "phoneme store" -- a black box somewhere in the head. The hypothetical elements so chosen are then recoded successively as neural messages, muscular contractions, changes in the shape of the articulatory tract, and finally as the acoustic stream of speech. We may even recode once more (i.e., sound into spectrogram) for convenience in dealing with speech as a visual pattern.

Now what has happened, in these successive recodings, to the hypothetical elements selected from the phoneme store? One might reasonably expect to find separate and distinct entities at the neural level and perhaps in the action patterns of muscular contraction but certainly we would not expect a one-to-one correspondence to carry over into the shape of the vocal tract. The shape at any one instant will depend not only on the information arriving at that instant but on past history as well. In short, we cannot expect the boundaries of our hypothetical elements to survive the stage of encoding into articulatory shape. It follows that the acoustic output, too, will be "smeared", even though the input was discrete, by assumption. It follows also that we cannot infer -- as Mr. Ladefoged seems to have done -- that, simply because the output is found to be smeared, the input could not have been discrete.

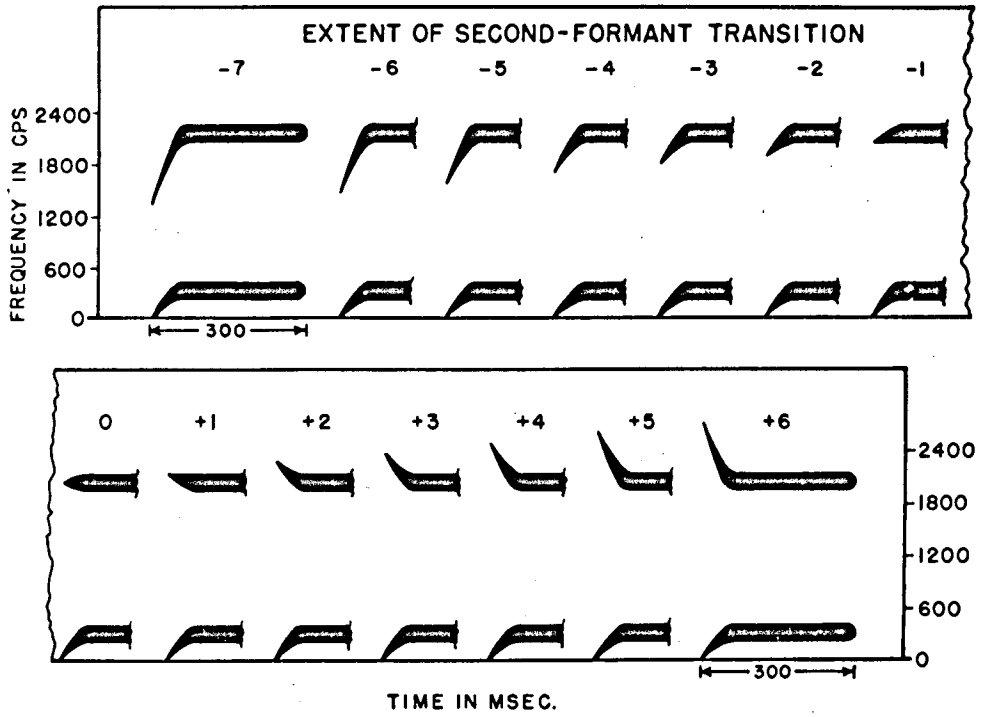


Fig. 1. The synthetic speech patterns arranged in order of direction and extent of second-formant transition from -7 to +6, in arbitrary units of 120 cps difference in starting frequency.

Perhaps, as a constructive counterargument on behalf of the phoneme as a perceptual entity, I may be allowed to sketch a little experiment *(refs.2 & 3)*. My argument will be not unlike Mr. Ladefoged's in that I shall invoke discreteness as evidence for perceptual reality. I shall not, however, seek this discreteness along the time axis, where it is not to be expected in any case, but rather along an acoustic dimension on which a set of spectrographic patterns can be arranged.

The experiment goes like this: we draw a set of synthetic speech patterns, like those in the *fig.1*, that differ systematically in the shape (or starting frequency) of their second-formant transitions. This is the physical variable. When one listens to the sequence of sounds that are produced from these patterns, he hears syllables that start with "ba" and then shift rather abruptly to "da" and again to "ga". In the first half of the experiment, we present these sounds in random order to a group of listeners and ask them to identify the consonant as "b", "d", or "g". The group agreement is quite good; it divides the stimulus dimension into three regions, as shown in the upper part of the *fig.2*. In the second half of the experiment, the same sounds are presented in ABX format (where A and B are adjacent or nearly adjacent patterns and X is a repetition of either A or B). The subject is

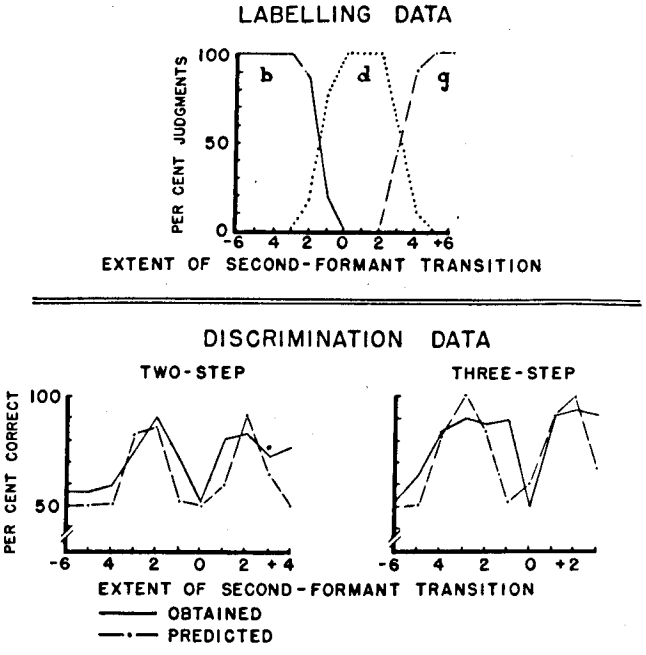


Fig.2. Upper: group identifications of the sounds illustrated in diagram 1. Lower: Discrimination of pairs of sounds that differ by two (or three) units in extent of second-formant transition. The predicted values are based on the labelling data in the upper half of the diagram.

asked to say whether X is A or B, basing his choice on any difference whatever that he can hear in the sounds. The results, given in the lower half of the diagram, show that it is very difficult to discriminate sounds that lie within the same phoneme category but relatively easy to hear differences of comparable size at a boundary between categories. If we ask how much of the discriminative behavior is due to the listener's ability to name the sounds and how much is due to any other cause, we find that the margin left over for improvement on any basis other than naming is very small. In short, the listener behaves almost as if the only basis he has for hearing differences in the sounds lies in the different names he can assign to them.

What does such an experiment mean? It tells us that discrete categories at the level of behavior have been carved out of something that varied continuously at the acoustic level. I am inclined to interpret this as evidence for the perceptual reality of these categories and of the phonemes whose names they bear.

DR. D. M. MACKAY: Despite the delight which all of us must have felt on the evidence of this demonstration, I am not sure that I agree with one conclusion drawn from it. This is really an experimental question. Suppose you were to articulate your test-sentence more and more slowly, and superimpose the hiss in the same place each time, then presumably there would come a point below which your subject could correctly locate the hiss every time. The fact that human beings cannot hear things phoneme by phoneme at normal speeds, therefore, would not seem to prove that an electronic mechanism, which is not speed-limited, could not discriminate, and if necessary type, phoneme by phoneme.

PROF. D. B. FRY: The major objection to Mr. Ladefoged's paper is the weakness of its logic. He says at one point that items of speech are perceived as "a temporal pattern as a whole"; this implies discreteness since, if there are "wholes" they must be discrete wholes. The clue to this seeming contradiction is to be found in another place where he says it is unwise to work with segments smaller than phrases, words or syllables. His objection therefore concerns, not the principle of discreteness, but the size of the

REFERENCES

1. LIBERMAN, A. M. Some results of research on Speech Perception, *J. Acoust. Soc. Am.*, 1957, 29, 117.
2. LIBERMAN, A. M., DELATTRE, P. C., COOPER, F. S. and GERSTMAN, L. J. The Role of Consonant-Vowel Transitions in the Perception of the Stop and Nasal Consonants. *Psychological Monographs*, 1954, 68, No. 8.
3. LIBERMAN, A. M., HARRIS, KATHERINE S., HOFFMAN, H.S., GRIFFITH, B. C. The Discrimination of Speech Sounds within and across Phoneme Boundaries. *J.exp. Psychol.*, 1957, 54, 358.

unit adopted as a basis for analysis. It is quite clear, however, and must indeed be well-known to Mr. Ladefoged, that once you accept discrete units as a basis, the same difficulties will re-appear no matter what the size of the unit. He seems to be urging us to discard the phoneme entirely in favour of the syllable, but exactly the same difficulties arise in the case of the syllable as in the case of the phoneme.

With respect to the demonstration he has given us, I would suggest an interpretation very different from his. We hear continuous speech into which a short burst of noise is injected at some point. The appreciation of the noise is immediate for everyone but taking in the speech involves precisely the process of phoneme scanning whose existence Mr. Ladefoged denies. The rate at which this phoneme scanning is carried out varies with individuals and each of us assigns the occurrence of the noise to the point he happens to have reached in his scanning of the speech.

MR. R. L. GREGORY: Dr. R. Warren and I have recently done an experiment in Cambridge which seems relevant here. I will not attempt to *explain* it, but I think the finding might be of some interest. The experiment is simply to have a word repeated over and over again. We had a word on an endless loop on the tape recorder - some such word as "apple" - and a very interesting point is that the word is heard to change spontaneously. It may have 30 different variations; it may be heard, for example, as *a-pull*, and then quite suddenly it may be heard with, say, a pronounced foreign accent. We think this may be analogous to the Neckar cube phenomenon, where you get spontaneous reversal of visual figures. We suspect that this auditory effect is relevant to the recognition of speech, for it suggests a searching for the "best bet" on the available information.

MR. J. MORTON: The main point I would like to make is that, in investigating this sort of field, there seems to be a great danger that experiments using individual phonemes will give results which indicate that we perceive sound in terms of these individual phonemes, which is what Professor Fry seems to be saying. If, on the other hand, you are considering the total perceptual process, then the results will indicate that we are dealing with larger patterns of sound at any one time. In this way contradictory results will arise depending upon the approach.

The superiority of the more wholistic approach was illustrated subjectively for me in Mr. Ladefoged's demonstration. Originally I thought the sound had come from a natural break in the sentence, - "We thought it was Jane . . pause . . who could be brave and in the team", and I thought

REFERENCE

1. WARREN, R. M. and GREGORY, R. L. An Auditory Analogue of the Visual Reversible Figure. *Amer.J.Psychol.*, 1958, 71, No. 3, 612.

the 's' came after 'Jane . .'. Then it occurred to me that I had chosen the natural break and came to the conclusion that this was because I had perceived the words preferentially as whole phrases. On re-scanning my memory I changed the position of the 's' to above 'could', much nearer the true position. One might say that in order to do this I had deliberately to drop the point of my perception below the level of integration at which I normally perceive, i.e. take an analytical attitude.

This point also arises in the study of reading. You find the most powerful readers, in terms of efficiency, can take in at a single glance, phrases of an increasing number of words. Very slow readers, reading word for word, usually vocalise considerably and seem to transfer the printed symbol to an auditory one before dealing with the content. Faster readers appear to jump this step and go straight on to a more comprehensive grouping: of the ideas behind the words, rather than the words themselves - and fit these ideas into some kind of existing thought scheme. I hope to get some experimental evidence on this point later. This concept is very close to what Dr. Mackay considered in his paper when talking about analogue representation in the brain.

The confusion may most simply be resolved by distinguishing clearly between mechanisms in the brain concerned with various kinds of analysis and integration at different levels of complexity, and perception as a normal activity. To investigate the latter, one must use normal stimuli, which does not mean isolated phonemes.

MR. G. L. SHAW: (Written contribution to Session 3, papers 2 and 3 - See discussion on Session 3, paper 2.

MR. P. LADEFOGED (in reply): Dr. Cooper gave us a masterful account of an experiment which, he suggested, showed that listeners were perceiving separate phonemes. But in fact his subjects were listening to sounds like whole syllables; and their responses indicate whether they can tell two syllables apart. It is interesting that when subjects are trying to distinguish between members of a range of acoustic stimuli in terms of the three possibilities /ba, da, ga/ they can more easily distinguish between some of the stimuli than between others. But this does not show that they interpret each of these acoustic stimuli as being composed of two elements, e.g. a /b/ and an /a/.

In reply to Professor Fry and Dr. Mackay, the point about the confusion between discreteness and the size of the discrete unit is well taken. Prof. Fry is quite right: I am not objecting to discreteness, as such. My real objection is to considering that the units involved in perception are anything like the same size as phonemes. I base this objection both on the results of perceptual experiments such as I demonstrated earlier on, and

on evidence, which I cannot present in detail now, that in talking long complex muscular patterns are triggered off as a whole. As Dr. Mackay has rightly pointed out, this is largely irrelevant to the problem of constructing a machine to recognise speech. It may well be that the phoneme is the most suitable unit for this purpose. But my basic disagreement with Prof. Fry remains; I think that it is very unlikely that human beings recognise speech in terms of linguistic units such as phonemes.

SESSION 3

PAPER 4

TWO THEOREMS OF STATISTICAL SEPARABILITY IN THE PERCEPTRON

by

DR. F. ROSENBLATT

BIOGRAPHICAL NOTE

Frank Rosenblatt, born in New Rochelle, New York, U.S.A., July 11, 1928, graduated from Cornell University in 1950, and received a PhD degree in psychology, from the same university, in 1956. He was engaged in research on schizophrenia, as a Fellow of the U.S. Public Health Service, 1951-1953. He has made contributions to techniques of multivariate analysis, psychopathology, information processing and control systems, and physiological brain models. He is currently a Research Psychologist at the Cornell Aeronautical Laboratory, Inc., in Buffalo, New York, where he is Project Engineer responsible for Project PARA (Perceiving and Recognizing Automaton).

TWO THEOREMS OF STATISTICAL SEPARABILITY IN THE PERCEPTRON

by

DR. FRANK ROSENBLATT

SUMMARY

A THEORETICAL brain model, the perceptron, has been developed at the Cornell Aeronautical Laboratory, in Buffalo, New York. The perceptron is a probabilistic system, capable of learning to recognize and differentiate stimuli in its environment. Previous reports have covered the theory of a class of perceptrons based on fixed-threshold neurons, similar to the McCulloch-Pitts model. The present paper introduces the concept of a "continuous transducer neuron", and outlines the proof of two theorems which indicate that a properly designed perceptron will be capable of spontaneously forming meaningful classifications of the stimuli in its universe, without being taught by an experimenter.

---

1. PROBABILISTIC MATHEMATICS VS. SYMBOLIC LOGIC

ONLY a few months before the Office of Naval Research began its support of the perceptron program, at the Cornell Aeronautical Laboratory, John von Neumann, one of the most outstanding advocates of the proposition that man might some day achieve an artificial device working on the same principles as the human brain, wrote the following prophetic passage *(ref. 4)*:

> "Logics and mathematics in the central nervous system...must structurally be essentially different from those languages to which our common experience refers ... When we talk mathematics, we may be discussing a *secondary* language, built on the *primary* language truly used by the central nervous system. Thus the outward forms of *our* mathematics are not absolutely relevant from the point of view of evaluating what the mathematical or logical language *truly* used by the central

nervous system is... Whatever the system is, it cannot fail to differ considerably from what we consciously and explicitly consider as mathematics."

What von Neumann is saying here deserves careful consideration. The mathematical field of symbolic logic, or Boolean algebra, has been eminently successful in producing our modern control systems and digital computing machines. Nevertheless, the attempts to account for the operation of the human brain by similar principles have always broken down under close scrutiny. The models which conceive of the brain as a strictly digital, Boolean algebra device, always involve either an impossibly large number of discrete elements, or else a precision in the "wiring diagram" and synchronization of the system which is quite unlike the conditions observed in a biological nervous system. I will not belabor this point here, as the arguments have been presented in considerable detail in the original report on the perceptron *(ref. 3)*. The important consideration is that in dealing with the brain, a different kind of mathematics, primarily statistical in nature, seems to be involved. The brain seems to arrive at many results directly, or "intuitively", rather than analytically. As von Neumann has pointed out, there is typically much less "logical depth" in the operations of the central nervous system than in the programs performed by a digital computer, which may require hundreds, thousands, or even millions of successive logical steps in order to arrive at an analytically programmed result.

Those readers who are familiar with the concept of the perceptron know that it is a model of a system which is primarily concerned with the recognition of the forms, sounds, and other stimuli which make up the ordinary physical world, as we know it through our own senses. The theory upon which this system is based is called the "theory of statistical separability". The mathematics upon which this theory stands, has much more in common with the mathematics of particle physics than with the mathematics of digital computers. The reason for this is fundamental: Boolean algebra, or symbolic logic, is well suited to the study of completely describable logical systems, but breaks down as soon as we attempt to apply it to systems on which complete information is not available. If we lack a detailed wiring diagram, but know only the statistical parameters, or probabilities of connection within a logical net, then the only way we can use Boolean algebra to determine the probable response of the system would be by complete enumeration of every possible connection diagram which meets the parametric constraints, whereupon we could actually count the number of alternatives which respond in each of the logically possible ways. In dealing with systems of any complexity, the number of possible connection diagrams becomes, for all practical purposes, infinite, so that we can not use this enumerative approach in practice, even though it may be possible in principle. Probability theory, on the other hand, is specifically designed to permit us

to make precise statements in the absence of complete information. If we know only a few parameters of a statistical distribution, we may be able to make highly accurate statements about the mass behavior of a collection of events, a logical network, or a nervous system, *even though we have never unravelled the detailed wiring diagram for any particular case.*

2. THE IMPORTANCE OF PERCEPTUAL PROCESSES FOR AUTOMATA

We have said that the perceptron is primarily concerned with the recognition of stimuli, or patterns, in its environment. In this, it is fundamentally different from any digital computer. Since computers are very good at something which people, by and large, do very badly, that is, arithmetic, they have been popularly represented as "giant electronic brains". This comparison seems to me an unfortunate one. It suggests to many people that because a computer does certain things that the brain can do, the brain must work something like a computer. I have already indicated above that I consider this position to be untenable. But in order to understand the unique capabilities of the perceptron, it might be helpful to consider the sort of thing that *can* be done by digital computers.

Computers, in general, are designed to follow rules. If we can set up rules for multiplication, we can design a computer to multiply. If we give this computer any two numbers, even if these specific numbers have not been considered by the designers of the machine, it can multiply them, and form the correct product. But this does not really satisfy our idea of original thinking, or intelligence. There has been no *discovery* involved; the correct answer simply follows from the fact that the rules of multiplication are completely universal, and apply to all numbers. Similarly, if we are ingenious enough to write a set of exact rules for minimizing the cost of some business operation, we can program a computer to minimize cost, and other such complex problems. In fact, if we can analyze the way in which people play chess sufficiently well to write an explicit set of rules for chess strategy, we can get a computer to play chess, as is being done with increasing success in recent programs.

But in all of these examples, the computer is following rules which are the result of human observation and analysis. Computers seem to share two main functions with the brain:

(a) Decision making, based on logical rules
(b) Control (as in guidance systems, automatic assembly lines, etc.) again based on logical rules.

The human brain performs these functions, together with a third: *interpretation of the environment.*

Why do we hold interpretation of the environment to be so important?

The answer, I think, is to be found in the laws of thermodynamics. A system with a completely self-contained logic can never spontaneously

improve in its ability to organize, and to draw valid conclusions from information. While there are certain trivial cases where this may appear questionable (for example, we might deliberately design a program to give wrong answers for the first ten trials and then modify itself to perform correctly from the eleventh trial on), by and large it seems to be a valid generalization. Even in the above case, the improvement after the tenth trial was anticipated and deliberately built into the program, so that it did not actually arise spontaneously. Spontaneous changes in such a system will, in general, lead to a deterioration, rather than an improvement, in its performance. On the other hand, a system which is capable of reorganizing its own logic, to correspond to a logical organization which already exists in the universe around it, takes on very different properties indeed. Such a system *can* improve (if it is properly constructed) by observing and learning from the organization of the surrounding world. The human brain is such a system. It is this ability to interpret the environment which allows the human brain to recognize and devise the logical rules which are applied by the computer. Conceptualization of the environment is the first step towards creative thinking.

So far as I know, the only machine prior to the perceptron which has shown itself to be capable of *spontaneous* improvement (as opposed to learning under the tutelage of an experimenter) has been Ashby's homeostat *(ref. 1)*. The homeostat, however, is not really a case in point here, because it is not really concerned, as we are, with the representation of meaningful information, but rather with the maintenance of an optimum state within the system.

The perceptron, as originally developed, would also have been incapable of spontaneous concept formation. The original perceptron could be taught to perceive differences between stimuli, by a process similar to that employed in training a dog. If you want to train a dog to come when called, you put him on a long leash, you say, "Come here", or "Come, Rover", and you pull him towards you. If you want him to sit down, you say, "Rover, sit", and push his tail down. In other words, we *force* the desired response. When he begins doing these things spontaneously, this indicates that we have passed on our own recognition of the difference between the words "Come here" and "Sit down", and that the dog can now perceive the distinction. This is really identical with the process that we originally studied with the perceptron.

It actually took only a very slight change in the dynamics of the perceptron to convert it into a system with very different capabilities, actually the first machine which is capable of having an original idea. It is the conversion to this new system which I wish to discuss for the remainder of this paper.

The key to this transition to a spontaneously organizing perceptron came with a recasting of our original mathematical analysis in a new and more elegant form, made possible by a revised concept of the basic unit, the

neuron. Let me begin, therefore, by presenting the rationale for this new concept.

3. THE CONTINUOUS TRANSDUCER NEURON

The logical model of the neuron which has been used almost universally in theoretical brain models during the last few decades is the "McCulloch-Pitts neuron" *(ref. 2)*. This assumes an all-or-nothing response, depending on whether the signal at some time, t, is above or below threshold. Moreover, and this is the most critical simplification assumed by the model, the cell is generally allowed to fire only at integral values of t, where time is measured in some convenient unit. Now actually, as is well known, a cell in the central nervous system does not respond in so simple a fashion. Such a cell is typically under continuous bombardment by a great number of impulses, which summate temporally as well as spatially. These impulses, eventually, may produce a region of depolarization in the membrane, and when this happens, the cell fires. As a consequence of this, the *frequency* of the cell is likely to be roughly proportional to the net, or mean intensity of the stimulation received. Lightly stimulated, the cell may respond at a very low frequency; under intense stimulation, the frequency increases. Thus, even though each impulse generated by the cell may be of the same amplitude, the cell is actually acting as a continuous transducer of the stimulus energy, if we measure the energy transmitted as a rate per unit time. Such a cell may carry considerably more information than we would suppose from a simple consideration of its discontinuous on-off properties.

The transfer function of a continuous transducer neuron, a_i, at time t, is equal to $\alpha_i(t) \cdot v(t)$, where v_i (the "value" of the neuron a_i) is a scale factor which may fluctuate with time, and which determines the relative amplitude of the output. We will have more to say about this "value" presently. A simplified continuous transducer neuron is shown in *fig. 1* (a). Note that there are an equal number of excitatory and inhibitory inputs, so that the input signal α, will be zero if stimulation is uniform over the entire sensory field. If there is a gradient of stimulation, or a localized region of stimulation such that the total excitatory component (Σx) is greater than the total inhibitory component (Σy), the input signal, α, will be positive. If the input from a given stimulus is primarily inhibitory, a will be negative.

For use in the perceptron, one additional assumption must be made about the logic of the neuron; it must be given a memory. This memory takes the form of changes in the magnitude of the "value", v. The value, v_i, of the neuron a_i, is represented by a stored quantity, which might be represented physically by an electric charge, the position of a potentiometer, the light

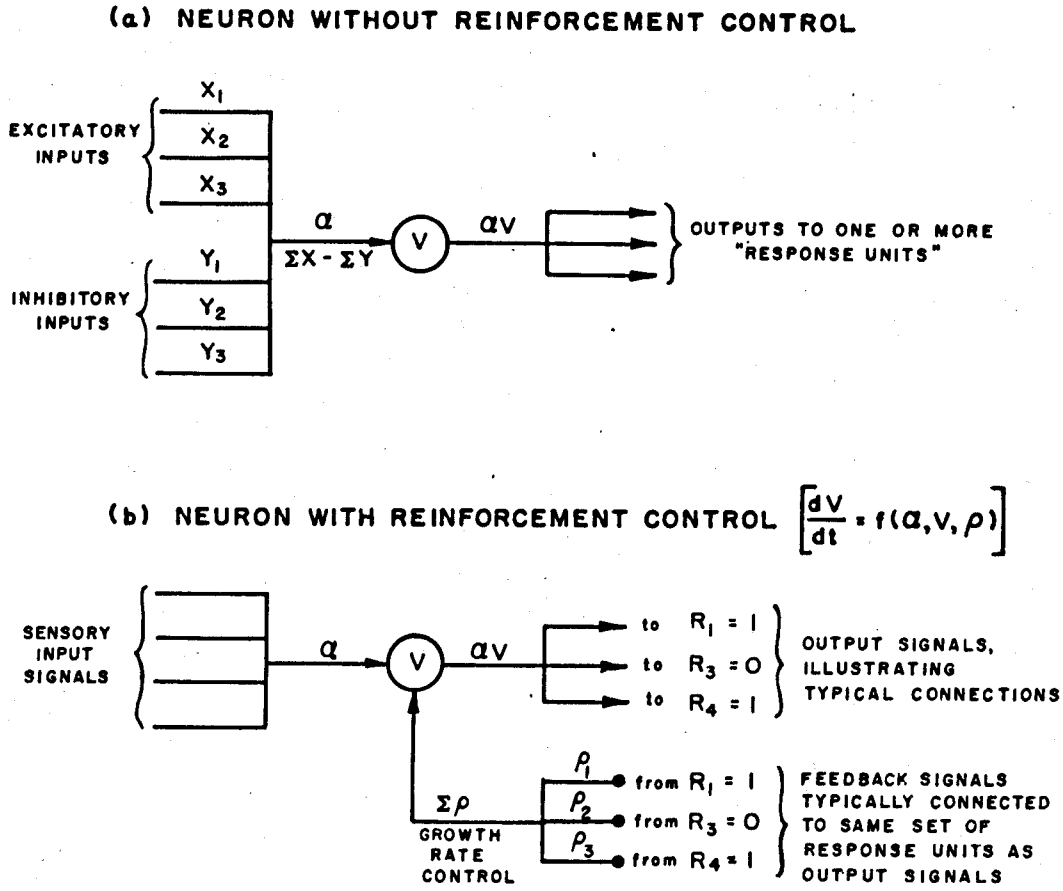


Fig. 1. Continuous Transducer Neurons

flux through a variable aperture, or any other convenient means. If we are thinking in terms of a biological analog, the "value" might be the degree of polarization of the membrane, the energy reserve of the cell, the volume of cytoplasm, or any other enduring condition which will affect the potency, amplitude, or frequency of the cell's output. The higher the "value", the more "powerful" the output of the cell, measured by whatever variable happens to be relevant for the system in question. For a perceptron, where the output function is equal to $\alpha\nu$, we assume that the amplitude of the output pulses will be proportional to the value.

The important question about the value, from the standpoint of memory storage, is how it changes, as a function of the activity of the system. *Figure 1(b)* shows a continuous transducer neuron with an additional input signal, labelled $\Sigma\rho$, which controls the growth rate of the neuron. The components of $\Sigma\rho$ (i.e., ρ_1, ρ_2, etc.) are typically feedback signals, which originate from the next logical layer of cells, which are called "Response units", or *R*-units. In the fixed threshold neurons used in the earlier perceptron systems *(ref. 3)*, the growth of the value was a function which depended on whether or not the cell was active. If the cell was active (α greater than threshold), an increment would be added to ν, which depended primarily on the presence or absence of a feedback signal corresponding to $\Sigma\rho$. In the continuous transducer system, three basic forms of growth function are of particular interest:

$$\frac{d\nu}{dt} = \alpha\Sigma\rho \tag{1}$$

$$\frac{d\nu}{dt} = \alpha\Sigma\rho - \delta\nu \tag{2}$$

$$\frac{d\nu}{dt} = \Sigma\rho(\alpha - \delta\nu) \tag{3}$$

The first of these equations is a growth function with a zero decay component. It is equivalent to (2) and (3) with δ set at zero. In the second two cases, there is a decay component proportional to the current magnitude of the value. δ, the decay coefficient, is a constant less than 1. All fixed-threshold perceptrons considered previously *(ref. 3)* share the characteristic of the first of these growth functions that the value of a neuron, or association cell, can continue to grow without bound. Note that over a set of neurons, since α is as likely to be negative as positive, the expected rate of growth will remain zero. Nonetheless, the *variance* of the value over the set of cells will tend to increase towards infinity, as time goes on, unless there is a decay function which increases monotonically with ν, as in (2) and (3). The importance of this distinction will become clear presently.

4. ORGANIZATION OF A PERCEPTRON

Let us now consider the organization of a very simple perceptron, made up of such "continuous transducer neurons" *(fig.2)*. Note that the logical depth of this system is no greater than two; the first step is from the sensory system to the association system, and the second is from the association to response system. The association system is composed of association cells (A-units) which are continuous transducer neurons. The inputs of these A-units come from origin points (which may be randomly distributed) in the sensory system, or "retina", and their outputs go to

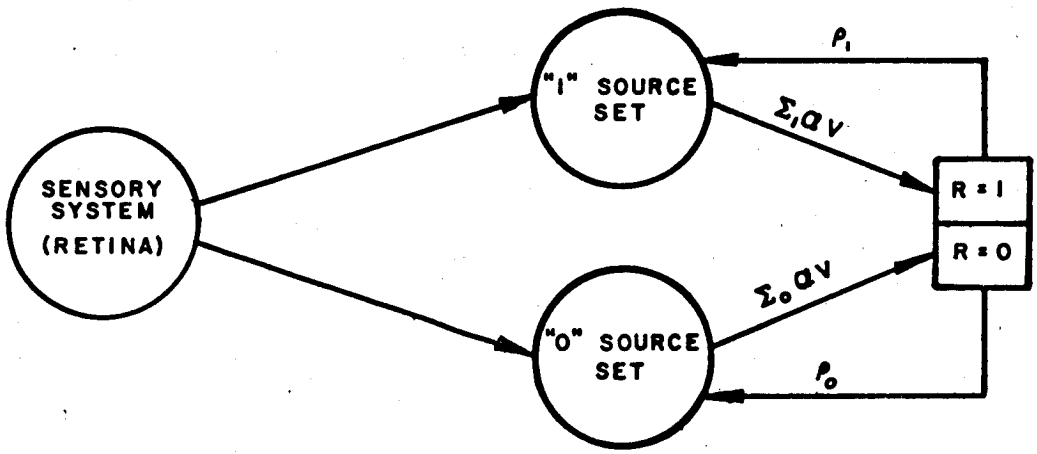


Fig.2. Organization of a Simple Perceptron

the single response unit which is shown in this system. (The organization of systems with large numbers of response units has been considered in previous work *(ref.3)*. Those A-units which tend to turn the response "on" (to the condition $R = 1$) are collectively designated the "1 source-set", while those A-units which tend to turn off, or inhibit, the response, are designated the "0 source-set". If $R = 1$, a reinforcement signal, ρ_1, is transmitted to all members of the 1 source-set. If $R = 0$, a reinforcement signal, ρ_0, is transmitted to all members of the 0 source-set. When a stimulus is presented to the visual system, the output signals from each of the source-sets are summed, and the sign of the difference, $\Sigma_1 \alpha v - \Sigma_0 \alpha v$, determines whether the response is 1 or 0. In other words, if the 1 source-set delivers a stronger signal, the response will be "1"; if the 0 source-set delivers a stronger signal, the response will be "0". Note that the

"reinforcement operator", $\Sigma\rho$, which appears in the growth rate equations (1), (2) and (3) will, in this system, be either 1 or 0, depending on the current state of the response.

Corresponding to each of the three growth rate equations, we have the three limit equations, for the value of an A-unit:

$$\nu \longrightarrow \pm\infty \tag{1a}$$

$$\nu \longrightarrow \frac{E\alpha\rho f_\rho}{\delta} \tag{2a}$$

$$\nu \longrightarrow \frac{E\alpha\rho}{\delta} \tag{3a}$$

where $E\alpha$ is the expected value of the input signal to the A-unit in question, and f_ρ is the mean frequency with which the A-unit is reinforced (i.e., the frequency with which $\Sigma\rho = 1$). The first and third of these cases will be the ones with which we are primarily concerned. It can be shown that the second case (which is more plausible for biological units, since the rate of decay does not depend on the feedback signal, $\Sigma\rho$) will merely lead to a somewhat weaker form of the system represented by the third alternative.

5. SOME IMPORTANT CORRELATION COEFFICIENTS

In the analysis of fixed threshold perceptrons *(ref. 3)*, it was shown that the ability of the perceptron to discriminate between similar forms depends on the proportion of A-units activated in common by each of the forms to be distinguished. The expected value of this proportion is called P_c. In dealing with classes of forms, it was found possible to measure the "similarity" of the classes in terms of P_c. In the case of continuous transducer perceptrons, however, where every A-unit may be "active", albeit at widely scattered frequencies, measurements based on probabilities of activation are no longer suitable. In place of P_c, we will make use of several product moment correlation coefficients to measure the similarity of the activity induced by different classes of stimuli. A class of stimuli, in this context, means any set of forms, or retinal projections, which we will regard, arbitrarily, as the same "kind" of visual object, e.g., letters of the alphabet, geometrical shapes such as squares and triangles, dogs and cats, etc. We will be concerned, particularly, with the binary discrimination of stimuli into two classes, which we will designate S_1 and S_2.

Two kinds of correlations are of particular importance in the analysis of similarity, for a continuous transducer perceptron. These correlations can be written, in abbreviated form, as $r_{\alpha\bar{\alpha}}$ and $r_{\bar{\alpha}\bar{\alpha}}$. $r_{\alpha_1\bar{\alpha}_1}$ is the correlation, measured over the set of all A-units, of the input signals, α, from some particular stimulus of class S_1, with the expected value of the input

signals from the class S_1. In other words, $r_{\alpha_1\bar{\alpha}_1}$ (which is roughly analogous to $P_{c_{11}}$, in the old system of notation) measures the similarity of any arbitrary stimulus of a class to the remainder of the class; it is a measure of the *coherence* of the stimulus classes. Note that this measure of similarity is not based on the stimuli themselves, but on the signals received by the association system. $r_{\alpha\bar{\alpha}}$, the second of the two basic correlations, is the correlation of the expected values of α for the two classes of stimuli, S_1 and S_2*. Both correlations are measured over the set of all admissible A-units, i.e., all units which meet the parametric constraints of the system.

It should be emphasized that the kinds of similarity relationships which emerge in a system of this sort will be a function of the organization of the connection system, by which signals are transmitted from sensory points to A-units. In our original studies of the perceptron, it was assumed that the origin points for the input fibers to an A-unit were randomly scattered throughout the retina, or sensory mosaic. This system, as might be expected, leads to a high degree of sensitivity to the *location* of a stimulus in the visual field, and tends to create a bias such that the influence of large stimuli outweighs the effect of small stimuli, in the development of learning and memory. An alternative system, now being investigated, calls for a polarization of the origin points for an A-unit about an arbitrarily selected line in the retinal field, such that all connections originating on one side of the line are excitatory, while all connections originating on the other side are inhibitory. The density of connections, in this system, increases in the neighborhood of the line, or "axis of polarization". Such a system is primarily sensitive to the location and direction of contours, rather than to illuminated areas per se, and the resulting measures of similarity, using the above correlation coefficients, will naturally be quite different. In order to minimize sensitivity, to the *location* of a figure, it is generally desirable to define α in such a way that negative values are eliminated. This is equivalent to setting a zero threshold for the A-unit; otherwise the expected value gain, following the translation of a figure over the field, will always be zero, if the numbers of excitatory and inhibitory connections are equal.

Despite their relativity with respect to the connection system, these "similarity correlations" ($r_{\alpha\bar{\alpha}}$ and $r_{\bar{\alpha}\alpha}$) appear to bear a definite relationship to our intuitive, phenomenological concept of "similarity". Particularly in the case of the contour-sensitive connection system mentioned above, it can be shown that classes of forms which we tend to regard as "similar" (e.g., different sizes and locations in the field of the same letter of the alphabet) will have a high value of $r_{\alpha\bar{\alpha}}$, while for dissimilar, or randomly selected forms, $r_{\alpha\bar{\alpha}}$ will be low. Similarly, if we consider $r_{\bar{\alpha}\alpha}$

---

* $r_{\bar{a}a}$ is used as an abbreviated notation for $r_{\bar{a}_1\bar{a}_2}$

between two strongly dissimilar classes of forms (such as horizontal and vertical bars, or the letters E and X) we will generally obtain a lower value than between similar classes (such as lower case "x" and upper case "X").

Now let us consider how these correlations come to be represented in the perceptron.

If any association cell is exposed exclusively to members of a single stimulus class, or if the reinforcement operator, ρ, is equal to 1 for all members of class S_1 and 0 for all members of class S_2, then the value of this cell will grow in a direction and at a rate which is determined by $\bar{\alpha}_1$, the expected value of the input signal from S_1 stimuli. Over a set of such units, therefore, as time goes to infinity, we would expect

$$r_{\alpha\nu} \longrightarrow r_{\alpha\bar{\alpha}} \tag{4}$$

where $r_{\alpha\nu}$ is the correlation of the input signals from any one stimulus (selected at random from the class in question) with the value currently stored in the A-units.

Now, if we "force" the responses of the system, so that, for example, in an environment of triangles and circles, the perceptron is made to give the response $R = 1$ consistently for all triangles, and $R = 0$ consistently for all squares, then in the 1 source-set we get

$$\begin{cases} r_{\alpha_1\nu_1} \longrightarrow r_{\alpha_1\bar{\alpha}_1} \\ r_{\alpha_2\nu_1} \longrightarrow r_{\alpha_2\bar{\alpha}_1} \end{cases} \qquad (Er_{\alpha_2\bar{\alpha}_1} < Er_{\alpha_1\bar{\alpha}_1}) \tag{5a}$$

and in the 0 source-set

$$\begin{cases} r_{\alpha_1\nu_0} \longrightarrow r_{\alpha_1\bar{\alpha}_2} \\ r_{\alpha_2\nu_0} \longrightarrow r_{\alpha_2\bar{\alpha}_2} \end{cases} \qquad (\mathrm{Er}_{\alpha_1\bar{\alpha}_2} < \mathrm{Er}_{\alpha_2\bar{\alpha}_2}) \tag{5b}$$

In other words, the values of the A-units in each source-set become correlated with the expected values of the input signals from the class of stimuli which has been "associated" to the corresponding response, and the correlation with signals from stimuli of the opposite class are expected to be smaller by a factor which varies with $r_{\bar{\alpha}\bar{\alpha}}$. Thus, if the stimulus classes are completely independent ($r_{\bar{\alpha}\bar{\alpha}} = 0$) we should expect (for the "1" source-set)

$$\begin{cases} r_{\alpha_i\nu_i} \longrightarrow r_{\alpha_i\bar{\alpha}_i} \\ r_{\alpha_j\nu_i} \longrightarrow 0 \end{cases}$$

Now consider the signals which are transmitted from the association system to the *R*-unit *(fig. 2)*. These signals are equal to the sum of $\alpha\nu$ for the source-set in question. But the correlation coefficient, $r_{\alpha\nu}$, for either source-set, has the formula:

$$r_{\alpha\nu} = \frac{\Sigma(\alpha - \bar{\alpha})(\nu - \bar{\nu})}{N\sigma_\alpha\sigma_\nu} = \frac{\Sigma\alpha\nu + [N\bar{\alpha}\bar{\nu} - \bar{\nu}\,\Sigma\alpha - \bar{\alpha}\Sigma\nu]}{N\sigma_\alpha\sigma_\nu} \tag{6}$$

where N = the number of A-units in the source-set.
If we assume that the two classes of stimuli, S_1 and S_2, produce the same distribution of signal amplitudes at the A-units, then the expected values of σ_α and $\bar{\alpha}$ will be identical for the 1-source-set and the 0-source-set. If, in addition, the same number of stimuli have been associated to each of the two responses, the variable ν will also have the same expected distribution in the two source-sets (under the experimental conditions for "forced learning", described above) so that σ_ν and $\bar{\nu}$ can be considered equal for the two cases. N is also assumed to be equal for the two source-sets. Thus, the term in brackets in the numerator of (6) reduces to a constant, which will be zero if $\bar{\alpha}$ and $\bar{\nu} = 0$. The denominator also becomes a constant, so that the correlation can be expressed in the form:

$$r_{\alpha\nu} = \frac{\Sigma\alpha\nu}{C_1} + C_2 \tag{7}$$

Consequently, the expected difference of the signals from the two source-sets, $\Sigma_1\alpha\nu - \Sigma_0\alpha\nu$, will be directly proportional to the difference of the correlations $r_{\alpha_i\nu_1}$ and $r_{\alpha_i\nu_0}$. We have already seen [in equations (5a) and (5b)] that one of these correlations will be proportional, in the limit, to $r_{\alpha_i\bar{\alpha}_i}$ and the other to $r_{\alpha_i\bar{\alpha}_j}$. The difference will be positive if the subscript i refers to the first class of stimuli, negative if i refers to the second class.

Under these conditions, therefore, we would expect that the presentation of any arbitrary stimulus from class S_1 will tend to evoke the response which has been formerly associated to S_1 stimuli, while the presentation of an S_2 stimulus should evoke the opposite response. This, actually, is the essence of the theory of "forced learning" in a continuous transducer perceptron. Predictions of the reliability of the response, as a function of the number of A-units in the system, can be made by making use of standard techniques for determining the probability of an error in the difference of two correlation coefficients, equal to or greater than the expected difference, $r_{\alpha_i\bar{\alpha}_i} - r_{\alpha_i\bar{\alpha}_j}$.

Simulation experiments, using the IBM 704 digital computer at the Cornell Aeronautical Laboratory, have substantiated the predictions of this theory. It has been possible to teach a simulated perceptron to recognize

the difference between geometrical forms, letters of the alphabet, and positions on the retina, using continuous transducer neurons. The most important question, however, still remained unanswered at this point in the program: What would such a perceptron do if, instead of forcing the desired response during a "training period", we simply turned it loose in an arbitrary environment? Would it ever, spontaneously, arrive at the desired concepts?

6. THE THEORY OF CLASS C PERCEPTRONS

The problem of spontaneous organization was originally investigated with respect to perceptrons in which the growth function of the value corresponds to equation (1). This study led to the following theorem and corollary:

THEOREM 1:

There exists a class, C, of perceptrons, which tend toward a statistically stationary state such that each binary response (R = 1, 0) becomes established as either 1 or 0 universally, for all stimuli, with an error probability, ϵ, which approaches zero as $t \rightarrow \infty$. In the terminal condition, each binary response gives 0 bits of information with regard to the current stimulus.

COROLLARY:

A perceptron in which the values of the A-units are allowed to grow without bound, and in which the sensory origin points of the A-units are randomly located, is a member of the class C.

In order to prove this theorem, it is clearly sufficient to prove the corollary for any particular case. We will consider, as a perceptron of this type, a continuous transducer perceptron with a single binary response, such as the one shown in *fig. 2*, with growth function (1). As there is no decay in the values of the A-units, they can continue to grow indefinitely, as indicated by equation (1a). The environment of this system is assumed to consist of two arbitrary stimulus classes, S_1 and S_2. Stimuli are assumed to occur in a random sequence, where the probability of an S_1 stimulus is $P(S_1)$, and the probability of an S_2 stimulus is $P(S_2)$. We must show that whatever the relationship of the stimulus classes, the condition stated in the theorem will tend to occur.

In order to develop a proof for this theorem, it will be necessary, first, to define some terms. In particular, we must consider several different methods of representing and classifying the "state of the system" at time t.

The description of the "state of the system", which gives us the information which we ultimately want, takes the form of a 2 x 2 probability matrix,

$$P = \begin{pmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{pmatrix}$$

where the first subscript denotes the stimulus class (S_1 or S_2) and the second denotes the response ($R = 0$ or $R = 1$). Thus P_{21} = the probability that a Class 2 stimulus would evoke the response $R = 0$, at time t.

The matrix P can be classified into three cases, as follows:

P is *horizontally symmetric* if both elements of one column are greater than 0.5, and both elements of the other column are less than 0.5, i.e.,

$$P = \begin{pmatrix} .5 + \Delta & .5 - \Delta \\ .5 + \Delta & .5 - \Delta \end{pmatrix} \quad \text{or} \quad P = \begin{pmatrix} .5 - \Delta & .5 + \Delta \\ .5 - \Delta & .5 + \Delta \end{pmatrix}$$

P is *diagonally symmetric* if one pair of diagonal elements are greater than 0.5, and the other pair are less than 0.5, i.e.,

$$P = \begin{pmatrix} .5 + \Delta & .5 - \Delta \\ .5 - \Delta & .5 + \Delta \end{pmatrix} \quad \text{or} \quad P = \begin{pmatrix} .5 - \Delta & .5 + \Delta \\ .5 + \Delta & .5 - \Delta \end{pmatrix}$$

P is *neutral* if the elements of one or more rows are both equal to 0.5.

Theorem 1, stated in terms of these definitions, predicts that in a Class C perceptron the terminal state will always tend to be horizontally symmetric, and that Δ will approach 0.5 as $t \to \infty$.

In order to evaluate P at some arbitrary time, t, and to show the trend in its development through time, two other matrices are helpful. The first of these is the 2 x 2 matrix, N, where

$$N = \begin{pmatrix} n_{11} & n_{12} \\ n_{21} & n_{22} \end{pmatrix}$$

Here the row subscript again refers to the stimulus, and the column subscript to the response. Each n is equal to the number of times that the indicated response has occurred in response to a stimulus from the indicated class. n_{21}, for example, is the number of times that a stimulus of Class 2 has evoked the response $R = 0$, throughout the history of the system up to time t.

The remaining matrix, K, is again a 2 x 2 matrix of the same form:

$$K = \begin{pmatrix} k_{11} & k_{12} \\ k_{21} & k_{22} \end{pmatrix}$$

where $k_{ij} = r_{\alpha_i \nu_j} \sigma_{\nu_j}$. The elements of this matrix are proportional to the expected values of the output signals from the j-source-set, in response to an arbitrary stimulus of Class i. The significance of the correlations $r_{\alpha\nu}$ has already been indicated in the preceding section. For simplicity, we assume a perceptron so designed that $\bar{v} = 0$. The conditions assumed for

equation (7) all apply, except for the fact that σ_{ν} can no longer be assumed equal for the two source-sets, since the sets may not have been "reinforced" an equal number of times. The elements of K have, therefore, been corrected for the effect of σ_{ν}.

Now, it can be shown that $K(t)$ is a function dependent only on $N(t)$, and that $P(t)$ depends exclusively on $K(t)$. Moreover, we can show that $\{N(t)\}$ is a function exclusively of $\{N(t-1)\}$, $\{P(t-1)\}$, $P(S_1)$ and $P(S_2)$, where the last two probabilities have been defined as constants. Consequently, the matrix N is a Markovian process, and P is a function of this Markovian process.

Our general procedure will consist of showing that there exists a path by which any state of the system can progress to one of the two admissible terminal states,

$$P = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} \text{ or } \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}$$

It can be shown that, since the system is fundamentally Markovian, these states are "trapping", i.e., once the system has arrived in either of these states, by any path whatever, the probability of departing from that state is zero. The only other states that might be trapping are the diagonal states, $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ and $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$. The proof that the system will always, given sufficient time, tend towards one of the horizontal rather than one of the diagonal terminal states, rests on the fact that as long as t remains finite, the states will never be *absolutely* trapping, i.e., there always remains some error probability, and that this error probability tends to be greater for the diagonal than for the horizontally symmetric states.

The initial state of the system, at $t = 0$, can generally be characterized by the conditions:

$$N(0) = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

$$K(0) = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

$$P(0) = \begin{pmatrix} .5 & .5 \\ .5 & .5 \end{pmatrix}$$

or, in abbreviated notation,

$$\begin{pmatrix}0&0\\0&0\end{pmatrix}, \begin{pmatrix}0&0\\0&0\end{pmatrix}, \begin{pmatrix}.5&.5\\.5&.5\end{pmatrix}$$

where the matrices are assumed to be written in the order N, K, P. If we now treat the occurrence of a stimulus, S_1 or S_2, as an operator on the state of the system, we can indicate the development of the system over any period of time as a tree, with a fourfold branching of possibilities at each successive moment in time (assuming time, for the sake of convenience, to be quantized rather than continuous):

$$\begin{pmatrix}0&0\\0&0\end{pmatrix}, \begin{pmatrix}0&0\\0&0\end{pmatrix}, \begin{pmatrix}.5&.5\\.5&.5\end{pmatrix} \quad \begin{array}{l} \xrightarrow{P_{11}\,P(S_1)} \begin{pmatrix}1&0\\0&0\end{pmatrix} \\ \xrightarrow{P_{12}\,P(S_1)} \begin{pmatrix}0&1\\0&0\end{pmatrix} \\ \xrightarrow{P_{21}\,P(S_2)} \begin{pmatrix}0&0\\1&0\end{pmatrix} \\ \xrightarrow{P_{22}\,P(S_2)} \begin{pmatrix}0&0\\0&1\end{pmatrix} \end{array} \qquad (8)$$

To each of these new N matrices, there again corresponds a K and a P matrix, which governs the probabilities on the next split, etc. In general, $N(t) = N(t-1)$, with one of the four elements incremented by 1 for each of the four possibilities, as indicated in equation (8), with the probabilities indicated by the expressions above the arrows. Thus, the sum of all the elements in N is always equal to t, and the sum of the elements in either column is equal to the number of occurrences of the corresponding response, in the prior history of the system.

The number of terminal branches in a tree of the form indicated in (8) will be equal to 4^t, a number which grows large far too rapidly to permit exhaustive computation of the spectrum of terminal states beyond the first few generations. The probability of any given terminal state or branch is equal to the product of the branch probabilities $(P_{sr}P(S_s))$ for each step of the sequence prior to t, where S_s is the stimulus, and R_r the response which occurred. These "world lines" of the system yield the expected values of the elements in the terminal P matrices, and the total probability that the system will terminate in a horizontal or diagonal state can be obtained by summing the branch probabilities of the terminal branches of each class.

In order to prove the theorem, which is our primary concern here, we must first show how the P matrix depends on N. The first step will be the development of the elements, k_{ij}, of the K matrix.

We have said that $k_{ij} = r_{\alpha_i \nu_j} \sigma_{\nu_j}$, which we have seen to be proportional to $\Sigma_j \alpha\nu$, the expected output signal from the j-source-set, in the presence of a stimulus of Class i. If we let $\sigma^2_{\nu\alpha}$ be the component of the variance of ν which is correlated with α, then

$$r^2_{\alpha\nu} = \frac{\sigma^2_{\nu\alpha}}{\sigma_\nu{}^2}$$

and

$$k^2_{ij} = \frac{\sigma^2_{\nu_j \alpha_i}}{\sigma^2_{\nu_j}} \cdot \sigma_{\nu_j}{}^2 = \sigma_{\nu_j}{}^2_{\alpha_i} \tag{9}$$

In other words, the elements, k_{ij}, are measures of the expected covariance of α and ν, for stimulus Class i and source-set j. Now this covariance comes from the n_{1j} increments to the value from S_1 stimuli, and the n_{2j} increments from S_2 stimuli. It is convenient, for purposes of analysis, to factor the square of the required covariance in an approximate fashion as follows:

$$\sigma_{\nu_j}{}^2_{\alpha_i} \approx r^2_{\alpha_i \bar{\alpha}_i} \sigma_{\nu_j}{}^2_{\bar{\alpha}_i} \tag{10}$$

which means that we need only find the covariance of ν_j and $\bar{\alpha}_i$, the expected value of α for the stimulus class S_i. Now each increment to the value of the source-set has an expected value equal to $\Sigma\alpha$, for all A-units in the source-set. This increment has an expected variance which we can arbitrarily set equal to unity. This variance can be divided into four components, as follows:

$\Delta_1 \equiv$ Variance which is "unique" to the particular stimulus, and uncorrelated with either $\bar{\alpha}_1$ or $\bar{\alpha}_2$

$\Delta_2 \equiv$ Variance which is correlated with $\bar{\alpha}_1$ but not with $\bar{\alpha}_2$

$\Delta_3 \equiv$ Variance which is correlated with both $\bar{\alpha}_1$ and $\bar{\alpha}_2$

$\Delta_4 \equiv$ Variance which is correlated with $\bar{\alpha}_2$ but not with $\bar{\alpha}_1$

The magnitudes of these four components will be:

$$\Delta_1 = 1 - (\Delta_2 + \Delta_3 + \Delta_4)$$

$$\Delta_2 = r^2_{\alpha_i \bar{\alpha}_i} \ (1 - r^2_{\bar{\alpha}\bar{\alpha}})$$

$$\Delta_3 \equiv r^2_{\alpha_i \bar{\alpha}_i} \ r^2_{\bar{\alpha}\bar{\alpha}}$$

$$\Delta_4 = r^2_{\alpha_i \bar{\alpha}_j} (1 - r^2_{\bar{\alpha}\bar{\alpha}})$$

The first and fourth of these components clearly makes no contribution to $\sigma^2_{\nu\bar{\alpha}}$. The second two components each make a contribution, which combine in the covariance as follows:

$$\sigma^2_{\nu_j \bar{\alpha}_i} = n^2_{ij}\,\Delta_{2(i)} + \left(\sqrt{n^2_{1j}\,\Delta_{3(1)}} + \sqrt{n^2_{2j}\,\Delta_{3(2)}}\right)^2$$

$$= n^2_{ij}\, r^2_{\alpha_i \bar{\alpha}_i}\left(1 - r^2_{\bar{\alpha}\bar{\alpha}}\right) + \left[n_{1j}\, r_{\alpha_1 \bar{\alpha}_1}\, r_{\bar{\alpha}\bar{\alpha}} + n_{2j}\, r_{\alpha_2 \bar{\alpha}_2}\, r_{\bar{\alpha}\bar{\alpha}}\right] \quad (11)$$

Multiplying through by $r^2_{\bar{\alpha}\bar{\alpha}}$, as indicated in equation (10), we obtain:

$$k^2_{ij} = n^2_{ij}\, r^4_{\alpha_i \bar{\alpha}_i}\left(1 - r^2_{\bar{\alpha}\bar{\alpha}}\right) + r^2_{\alpha_i \bar{\alpha}_i}\, r^2_{\bar{\alpha}\bar{\alpha}}\left[n_{1j}\, r_{\alpha_1 \bar{\alpha}_1} + n_{2j}\, r_{\alpha_2 \bar{\alpha}_2}\right]^2 \quad (12)$$

One point which has been overlooked thus far concerns the signs of the k_{ij}. Since k_{ij} has been squared in the course of this analysis, the sign of the element, as obtained from equation (12), might be either positive or negative. Now it must be remembered that k_{ij} is intended to be proportional to the output signal of the j-source-set, $\Sigma_j \alpha\nu$, and to $r_{\alpha_i \nu_j}$. If the values, ν_j, have been built up exclusively by increments from stimuli of the opposite class from Class i, then it is quite conceivable that $r_{\alpha_i \nu_j}$ could be negative, if $r_{\bar{\alpha}\bar{\alpha}}$ were negative. $r_{\alpha_i \nu_j}$ could also be negative, for some specific stimulus of Class i, if $r_{\alpha_i \bar{\alpha}_i}$ were negative for the stimulus in question. Under all other conditions, the signs will remain positive. Now, if we limit our discussion to perceptrons with random spatial distributions of origin points for the A-unit input fibers, it can be shown that $r_{\bar{\alpha}\bar{\alpha}}$ will always be positive. If the stimulus classes, S_1 and S_2, are totally disjunct, $r_{\bar{\alpha}\bar{\alpha}}$ may be equal to zero, but it will never go negative. If, on the other hand, the distribution of origin points is organized in some special pattern, such as the polarized contour-sensitive distribution described above, it is possible to obtain negative values of $r_{\bar{\alpha}\bar{\alpha}}$. If this should be the case, then there will be certain environments in which the perceptron which has been described will not behave as a Class C system, but will begin to act as a Class C' system (to be described in the following section) instead. Let us therefore limit ourselves to perceptrons with random spatial distributions of origin points (as indicated in the Corollary), for which the elements k_{ij} will always be positive. We should also assume a "well behaved" system in which $r_{\alpha_i \bar{\alpha}_i}$ is always positive, for any stimulus of Class i.

Now the difference between the two k's of the same row, $k_{i2} - k_{i1}$, will clearly be proportional to the difference, $\Sigma_1 \alpha\nu - \Sigma_0 \alpha\nu$, between the signals from the two source-sets, which determines the response. The probability that the response $R = 1$ will occur is a positive monotonic function of this difference. It is also clear from equation (12) that if $r_{\bar{\alpha}\bar{\alpha}} \neq 0$, the magnitude of the k_{ij} will increase monotonically with $n_{1j} + n_{2j}$, i.e., with the column sums of the N matrix. Consequently, if the ratio of the sum of the i-column to the sum of the j-column should increase without bound, it is

clear that p_{1i} and p_{2i} will tend towards unity, while p_{1j} and p_{2j} tend towards zero, which is the condition of horizontal symmetry predicted by the theorem. It is also clear that, if the elements of one row of the N matrix are equal, then an inequality between the elements of the remaining row will cause the P matrix to be horizontally symmetric.

Let us now consider what happens as soon as a stimulus of either class is presented to the perceptron [equation (8)]. It is clear that the first increment will immediately introduce a bias towards either the left column or the right column, so that equation (8), in more complete form, can be written:

$$\begin{pmatrix}0&0\\0&0\end{pmatrix},\ \begin{pmatrix}0&0\\0&0\end{pmatrix},\ \begin{pmatrix}.5&.5\\.5&.5\end{pmatrix} \longrightarrow \begin{cases} .5P(S_1) \longrightarrow \begin{pmatrix}1&0\\0&0\end{pmatrix}, \begin{pmatrix}k&0\\k&0\end{pmatrix}, \begin{pmatrix}.5+\Delta&.5-\Delta\\.5+\Delta&.5-\Delta\end{pmatrix} \\ .5P(S_1) \longrightarrow \begin{pmatrix}0&1\\0&0\end{pmatrix}, \begin{pmatrix}0&k\\0&k\end{pmatrix}, \begin{pmatrix}.5-\Delta&.5+\Delta\\.5-\Delta&.5+\Delta\end{pmatrix} \\ .5P(S_2) \longrightarrow \begin{pmatrix}0&0\\1&0\end{pmatrix}, \begin{pmatrix}k&0\\k&0\end{pmatrix}, \begin{pmatrix}.5+\Delta&.5-\Delta\\.5+\Delta&.5-\Delta\end{pmatrix} \\ .5P(S_2) \longrightarrow \begin{pmatrix}0&0\\0&1\end{pmatrix}, \begin{pmatrix}0&k\\0&k\end{pmatrix}, \begin{pmatrix}.5-\Delta&.5+\Delta\\.5-\Delta&.5+\Delta\end{pmatrix} \end{cases} \tag{13}$$

Each of the resultant first generation matrices is of horizontally symmetric form. (In the event that $r_{\overline{\alpha}\overline{\alpha}} = 0$, the first generation matrices would be neutral, but this can be regarded as a limiting case for the Class C perceptron, in which horizontally and diagonally symmetric states remain equally probable, no matter how far t is extended.) It is now clear that, *whichever stimulus is the next to occur*, the probability is greater than 0.5 that the same response will occur which occurred originally; and if this does happen, then it is evident that the same column of the N-matrix which was incremented before will be incremented again, leading to an increase in Δ, and strengthening the tendency towards the horizontal form. Thus, we find that a matrix which is already a horizontally symmetric case tends to remain so, and that the application of further stimuli will always tend to increase the existing tendency. In the limit, as Δ approaches 0.5, it is clear that the probability of ever gaining an increment in the "weak" column of the N-matrix goes to zero, and consequently, the state can be considered strongly trapping.

We must still consider, however, the eventuality that, before becoming "trapped" in a horizontal state, sufficient responses should occur in each column so that the system goes into a diagonally symmetric state. Consider, for example, the second generation matrix $N = \begin{pmatrix}1&0\\0&1\end{pmatrix}$, for which we would have:

$$P = \begin{pmatrix}.5+\Delta&.5-\Delta\\.5-\Delta&.5+\Delta\end{pmatrix}$$

Note, however, that the K-matrix, in this case, takes the form $\begin{pmatrix} k & k \\ k & k \end{pmatrix}$, with a nonzero element in each cell, whereas in a second generation horizontal case, $K = \begin{pmatrix} k & 0 \\ k & 0 \end{pmatrix}$. Moreover, it is clear from equation (12) that each k_{ij} in the diagonal case will be less than the dominant k's in the horizontal case, since the magnitude of the k's tends to increase with the square of the column sums of the N matrix. Consequently, the row differences, $k_{i2} - k_{i1}$, which determine the p_{ij}, are bound to be greater for the horizontally symmetric cases than for the diagonally symmetric cases; in other words, the horizontal cases are "stronger", or more binding, than the diagonal cases, and the probability of an "atypical" response (one which goes counter to the current tendency of the P matrix) is greater for the diagonal than for the parallel case. This argument can clearly be extended for any generation matrix. Consequently, no matter how far the matrices progress towards a perfectly trapping condition, the horizontal cases will always be more strongly trapping than the diagonal cases.

It remains only to be shown that any diagonal state of the system can be transformed into a horizontally symmetric state, in order to prove the theorem. Consider the arbitrary diagonal state, $N = \begin{pmatrix} n' & n \\ n & n' \end{pmatrix}$, where the primed elements indicate large integers, and the unprimed letters indicate smaller integers. To each of the large elements corresponds some $p_{ij} < 1$, and to each of the small elements corresponds some $p_{ij} > 0$. Consequently, on the subsequent cycle, $(t + 1)$, there is a nonzero probability that $n \rightarrow n + 1$. This yields a new matrix, which is subject to the same argument. It is therefore clear that after $n' - n$ such events, either of the small elements (n) might grow to the magnitude of the large elements (n'), which is a sufficient condition for a horizontally symmetric state (since the n's of one row are now equal, and the n's of the other row unequal).

This completes the proof for Theorem 1. The behavior of such a Class C system can be most clearly represented by an equilibrium diagram *(fig.3)*, in which heavy arrows indicate strong tendencies, and light arrows indicate weak tendencies. In general, as time goes to infinity, all systems will eventually assume a horizontally symmetric form, with continually decreasing probability of escaping to either a neutral or a diagonal condition.

7. THE CLASS C' THEOREM

Theorem 1 indicates that the type of perceptron which had been considered at the outset would be incapable of "spontaneous organization". In fact, such a system, even if it has been deliberately taught to associate opposite responses to two classes of stimuli, is likely to degenerate to the terminal condition described by the theorem, if it is subsequently left on its own,

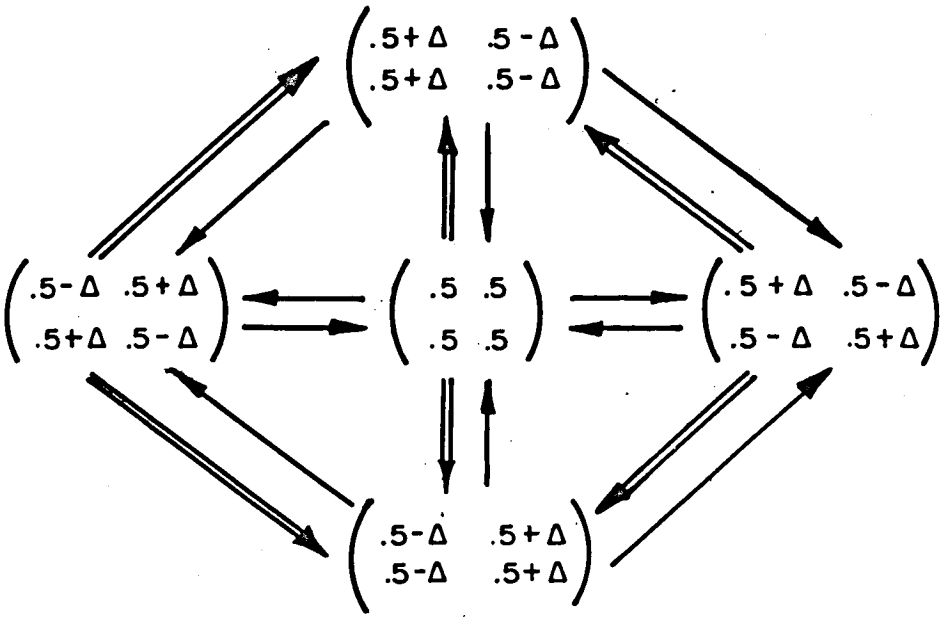


Fig.3. Equilibrium diagram for a Class C System
(States of the System are represented by P Matrices)

without further training or supervision by a human operator. This discovery led to a search for a system which would not be subject to the Class C effect, and which would tend, instead, to arrive at a "useful" division of its environment, without human intervention. The following existence theorem states the conditions with which we are specifically concerned:

THEOREM 2

There exists a class, C', of perceptrons, which always tend toward a statistically stationary state, such that for each binary response (R = 1, 0) the environment will be dichotomized into the two stimulus classes, S_1 (whose members evoke the response R = 1) and S_2 (whose members evoke the response R = 0). The dichotomy which is established will, in general, be characterized by a high degree of similarity between stimuli of the same class, and marked dissimilarity between stimuli of opposite classes. In the limit, each response of such a system contains 1 bit of information, with respect to the current stimulus.

Actually, there are many systems which meet the conditions of this theorem. The simplest of these, mathematically, is one in which the output signals of each source-set of A-units are divided by the current value of σ_ν, thus making the elements of the K-matrix (defined above) proportional to $r_{\alpha\nu}$. A system of much greater practical interest, however (which, moreover, converges to its terminal condition more rapidly), results from the growth functions defined by equations (2) and (3). As indicated in (2a) and (3a), the values of the A-units tend towards a limit which is proportional to the expected value of the input signal times the associated magnitude of the reinforcement operator, which is assumed here to be 1 or 0. Hence, we have the following corollary:

COROLLARY:

A perceptron in which the values of the A-units decay at a rate proportional to their current magnitude is a member of the Class C'.

With the proof of the Class C theorem as a model, we are now prepared to undertake the rather more subtle proof of the Class C' theorem. As before, it will be sufficient to prove the corollary, for any particular system, in order to prove the theorem. Actually, this proof involves two distinct points:

1. We will prove that the system always tends to dichotomize the environment into *some* two classes. To prove this, it is sufficient to show that, given *any* two stimulus classes, S_1 and S_2, the system will tend towards a diagonally symmetric terminal state in preference to a horizontally symmetric terminal state.
2. We must then prove that the dichotomy which is most strongly preferred by the system, in any environment, will correspond to our concept of "good similarity and dissimilarity", i.e., that it will tend toward a dichotomy which maximizes $r_{\alpha\bar{\alpha}}$ and minimizes $r_{\bar{\alpha}\bar{\alpha}}$.

The analysis will be carried out for a perceptron having the same organization as before, and characterized by the growth function stated in equation (3).

As in the case of the Class C perceptron, the state of the system will be described in terms of three matrices. The matrix P has precisely the same meaning as before, and can be classified, in the same manner, into horizontally and diagonally symmetric cases, and neutral cases. If we try to use the N matrix as a basis for calculating the state of the system, however, we will run into serious difficulties, for N is no longer Markovian. In the Class C perceptron, it makes no difference in what sequence the contributions to the elements n_{ij} occur; a given number of reinforcements in a specified category will always have the identical effect, regardless of the time at which they occur. In the Class C' system, however, this is no longer the case. The elements of the N matrix no longer unambiguously determine the P matrix, and consequently, the probability of each subsequent N matrix can

no longer be predicted as before. Fortunately, we can replace N with an alternative matrix, which is Markovian, and which we will call V. The V matrix is a 2 x 2 matrix with elements v_{ij}, where, as before, the row corresponds to the stimulus class, and the column to the response. The element v_{ij} can be interpreted as a measure of the residual value, at time t, which is due to the past reinforcements of the j source-set by stimuli of Class i. In accordance with equation (3), the elements of V will decay by a fraction δ, whenever the source-set j is reinforced (i.e., whenever the response $R = R_j$). The column sums of V correspond to the total values of the elements in the j source-set at time t due to all previous stimuli. It is clear that if the proportional decay, δ, is applied to each component of the sum individually (i.e., to the elements v_{1j} and v_{2j}) the net effect will be the same as if the decay had been applied to the column sums at each step, so that the analysis of the source-set value into two components does not introduce any error in the stochastic process. Specifically, the elements of each successive V matrix can be obtained by the rule:

$$v_{ij}(t) = \begin{cases} v_{ij}(t-1) & \text{if } R(t) \neq R_j \\ (1-\delta)v_{ij}(t-1) + 1 & \text{if } R(t) = R_j \end{cases} \qquad (14)$$

Now we have indicated that the terminal values of the A-units in a Class C' perceptron will be proportional to the expected values of α, measured over the set of stimuli to which the A-units in question have been exposed. The variance of the value, σ_v^2, will therefore tend towards a statistically stable magnitude. If we assume that the initial value distribution over the set of A-units has the same variance as this terminal distribution, then we would actually not expect the variance to change at all, as a result of reinforcement of the A-units; in effect, we will be changing the *correlations* of the A-unit values without changing the moments of the distribution. Consequently, the covariance matrix, K, which was used in the preceding analysis, can be replaced with the correlation matrix R, which consists of the coefficients $r_{ij} \equiv r_{\alpha_i v_j}$. These correlation coefficients are clearly equal to the covariance which can now be obtained in terms of the V matrix just as it was before from the N matrix) divided by σ_v. We now require an explicit expression for σ_v. Specifically, the variance σ_v^2 can be divided into five components as follows:

$\Delta_0 \equiv$ residual of initial noise component, present at time $t = 0$.
$\Delta_1 \equiv$ variance correlated with $\bar{\alpha}_1$ but not with $\bar{\alpha}_2$.
$\Delta_2 \equiv$ variance correlated with $\bar{\alpha}_2$ but not with $\bar{\alpha}_1$.
$\Delta_3 \equiv$ variance which is correlated with both $\bar{\alpha}_1$ and $\bar{\alpha}_2$.
$\Delta_4 \equiv$ variance which is "unique" to the individual stimulus (corresponding to Δ_1 in the previous analysis)

The magnitude of these components is given by the expressions:

$$\Delta_0 = \frac{\bar{\nu}_0^2}{\bar{\alpha}^2}(1-\delta)^{N_j}$$

$$\Delta_1 = \nu_{1j}^2\, r_{\alpha_1\bar{\alpha}_1}^2 (1 - r_{\bar{\alpha}\bar{\alpha}}^2) + \nu_{2j}^2\, r_{\alpha_1\bar{\alpha}_2}^2 (1 - r_{\bar{\alpha}\bar{\alpha}}^2)$$

$$\Delta_2 = \nu_{2j}^2\, r_{\alpha_2\bar{\alpha}_2}^2 (1 - r_{\bar{\alpha}\bar{\alpha}}^2) + \nu_{1j}^2\, r_{\alpha_2\bar{\alpha}_1}^2 (1 - r_{\bar{\alpha}\bar{\alpha}}^2)$$

$$\Delta_3 \equiv r_{\bar{\alpha}\bar{\alpha}}^2 \left[\nu_{1j}\, r_{\alpha_1\bar{\alpha}_1} + \nu_{2j}\, r_{\alpha_2\bar{\alpha}_2}\right]^2$$

$$\Delta_4 = \nu_{1j}\,(1 - r^2_{\alpha_1\bar{\alpha}_1}) + \nu_{2j}\,(1 - r^2_{\alpha_2\bar{\alpha}_2})$$

where $\bar{\nu}_0$ is the magnitude of the initial mean value, at $t = 0$, and N_j is the number of times that the response $R = R_j$ has occurred. From these components, we can readily obtain the approximation:

$$r_{\alpha_i\,\nu_j} \approx r_{\nu_j\bar{\alpha}_i}\, r_{\alpha_i\bar{\alpha}_i} = r_{\alpha_i\bar{\alpha}_i}\sqrt{\frac{\Delta_i + \Delta_3}{\Sigma\Delta}} \tag{15}$$

In terms of these r's, it is easy to approximate the elements of the P matrix, through the use of Fisher's z-transformation:

$$\begin{cases} p_{i1} = \phi\left(\dfrac{z_{i1} - z_{i2}}{\sqrt{\dfrac{2}{N_{A_r} - 3}}}\right) \\ p_{i2} = 1 - p_{i1} \end{cases}$$

where $$z_{ij} = \frac{1}{2}\ln(1 + r_{\alpha_i\nu_j}) - \frac{1}{2}\ln(1 - r_{\alpha_i\nu_j})$$

$\phi(x)$ = normal probability integral from $-\infty$ to x

and N_{A_r} = number of A-units in a source-set.

We must now define the concept of a "saturated state" for a C' system. Such a system will be called "saturated" when the sum of the squares for each column of the R matrix is maximum, i.e., when

$$\left(\frac{r_{\alpha_1\nu_j}}{r_{\alpha_1\bar{\alpha}_1}}\right)^2 + \left(\frac{r_{\alpha_2\nu_j}}{r_{\alpha_2\bar{\alpha}_2}}\right)^2 = 1 + r_{\bar{\alpha}\bar{\alpha}}^2 \tag{17}$$

This expression represents the upper limit for the correlations, as $r_{\nu\bar{\alpha}} \to 1$ for the two elements of the column, taken jointly.

Now, it is clear that every increment to the values of the A-units will bring the system closer to this saturated condition. Moreover, *the degree of saturation can only increase;* once a system has become saturated, it can never become unsaturated, since whichever stimulus occurs, and whatever the sequence of events, every new increment will only increase the tendency of $r_{\nu\bar{\alpha}}$ to approach unity. If we now examine the saturated state, it is clear that in the limit a horizontally symmetric system will become neutral as the two columns saturate, while a diagonally symmetric system will tend to remain diagonally symmetric. Thus, the only acceptable terminal conditions, for a Class C' system, are either neutral or diagonally symmetric. In the neutral terminal state, each of the r_{ij} elements is equal to $r_{\alpha\bar{\alpha}}\sqrt{\frac{1+3r^2_{\bar{\alpha}\alpha}}{2+2r^2_{\bar{\alpha}\alpha}}}$, while in the diagonal terminal state, the strong elements go to $r_{\alpha_i\bar{\alpha}_i}$ and the weak elements go to $r_{\alpha_j\bar{\alpha}_i}$. Consequently, the diagonal terminal states tend to be "trapping", and the neutral states have a low probability. A Class C system can now be recognized as a limiting case of the Class C' system, in which the time for saturation goes to infinity, due to the fact that σ_ν never approaches a limit.

This completes the first part of our proof, i.e., we have shown that, given any two classes, S_1 and S_2, the perceptron will prefer a diagonal terminal state to a horizontally symmetric state. The equilibrium diagram which characterizes this type of system is shown in *fig.4*. It is now necessary to show that the particular dichotomy which is formed in an arbitrary environment will tend to be one which corresponds to our criterion of similarity. In order to do this it will be sufficient to show that the net terminal error probability (considering as an "error" any response which is contrary to the bias of the terminal P matrix) will be minimum for a dichotomy which meets our similarity criteria. If the system terminates in a state in which the error probabilities are not minimum, then it will always tend (given infinite time) towards a state in which the error probabilities are still smaller, i.e., it will tend towards a more "strongly trapping" terminal condition. Let us, therefore, examine the various possible dichotomies, in an arbitrary environment, from the standpoint of their error probabilities.

Consider, first, an environment consisting only of two stimuli (not two *classes* of stimuli), S_1 and S_2. Under these conditions, $r_{\alpha\bar{\alpha}}$ for each stimulus will be equal to 1, and $r_{\bar{\alpha}\alpha}$ is equal to $r_{\alpha_1\alpha_2}$. Here there is only one dichotomy possible, so there is no doubt about which will be selected. As long as the two stimuli do not produce identical signals, (i.e., as long as $r_{\alpha_1\alpha_2} \neq 1$), the perceptron should tend to give a 1-response for one stimulus,

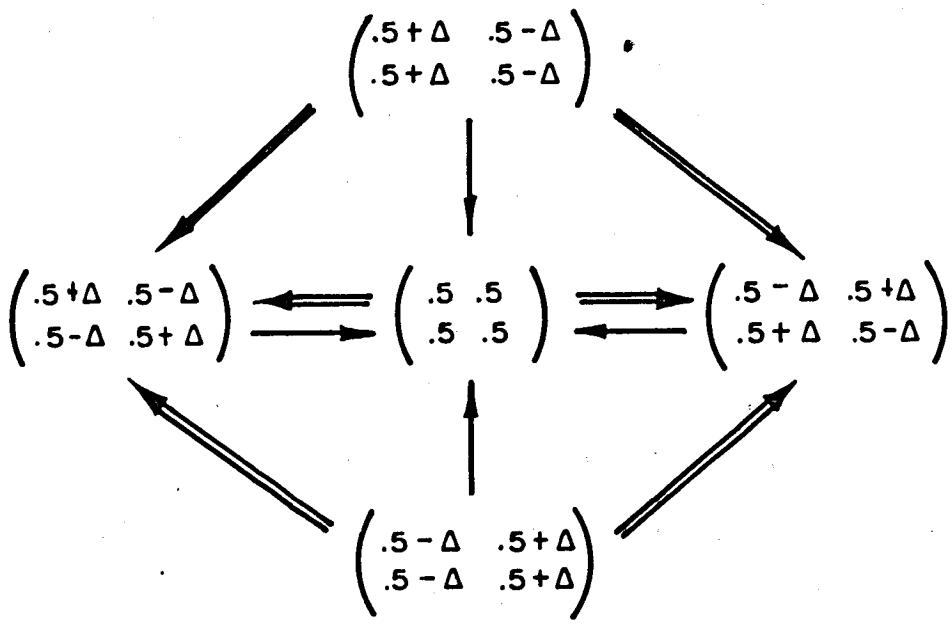


Fig.4. Equilibrium diagram for a Class C' System

and a O-response for the other. Now, if there are more than two stimuli, we can represent the state of the system unambiguously, without presupposing any particular division of the environment into classes, by adding a row to the V, R, and P matrices for each additional stimulus. Thus, for n possible stimuli, the P matrix becomes:

$$\begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \\ \vdots & \vdots \\ p_{n1} & p_{n2} \end{pmatrix}$$

This matrix can be classified as horizontally symmetric if all elements of one column are less than 0.5, and all elements of the other column are greater than 0.5. If for any row, $p_{i1} > p_{i2}$, and for any other row, $\rho_{i1} < \rho_{i2}$, the matrix is diagonal. All of the development which previously applied to the calculation of the P matrix from V and R can similarly be

restated in terms of this expanded matrix. Equation (16) still applies to the calculation of the p's, equation (14) to the ν's, and a suitable extension of equation (15) (in which the variance contributions from each element in the appropriate column of the V matrix are included) permits us to calculate the r's.

Now, the total error probability for such a system at time t will be

$$P_\epsilon = \sum_i g_i P(S_i) \tag{18}$$

where g_i is the "weak" probability (i.e., the probability which is less than 0.5) from the i row of the P matrix, and $P(S_1)$ is the probability that the particular stimulus, S_i, will occur. It is clear from the general dynamics of the Class C' systems, as discussed above, that the system will tend to keep changing until it arrives in a maximally stable terminal condition, and this maximally stable condition will be one in which P_ϵ, as defined by equation (18), is minimized.

Now consider any class, S, of stimuli. The probability, $P(S)$, that some member of this class will occur, is equal to the sum of the probabilities, $P(S_i)$, that the individual members of that class will occur. Consequently, if the system is in state P at time t, and if we define the two stimulus *classes*, S_1 and S_2, so that S_1 is the class of all stimuli for which the response $R = 1$ is preferred at time t, and S_2 is the class of all stimuli for which $R = 0$ is the preferred response, we can write [from equation (16)]:

$$P_\epsilon(t) = g_1 P(S_1) + g_2 P(S_2) \tag{19}$$

But we can write explicit expressions for g_1 and g_2 (the weak diagonal elements of a terminal P matrix) in terms of equation (16), i.e.,

$$g_i = 1 - \phi\left(\frac{Z(r_{\alpha_i\bar{\alpha}_i}) - Z(r_{\alpha_j\bar{\alpha}_i})}{\sqrt{2/N_{Ar} - 3}}\right)$$

where the z's are defined as for equation (16). Consequently, we have

$$P_\epsilon = \sum_{i=1,2}\left[1 - \phi\left(\frac{Z(r_{\alpha_i\bar{\alpha}_i}) - Z(r_{\alpha_j\bar{\alpha}_i})}{\sqrt{2/N_{Ar} - 3}}\right)\right] - P(S_i) \tag{20}$$

which is the function which must be minimized in order to find the most stable classification system for the environment in question. An examination of this expression shows the following:

1. The terminal condition depends on the frequency of stimuli in the two classes; a highly coherent class which is very small, or the members of which are all highly unlikely to occur, will be less likely to occur (other things being equal) than a larger class, or a class which includes more frequent stimuli. This means that the system will tend to favor dichotomies which divide the environment evenly between classes.

2. Other things being equal, the system will favor dichotomies such that the intraclass correlations, $r_{\alpha\bar{\alpha}}$, are large, and interclass correlations, $r_{\overline{\alpha\alpha}}$, are small. This is the similarity condition which we set out to establish.

Since we have shown that the error function, [equation (20)], will be minimum under the conditions which best satisfy our similarity criteria, we have effectively proven Theorem 2. The predictions of this theorem, as well as those of Theorem 1, have been successfully demonstrated using the IBM 704 computer to simulate the performance of a perceptron in a simple perceptual environment. In the first experiment to be successfully completed, a 500 A-unit perceptron spontaneously learned to distinguish the class of squares on the left from the class of squares on the right, in an environment in which squares were allowed to appear in random positions anywhere in the right or left halves of a visual field, being excluded only from positions in which they would overlap the center line. After being exposed to 100 squares chosen at random from this environment, the perceptron exhibited the P matrix (based on 100 test-stimuli of each class):

$$P = \begin{pmatrix} 0 & 1.00 \\ .94 & .06 \end{pmatrix}$$

in which the net error probability is only 0.03.

8. SIGNIFICANCE OF THE CLASS C' PERCEPTRON

In discussing the Class C' system, I have taken pains to avoid any reference to the concept of *entropy*. A number of theorists have felt that a system which is self-organizing, in the sense that the Class C' perceptron appears to be, is in contradiction to the second law of thermodynamics. Actually, of course, if we concern ourselves simply with the physical state of the system, entropy is clearly increased, as the perceptron is an energy consuming device. From the standpoint of *information*, on the other hand, we may still ask the question whether the total amount of information has somehow been increased as a result of the perceptron's organizing process.

Consider an environment of eight stimuli, $S_1 \ldots\ldots S_8$. At the outset, the response $R = 1$ might denote the presence of any of the stimuli S_1, S_3, S_5, or S_7, while in its terminal condition, $R = 1$ might indicate the presence of S_1, S_2, S_3, or S_4. Can we legitimately say that the information given by the response $R = 1$ in the terminal state is greater than the information given in the initial state? In any absolute sense, probably not. On the other hand, if we recognize that the first four stimuli are all triangles, while $S_5 \ldots S_8$ are all circles, we can legitimately say that the information carried by the terminal response is more meaningful than the information carried by the initial response. "Meaningful", in this sense, means that, while $R = 1$ conveys one bit of information in either case, the information

in the terminal condition has become correlated with certain aspects of the environment which we wish to have reported, whereas in the original state, the information conveyed by $R = 1$ was of no interest or utility.

This argument is, admittedly, a sketchy one, and may contain loopholes. In any case, the fact that such questions arise (and the difficulty in answering them convincingly), seems to indicate a certain ambiguity in our concept of "information". The fact that "information" is not equivalent to "meaning" has often been noted before. The performance of the Class C' perceptron seems to make this distinction even more apparent.

Be that as it may, it seems clear that the Class C' perceptron introduces a new kind of information-processing automaton: for the first time, we have a machine which is capable of having original ideas. As an analog of the biological brain, the perceptron, or, more precisely, the theory of statistical separability, seems to come closer to meeting the requirements of a functional explanation of the nervous system than any system previously proposed. The neuroeconomy, and the number of necessary constraints required in order to specify a system of this type, seems to be well within the bounds of biological plausibility. So far as we have been able to ascertain, no known facts about the central nervous system have been violated by our assumptions, even though at times we have used simplification and short cuts which are not available in a biological organism. Wherever we have made such assumptions, as in the simplification of a "response unit" to a single two-stage element, the biological equivalent can be clearly indicated. Thus, while we have not proven the validity of statistical separability as the explanation of an organic brain, this explanation now seems to be the most plausible of the available alternatives.

As a machine, the future of the perceptron will depend heavily upon our ability to achieve an efficient and inexpensive A-unit. Studies of components and circuitry are currently in progress, and success seems imminent. We will not speculate here upon the possible applications of such a device. The perceptron is in its infancy, and it would be a mistake to rush it too abruptly towards an adolescence which can still scarcely be foreseen.

As a concept, it would seem that the perceptron has established, beyond doubt, the feasibility in principle of nonhuman systems which may embody human cognitive functions at a level far beyond that which can be achieved through present day automata. The future of information processing devices which operate upon statistical, rather than logical, principles, seems to be clearly indicated.

REFERENCES

1. ASHBY, W. ROSS, Design for a brain, *Wiley & Sons, New York,* (1954).
2. McCULLOCH, W. S., and PITTS, W., "A logical calculus of the ideas immanent in nervous activity", *Bulletin of Math. Biophysics,* 1943, 5, p. 115.
3. ROSENBLATT, FRANK, The Perceptron - A theory of statistical separability in cognitive systems, *Cornell Aeronautical Laboratory, Inc. Report No. VG-1196-G-1,* January 1958.
4. VON NEUMANN, JOHN, The computer and the brain, *Yale University Press, New Haven,* (1958).

APPENDIX TO THE PAPER BY DR. F. ROSENBLATT ∅

Since the draft of the preceding paper was written, a much improved method of analysis has been developed, which should really supersede much of the foregoing material. This method is applicable to networks of "conventional" on-or-off neurones, rather than to neurones of the continuous transducer variety. Each A-unit, a_i, is characterized by a threshold, θ, and a variable output strength, v_i, which serves as the memory variable. If the algebraic sum of the excitatory and inhibitory input signals from the retina is greater than the threshold, an A-unit "fires", and delivers its output, v, to the R-unit to which it is connected. If the measure of the total input signal to the A-unit is α, then we can define the "activity function", $\alpha*$, as

$$\alpha* = \begin{cases} 1 \text{ if } \alpha \geq \theta \\ 0 \text{ if } \alpha < \theta \end{cases}$$

The output signal from every A-unit will thus be equal to $\alpha* v$, and we will assume that the R-unit is "activated" by a net input signal greater than zero, and "suppressed" by a net input signal less than zero. Specifically,

$$R = \begin{cases} 1 \text{ if } \sum_i \alpha_i* v_i \geq 0 \\ 0 \text{ if } \sum_i \alpha_i* v_i < 0 \end{cases}$$

where the summation is over all A-units, a_i.

Now let P_{a_i} = the proportion of A-units responding to stimulus S_i, and let $P_{a_{ij}}$ = the proportion of A-units responding *both* to stimulus S_i and to S_j. The expected values of these proportions are known functions, which have been previously described *(ref. 1)*. While a small perceptron may show appreciable deviations from these expected values, in a very large perceptron, the proportions of responding units should be very close to the expected values. In order to eliminate from consideration the variability of different perceptrons, the following remarks will be restricted to the assumption of a very large, or infinite, perceptron, in a finite universe of stimuli, S_1, S_2, ..., S_n.

If a perceptron is exposed to some stimulus, S_i, and its association system is reinforced, then the total amount of value gained by the set of A-units responding to S_i will be some function of P_{a_i}, which depends on

∅ Added after the Symposium

the particular rules of reinforcement which are applied. For example, if we simply have a rule that the active units gain some increment of value, Δv, and inactive units are unaffected then the total amount of value gained by the set of A-units responding to S_i will be kP_{a_i}, where k is a constant of proportionality. Such a system is called an "alpha system". Now suppose we present stimulus S_i, and apply the above rules for reinforcement. Each of the A-units thus reinforced is likely to respond not only to S_i, but to one or more other stimuli as well. Let us now examine the effect of the reinforcement of S_i upon the sets of A-units responding to each of the other possible stimuli. The measure of the value change in the set of units responding to S_j, as a result of having reinforced S_i, will be called the "generalization coefficient", g_{ij}. In the alpha system, considered above, g_{ij} will clearly be equal to $kP_{a_{ij}}$. If we arbitrarily define our units of reinforcement so that $k = 1$, then we have for the alpha system that $g_{ij} = P_{a_{ij}}$.

In a different system, which we have called a "gamma system", the reinforcement rules specify that the total value, measured over the entire set of the A-units, must remain equal to zero. In this case, if the active units gain a quantity equal to P_{a_i}, then this same quantity is subtracted uniformly over the association set as a whole, so that, again assuming the proportionality constant, k, to be equal to 1, we have for the gamma system: $g_{ij} = P_{a_{ij}} - P_{a_i} P_{a_j}$.

Now if there are exactly n possible stimuli, there will be an n by n matrix of generalization coefficients, g_{ij}. The *rows* of this matrix represent all of the "contributions" from a reinforcement of S_i to the sets of A-units responding to each possible stimulus, S_j (where S_j may be identical with S_i). The *columns* represent all of the possible contributions to the set of A-units responding to S_j, from reinforcements of each possible contributing stimulus, including itself. Let us designate the row vectors of this matrix G_i, and the column vectors G_j^*. Let us represent the expected frequency of occurrence of each of the n stimuli by a positive scalar number, f_i. The set of n frequencies, for the n stimuli, can be represented by a vector, F. The scalar product, FG_j^*, will then be equal to the expected rate of change for the total value of the set of A-units responding to S_j, as a result of being shown each of the possible stimuli at its expected frequency, each one being positively reinforced. Thus, over a period of time, t, assuming a zero decay rate, we would expect the value measured over this set of A-units to grow to tFG_j^*.

Let us now consider the effect of the reinforcement operator, ρ, which is set equal to +1 if the response $R = 1$ is to be "positively reinforced",

and -1 if the response is "negatively reinforced", for a given stimulus. Assume that this reinforcement is applied consistently for each stimulus, i.e., whenever stimulus S_i occurs, its responding A-set is either positively reinforced or else negatively reinforced on each occasion. Then the generalization coefficient, g_{ij}, must be multiplied by the sign of ρ, and the column vector now consists of elements $\rho_i g_{ij}$. If we sum the elements of each such column vector, multiplying each element by its appropriate frequency, f_i, then the array of column sums forms the vector:

$$V = \sum_i f_i \rho_i G_i = (v_1, v_2, \text{---} v_n)$$

Each element of this vector, v_j, represents the expected rate of change of the value of the A-set responding to S_j, as a result of exposure to a random sequence of stimuli from the universe in question. If v_j is positive, the stimulus S_j will tend to turn on the response $R = 1$, while if v_j is negative, it will induce the response $R = 0$.

Let us first consider the simple alpha system, with no decay. The generalization coefficients for this system, g_{ij}, are simply equal to $P_{a_{ij}}$, and consequently are all positive. Suppose S_i occurs, at time t_1. This immediately gives us the value vector, $V(t_1) = G_i$. Since every element of this vector is positive, it follows that whichever one of the n possible stimuli occurs at time t_2, the response $R = 1$ will occur. But the occurrence of this response means that we will again add an all-positive vector to V, and, in fact, it is clear that no element of V can ever become negative. Consequently, this system behaves as a "Class C perceptron", in accordance with our first theorem.

Now consider the gamma system, for which $g_{ij} = P_{a_{ij}} - P_{a_i} P_{a_j}$. It can be shown that for a perceptron in which the A-units are randomly connected to the retina, and each A-unit receives a fixed number of excitatory and a fixed number of inhibitory input connections, it will always be true that

$$P_{a_{ij}} = P_{a_i} P_{a_j}$$

provided that the area of the intersection of the two stimuli is equal to the product of their normalized areas (i.e., $C_{ij} = R_i R_j$, where C_{ij} is the common area, and R_i and R_j are the retinal areas of the stimuli S_i and S_j, with the area of the retina taken as unity). If stimuli can occur with equal probability in any retinal position, then their expected intersection will indeed be equal to this value of C_{ij}. If all intersections are equal to their expected value, the corresponding value of g_{ij} will be zero. If, however, two stimuli have a common area, C_{ij}, greater than this expected value, g_{ij} will be > 0, and if the stimuli are disjunct, or have a common area less than the expected value of C_{ij}, it follows that g_{ij}

will be negative. Even though the expected value of C_{ij} for a class of randomly placed stimuli is equal to zero, it can be shown that, due to nonlinear characteristics of the $P_{a_{ij}}$ function, the expected value of g_{ij} over the class of stimuli may be greater than zero. This is actually a necessary condition to permit a response to generalize consistently over the class of stimuli.

To analyze the simplest possible case, consider a stimulus universe consisting of two disjunct stimuli S_1 and S_2. From what has just been said, the generalization coefficients g_{11} and g_{22} will clearly be positive, and g_{12} and g_{21} will be negative. Consequently, each stimulus will generalize positively to itself, and negatively to the other stimulus, and the perceptron will necessarily behave in the fashion of a Class C' system, described in Theorem 2. It should be noted, however, that no decay has been postulated in this system, and none is necessary. Even though the magnitude of the values of the A-units, and consequently the variance of the value, will grow without bound, the system remains well-behaved, so that the Corollary of Theorem 1 is disproved, at least for this special case.

Due to the fact that the expected value of g_{ij} for stimuli of opposite classes, in an environment in which every placement of the stimulus on the retina is equally probable, is equal to zero, it is desirable in some cases to add a constant loss rate, ϵ, to the value dynamics. In this case, every A-unit always loses a slight decrement of value, in addition to any other changes upon which the loss may be superimposed. This gives us a modified gamma system, with the generalization equation:

$$g_{ij} = P_{a_{ij}} - P_{a_i} P_{a_j} - \epsilon$$

If a decay rate, δ, is also incorporated in the system, it is clear that in the absence of any reinforcements, the system will stabilize with all units at a level in which the values are proportional to $-\frac{\epsilon}{S}$, at which point the decay rate (which is now positive) will exactly balance the rate of loss. This system has the advantage of adding a slight negative interaction between stimulus classes, which otherwise might develop a positive relationship. Again, due to the nonlinearity of the $P_{a_{ij}}$ function, this negative interaction tends to affect "well separated classes" more strongly than classes of "similar" stimuli, and thus generally helps in establishing a desirable separation.

The performance of an infinite perceptron, in which g_{ij} is obtained from the above formula, is illustrated in *fig. 1*. The reinforcement operator, ρ, was set equal to +1 for $R = 1$, and to 0 for $R = 0$, i.e., the system was reinforced only for $R = 1$, and was left unchanged for $R = 0$. In this experiment, two classes of stimuli were assumed: one class consists of horizontal

SPONTANEOUS ORGANIZATION OF INFINITE PERCEPTRON IN ENVIRONMENT OF 4 x 20 VERTICAL & HORIZONTAL BARS

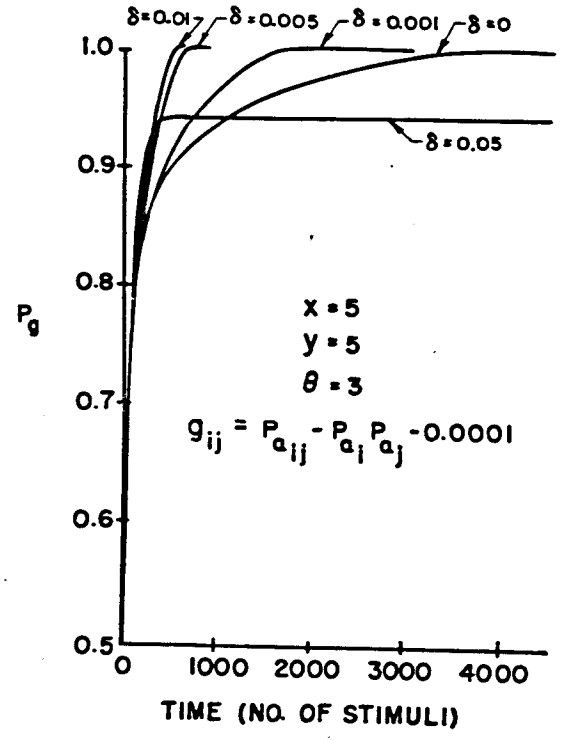


Fig.1

EXPECTED WAITING TIME TO PERFECT PERFORMANCE, AS A FUNCTION OF DECAY RATE (MEANS OF 10 RUNS)

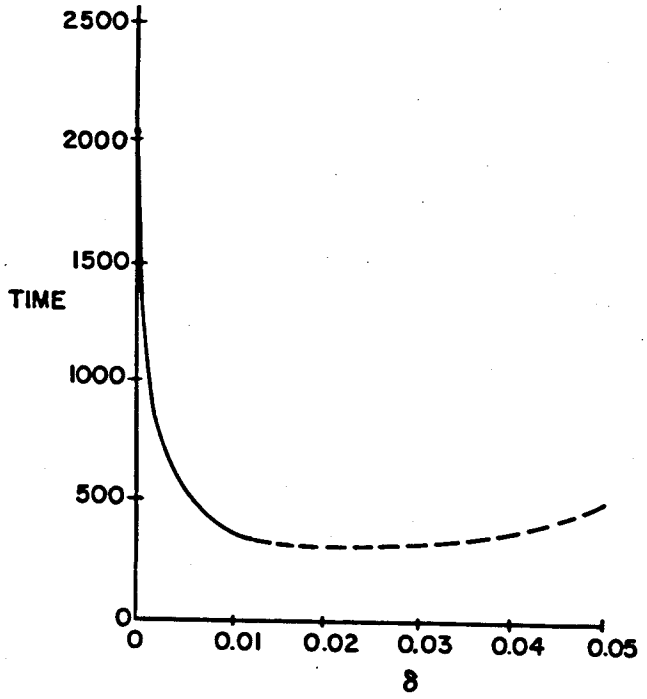


Fig.2

bars, covering an area of 4 x 20 retinal units, and the other class consists of vertical bars of the same area. The retina is a 20 x 20 mosaic, and is assumed to have a torroidal connectivity, i.e., if a stimulus image is shifted off of one edge of the retinal space, it re-enters at the opposite edge. This guarantees a uniform coverage of all retinal points by each class of stimuli. The first set of curves shows the probability of correct (i.e., consistent) generalization of the response $R = 1$ to one class of stimuli, and $R = 0$ to members of the other class, for different values of the decay rate, δ. Note that even with a zero decay rate, the system eventually discriminates perfectly between the two classes. As the decay rate increases, performance gradually improves, up to a point where, with $\delta > 0.01$, the perceptron begins to "forget" too rapidly, and performance becomes unstable. This condition is also reflected in *fig. 2*, which shows the expected waiting time to perfect performance, as a function of δ.

REFERENCE

1. ROSENBLATT, F., The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain; *Psychol. Review*, **65**, 1958, 386.

DISCUSSION ON THE PAPER BY DR. F. ROSENBLATT

MR. E. A. NEWMAN: Several years ago Mr. Day, of the N.P.L., and I thought up a device which I will say a little about. This assumed a matrix of photocells as a sensing network, random connected to an association network which in turn was connected to a response unit. There was feedback between the input and output of the association network via the response unit, arranged to reinforce association cell activity by reducing gate thresholds should the response to a stimulus be the desired one. Since in England we have not the large amounts of programming effort available to our friends abroad we never programmed the device.

We would have liked our device to possess properties, which, as far as we could see it did not have. The first is that if presented with a slowly changing shape it would infer that all the shapes covered by the transitions were related, the second was that if presented with a set of shapes all varying but slightly from each other, it would infer they were related, even if other shapes were perceived in between. What it did do was to relate in a random way some items out of a set of coincident ones, afterwards putting special stress on this accidental selection, and to be more likely to relate inputs which followed in time but had no continuity of spatial pattern than those that had. This I suggest is not quite what one wants to do.

Most patterns that interest us contain a great deal of special connectivity and continuity. The information needed to specify that a pattern is continuous, and not a snow storm is very large indeed. In our system information about spatial order was thrown away in the radom connection to the association network and had to be learnt. An association network which was connected to the photocell matrix in a spatially ordered manner would be better in this respect.

Our device took note of too much information. The only interesting part of a pattern is its edges, and these only if they alter in time. What we would have liked our device to do was to react to changes in pattern in space time. Apart from any other aspect of the matter, a device which takes note of other than space time changes rapidly uses up all its storage.

Our device reached different end states for different kinds of pattern, but these end states were patterns of dots which had to be recognised. There is not much point in knowing that a device gives a 1:1 correspondence between input and output state, if one cannot recognise the output states. I think the device has at least something in common with the perceptron. Would Dr. Rosenblatt explain how this latter device overcomes the limitations given above.

DR. SHUEY: I have four questions. I would like to say, first, that I think Dr. Rosenblatt's work has contributed a great deal to the analysis of networks of the neural type. My questions do not refer to the new material Dr. Rosenblatt has just presented, but to the material in the printed paper and other published material which people in both England and the United States have seen.

First, some of the initial reports made a great point of random distributions and probability. I have a question as to whether randomness and probability are philosophical essentials or merely a very significant feature of this particular analysis. I do not wish to de-emphasize the importance of analysis. I am merely raising the point as to whether probability is a philosophical necessity in networks of this type.

Second, it seems to me that the detailed connectivity either in a precise fashion, or in a statistical fashion, should be determined by the class of objects or environment that you want the device to recognize. I think I can illustrate my point quite simply if I take a pattern recognition system with three levels. Let the first level contain N binary receptors. Let the second level of binary elements have one element for each of the 2^N input states. Let the third level of binary elements have one element for each possible association that can be formed between input states; there are 2^{2^N} such associations. A specific input can belong to many associations. In other words, given a set of 2^N points, there are 2^{2^N} subsets that I can form from these points. 2^{2^N} soon becomes a very large number as N is increased and it is, of course, even larger in a non-binary system. I do not believe that you want a system that will be able to make all possible associations. I think the possible associations should be determined by what you want the machine to do.

It seems to me that, if you assign the original connections at random, you are in a sense going to make a machine that is equally sensitive to all classes of patterns. I do not believe you want to do this any more than you as a person want to be able to differentiate snowstorms on a TV screen. The differentiation of individual snowflake patterns is in general of no interest to you.

I think one has to place some constraint on the connection. In the perceptron, you do not have all possible connections. You can change the gain of the connections, but you cannot change the connections. If with a given number of elements and connections, I am allowed to change the connections, I shall have a much more flexible system. There is a parallel in the nervous systems of animals. If the connections of dendrites and axons are in part determined by environment, the biological machine will be more flexible than if environment can determine only the strength of those connections. I understand, although I am not a physiologist, that this is an unanswered question as far as the cortex is concerned.

In the example I took earlier, the 2^{2^N} possible associations is deceiving. This is not necessarily the number of elements needed to form that number of associations. The associations might be formed by the state of another system. In a binary system with N_1 elements, it is only required that $2^{N_1} \geq 2^{2^N}$. Possibly, the perceptron does this type of thing.

My third question is related to one of Mr. Newman's points. In the experiments which have been done, if you continuously distort one object into another object, does the probability of making a correct selection smoothly go over or does it jump as a step function?

Fourth, can you make any statement relating how the perceptron discriminates to how we as individuals discriminate? I realize that this is a speculative field.

DR. A. M. UTTLEY: I want to say something about this 2^n.

We know that a set of n binary inputs has 2^n logical states, 2^{2^n} logical functions of those functions, and so on indefinitely. But each of our "outputs" to the external world provides a set of proprioceptive "inputs" to the analyser, in fact, some of the 2^n inputs and *not* one of the 2^{2^n} logical functions which, anyway, is just 0 or 1 depending on the input state.

Consider now the problems of a computer learning to produce an output O_1 which is a logical function of n inputs, I_1, ... I_n. Let there be an input S_1 to indicate success in choosing O_1 at random. Also let the random choice of O_1 be fed back as an input to the computer. We now have a machine with $(n + 2)$ inputs and 2^{n+2} units which can solve the problem. If the machine has, in addition, to learn a second output O_2 which is another logical function of the inputs then there must be 2^{n+4} units, and so on.

MR. STAFFORD BEER: I should like to address my remarks to the written paper, rather than to any of the chimerical perceptrons which have been floating about outside it.

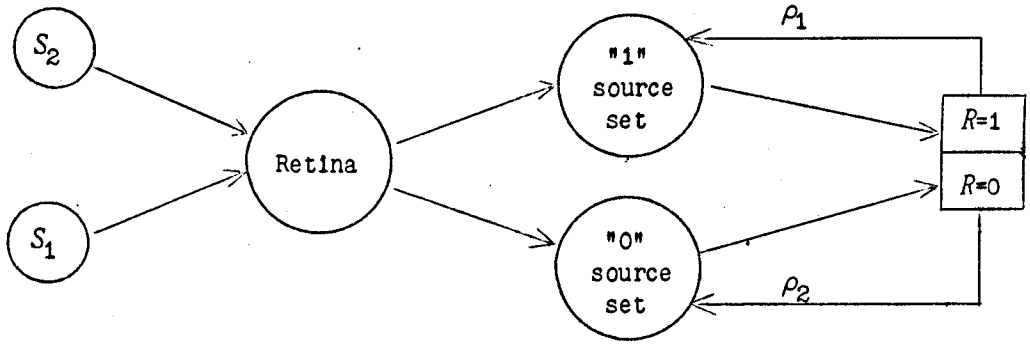
My criticism begins in the first section where Dr. Rosenblatt says "symbolic logic or Boolean algebra", and later on the same page "Boolean algebra or symbolic logic". That is, he presumably takes these terms as synonymous whereas they are not. Surely Boolean algebra is a small sub-set of the class of symbolic logics. There is a whole range of these, most of which are non-Boolean. For example, there is the predicate calculus.

The confusion the author has made leads him to a false dichotomy between Boolean algebra and statistical mathematics as tools. This excludes the most useful semantic tools at our disposal. For example, the kind of logic used by Ashby in this work is an algebra of classes, the precise virtue of which is that the classes need only be specified as bounded in dimensions that are of interest, and that their behaviour is entirely probabilistic. This gives us the virtues of statistical methods, without the lack of logical distinctness which seems to characterize this paper. I would

concede that this might be just a question of terminology (if it is, my point is merely captious) were it not for the disastrous consequences which seem to follow.

For example, an A-unit (one of the association cells) is conceived as a set of ν - values. These may be, it seems to me, *either* a point-set covering the neurons in the A phase-space, *or* a set of values through time taken up by one neuron, *or* again the values taken up over time by the phase averages. By spurning the kind of logic which would distinguish between these possibilities, Dr. Rosenblatt has left himself open to a charge of circularity. For he proposes two theorems which clearly purport to demonstrate ergodic properties in the perceptron, whereas the ergodicity is probably subsumed in the logically loose definitions of the ν - set from which he starts.

From that example, I shall now go on to generalize the charge of circularity. In Theorem 2, Dr. Rosenblatt distinguishes dichotomously between two classes of stimuli. I shall amend his diagram by doing the same. All stimuli must belong to one class or the other. This is an arbitrary division of the set of all stimuli, but one which I am entitled to make; note that I say nothing of the criteria by which a particular stimulus is identified as belonging to S_1 or S_2. I say only that any stimulus can be regarded as belonging to one of two classes and as feeding into the retina thus:-



This does *not* say what effect these stimuli have on the system.

The possibility is given in this system that events take place which cause the response to go to "1". Therefore we can collate these events, and assert that a class C of them exists if and only if the responses go to "1":-

$$\exists\ !\ C \equiv R \rightarrow 1$$

Equally, there exists another class of events if and only if the responses go to "0". We shall extend class C to include these too, and assert the existence of the class in which events take the response either to "1" or to "0", but not both:-

$$\exists\ !\ C \equiv R \rightarrow 1 \wedge 0$$

What information is given in the system about the role of the stimuli? In the system as shown there is nothing to show what kind of "switch" there may be across the retina; there is nothing to show that a member of either stimulus class will influence an "1" source set more than a "0" source set. So either kind of stimulus is capable of evoking either response. The account of class C may therefore be extended to:-

$$\exists\,!\ C \equiv S_1 \vee S_2\ (R \rightarrow 1 \wedge 0) \qquad \ldots\ldots \quad (1)$$

Since the system is no more restricted than this, a number of logically weaker proportions can be asserted. The one which interests me is this:-

$$\exists\,!\ C' \equiv S_1\ (R \rightarrow 1)\ S_2\ (R \rightarrow 0) \qquad \ldots\ldots \quad (2)$$

That is to say that since the first kind of stimulus is capable (expression 1) of evoking the unit response, then the cases of this happening can be collected as a sub-set. Similarly, since the second kind of stimulus is capable of evoking the zero response, these too can be collected as a sub-set. There is no reason why these two sub-sets should not be considered together as the class C′ (expression 2).

These two classes, C and C′, certainly exist in this system, although the second could be empty, and although there are many others. What do we know of them? For example, how much information can they each transmit? Class C transmits zero bits about a specific stimulus, because it may evoke either response. Class C′ transmits one bit about a specific stimulus, because that stimulus can only evoke *one* of the responses.

Looking at the diagram, how can this happen? What mechanism must be inferred at the retina to account for this information flow? If zero bits are transmitted from either S_1 or S_2, the retina must be acting as a kind of random switch. It is in fact uncoupling the stimulus system (S) from the response system (R). If one bit is transmitted, on the other hand, then in some sense S and R are coupled together. This coupling cannot be asserted categorically on the evidence; we can however say that a tendency must exist at the retina to couple (S_1R_1) and (S_2R_0), otherwise the system would not transmit.

Now this inference can be quantified. The conditional probability of the response given the stimulus must, for the class C, tend to 0.5; because there is nothing controlling the "switch". But for the class C′, the conditional probability of the response must tend to 1.0. That is:-

$$\exists\,!\ C : Pr\ (R|S) \rightarrow 0.5$$

$$\exists\,!\ C' : Pr\ (R|S) \rightarrow 1.0$$

The nature of the logical correspondence between S and R follows from this. In the first case, C, it is strictly unknown. But is is important to recognize that it *could* be a "one-many" correspondence. In the second case,

C′, however, it must tend to be a "one-one" correspondence, because the variety in the right-hand part of the system must now be limited by a function of the input α.

What limitations do these inferences about logical correspondence produce in the A-set? In the first case, C, the variance of the value of ν can very well increase. Since, as was shown, the *S-R* correspondence could be "one-many", or at least might include "one-many" transformations, this variance can go on growing indefinitely. In the C′ case, however, the variance of the value of ν must tend to a limit, because the variety is limited as $f(\alpha)$.

So the chain of logical properties propagates from the existential propositions with which we began. Now there is only one thing really going on in this simple version of the perceptron. This is given by the transfer function of the neuron α_i at time t, which the author tells us is equal to $\alpha_i(t).\nu(t)$. What is important about the system is the way it grows. The α input is the same for both classes of events considered, and so the growth function in which we are interested concerns the ν (t) element, and is expressed as $d\nu/dt$. In both the classes of events here considered, this function will involve the α input and the totality of feedback information ρ_1 and ρ_0 from the response sets to the source sets: that is, $\alpha\Sigma\rho$. Consider class C of events. There is an assumption of equal likelihood in the response alternative. Therefore the ρ feedbacks will frustrate each other in the long run, and the growth function $\alpha\Sigma\rho$ will permit the A-set ν - value system to increase indefinitely as prescribed by the inferences drawn above. In the class C′ case, however, this system tends to a limit, and $\alpha\Sigma\rho$ will not satisfy the conditions. If the ν - value system is to be limited while ν grows, the growth function must include a decay function arranged to increase monotonically with ν. Thus the growth function for class C′ may be expressed as $\alpha\Sigma\rho - \delta\nu$.

This completes *a possible description* of the Rosenblatt system. I say "a possible" because there are many more possible kinds of behaviour; I say "description" because this is nothing to do with proof - it is to do with displaying the system as given. I think the whole contents of the author's "theorems" are displayed by my description. A perceptron as defined is simply going to behave like this. It is also going to behave in other ways, because we could choose to define it in other ways. So the paper does not actually demonstrate anything. And since other kinds of behaviour are possible, no doubt the author has observed them. This may explain why his written accounts of the perceptron seem to be about different things, and why his verbal presentation today seemed to bear no resemblance to the written paper.

What then is the logical status of the two announcements which the author labels "theorems"? His method seems to be this. He asserts a

proposition about possible stochastic behaviour in the long run called a theorem. He asserts that another proposition is a corollary of this, and then has a discussion from which it emerges that the corollary class is not empty. Therefore the theorem class is not empty. This, it seems to me, tells us nothing more than that in the long run every configuration of which the system is capable will occur, and that Rosenblatt finds two of these configurations interesting. The "theorems" are, I would say, assertoric existential propositions about the system; and you neither prove nor disprove such descriptions. They are simply there, and they may or may not be interesting. What you *can* show about such propositions is that they are internally consistent, and this is I think precisely what I have done.

It follows, in my opinion, that the elaborate mathematical edifice from sections 5 to 8 in the paper is an array of tautologous remarks about the existential description given in sections 1 to 4. But as it does not actually say anything, there was no need to write it. What it does do is to build up an ethos of potency around the perceptron. On this account, someone might unfortunately be mesmerized into believing that this device is the ultimate unit of the brain that lies behind intelligence.

I feel sure that Dr. Rosenblatt would not want to make such an exaggerated claim as that; but I think that there are tendencies to exaggerated claims to be found in the paper. For instance, one which ought really to be mentioned is that, apart from the Homeostat, the perceptron is the first machine "to show spontaneous improvement". This is not so. But perhaps Dr. Rosenblatt rightly assumed that, after all, everyone here must certainly have heard of Dr. Uttley's work.

To sum up, I am saying that what Rosenblatt has shown is not that the perceptron is the necessary explanation of certain time trends observable in the brain, nor that it is a sufficient explanation. He has shown that, if we select various sub-sets of the stochastic behaviour of this worthwhile device, we shall find them interesting. And for this I, for one, would like to thank him.

Dr. W. K. TAYLOR: I think that a number of contributions to this Symposium have some similarity to Dr. Uttley's work on the classification of signals in the nervous system *(ref. 1)* and on probability computers *(ref. 2)*. Dr. Rosenblatt has introduced a difference that I described in connection with my work on the simulation of nervous systems *(ref. 3)*. This difference can be illustrated by considering an Uttley type model neuron unit with n binary inputs and a threshold m. The unit gives an output of "one" only if the sum of the "one" inputs is greater than or equal to m, otherwise the output is zero. Rosenblatt's unit, however, is equivalent to the one that I described at the Third London Symposium. The input signals were effectively continuous variables represented by pulse repetition rates and the output

pulse repetition rate was proportional to the algebraic sum of the input signals, inhibitory inputs being given a negative sign. It was shown that a suitable network of the units could learn to classify shapes if the transmission strength (Rosenblatt's "value") of the connections increased with use. There was no need to introduce a separate "growth rate control" input. This latter postulate does not appear to have a physiological basis whereas recent electron-microscope pictures of the brain *(ref. 4)* tend to support the hypothesis that the area of contact at each synapse, and hence the transmission strength, is controlled by the impulses arriving at the synapse.

One thing that seemed to be missing from Dr. Rosenblatt's paper was an estimate of how many shapes a perception with a given number of units could be expected to classify. How many A-units would be required to recognise a typewritten alphabet, for example?

Finally, Dr. Rosenblatt suggests that if he could find suitable hardware for constructing A-units, he could actually build a perceptron. I would like to ask him why he considers this to be worth while when he has already shown that he can simulate its operation on a digital computer.

DR. J. McCARTHY: I would like to raise an issue which concerns all of the neural work on perception - that is the work on perception which is based on nerve models, and I believe it applies also to the pandemonium and some of the work on speech recognition which was described this morning. The problem of perception can be divided into parts, at least two of which are discrimination and description. Now all of the work which has been described has been on the problem of discrimination; that is, a finite number of classes of stimuli are discriminated, and in the training process at least one example from each class is presented. However, much of our own perception can be described as description, whereby we can perceive something and generate a description of it, and we may be able to do this without ever having seen the thing before. I would like to give an example from letter recognition. All of us can tell A from B, and the previous speakers have made it plausible that they can train various devices to tell A from B, but we can also do the following. We can take this figure (the Russian letter) and describe it by saying, for example, that the figure consists of

REFERENCES

1. UTTLEY, A.M.; The Classification of Signals in the Nervous System, *R.R.E. Memorandum No. 1047*, 1954.
2. UTTLEY, A. M.; Temporal and Spatial Patterns in a Conditional Probability Machine, (Automata Studies, Edited by Shannon, C. E. and McCarthy, J., *Princeton University Press*, 1956.)
3. TAYLOR, W. K.; Electrical Simulation of some Nervous System Functional Activities, Third London Symposium on Information Theory, Edited by Cherry, C., *Butterworths*, 1956.
4. GRAY, E. G.; *Journal of Biophysical and Biochemical Cytology*, In Press.

four line segments, one of which is horizontal and the other three project up from it, one from each end and one from the middle. We can describe this without ever having seen it before.

I have done some work on the question of generating descriptions of visual images, but this work has not had a neural basis; it was a proposal for a programme which takes a digital representation of the picture, walks around it, so to speak, and attempts to determine the line segments.

Some kinds of description can be obtained as sequences of discriminations. In particular, we can get descriptions of speech or writing in this form, because of their sequential nature provided we can make discriminations for each letter. This constitutes a description, but I think it can be shown that, in the case of two-dimensional data, it is not possible to get a description system entirely out of discriminations.

DR. L. M. SPETNER: I would like to ask Dr. Rosenblatt if what he called $r_{\alpha\bar{\alpha}}$, the correlation between the α inputs of the association cells and the average of α, is really a good measure of similarity of stimuli. In some special cases it may be all right, but it is not clear that it is sufficiently general. The use of such a criterion, particularly in the C' type perceptron, means that the machine will tend to classify its environment according to its own construction. For example, the experiment quoted on page 448 of Dr. Rosenblatt's paper, which showed that a class-C' perceptron correctly divided all squares into those which were on the left and those which were on the right, seemed to be a direct result of the way the sensory and association cells were connected. I suspect that this perceptron might do a much worse job if it were asked to distinguish on some basis other than position. It seems to me that organisms do not operate this way. That is, they do not really go off on their own and begin classifying things in a vacuum; but rather they act more like a perceptron whose response is being forced. I do not mean that all organisms always have a formal teacher, but rather, I believe there is a continual reinforcing of responses through some kind of feedback from the environment to the organism.

I would like to think of the perceptron as a self-organizing transformation from a stimulus to a response. Each stimulus that may be applied to a perceptron can be considered as a point in a multi-dimensional Stimulus space. We might consider the response somewhat more generally than Dr. Rosenblatt does by saying that a response will consist of a *set* of numbers in each of the various response cells. Then any particular response can be considered as a point in a multi-dimensional Response space. (See *fig. 1*). The association cells then would provide a mapping of the Stimulus space into the Response space. Now it may well be that stimuli which I wish to consider similar may be very far apart from each other in Stimulus space, but if the perceptron is to organize its environment into categories of

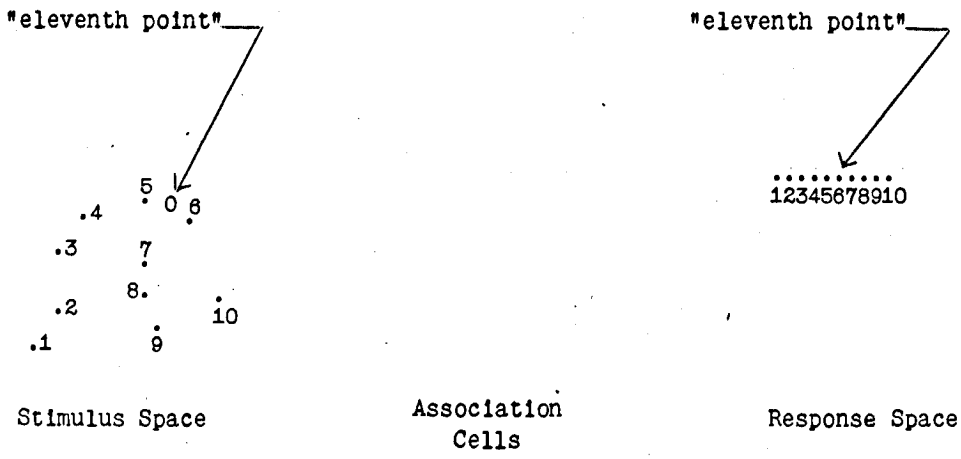


Fig.1

similarity which I choose to designate, then I shall force similar stimuli to have responses which are close together in Response space. I believe similarity of stimuli is really a rather arbitrary concept; at one time, I may wish to consider all triangles as similar and at another time I may want to consider all objects on the left side of the field of view as similar, whether they are triangles or not. Thus it is clear than arbitrary stimuli which I may wish to consider similar might well be very far apart from each other in Stimulus space. This allows for the concept of similarity to be imposed by an arbitrary experimenter rather than having it tied into the construction of the machine. Viewing the perceptron as a transformation from Stimulus to Response space also leads quite naturally to the concept of generalization. Referring again to the figure, let ten stimuli be represented by the ten points shown in the stimulus space. If I wish to consider these stimuli as similar, then I force their responses to lie in the restricted portion of Response space as shown, so that the perceptron shall learn the responses that correspond to these particular stimuli; in that it can reproduce each response for given each stimulus. Now we imagine that we present to the perceptron an eleventh stimulus that lies somewhere in between two that we have already given it in stimulus space. If our mapping has been sufficiently continuous, then it will appear as shown in the correct portion of Response space, and hence, we have reduced the concept of generalization to interpolation or extrapolation, provided one has achieved a continuous type of mapping.

DR. F. ROSENBLATT (in reply): I will try to touch very briefly on each of the points which have been raised.

The system which Mr. Newman described does indeed seem to bear a considerable resemblance to the perceptron, although the nature of the reinforcement function is somewhat different. Several of the comments which have been made here seem to reflect a common misimpression that a reinforcement function which changes thresholds is equivalent to a reinforcement function which changes the output of a neurone. This is not the case; by reducing the threshold, we are making a cell responsive to a greater variety of stimuli and thus changing the information transmitted by the cell, while in changing the output, we are changing the *weight* of the transmission channel of which the cell is a part, without in any way changing the signification of the information transmitted. A difficulty in systems which operate by changing thresholds is that as the threshold goes down, the cell becomes responsive to a greater and greater variety of stimulus events, and in the limit responds equally readily to any input, thus conveying no information at all. While such effects may not have been the source of the difficulty in Mr. Newman's device, they are apt to present an initial handicap to any system operating by general threshold reduction, rather than by the strengthening of specific transmission channels. I would predict that Dr. Taylor will encounter the same difficulty when he goes to larger models of his own system, if he continues to reduce the thresholds without a lower bound.

With regard to Mr. Newman's problem of recognizing slowly changing shapes as identical, I must say that the perceptron does indeed have such a tendency. For example, if we change a triangle gradually into a square by a series of progressive distortions, the perceptron will sometimes recognize the two stimuli as members of different classes, and sometimes will place them in the same class, depending on the frequency distribution of the various possible intermediate states. If there is a bimodal distribution, with more "good" triangles and "good" squares than distorted figures, then we can design a perceptron which will spontaneously arrive at the desired classification. If, on the other hand, all transformations (including those which are halfway between triangle and square) are equally likely, there is no logical basis for concluding that there are "two kinds of things" here, except by an arbitrary convention. How far we have to broaden the cleft between modes of the distribution in order to make the system discriminate reliably is something which we are now beginning to investigate. We are also examining the nature of the probability transition which Dr. Shuey asks about in his third question. In some cases, it seems possible to design a perceptron which will break up a continuum (say between squares on the extreme left and squares on the extreme right of the field, or between circles and ellipses) into a number of discrete domains, assigning a separate response to each. We find that the domain over which the

perceptron generalizes is governed in part by its design parameters, and in part by the particular environment of stimuli in which it is located.

The problem of the serial order in which events occur does not seem, in the perceptron, to present as serious a problem as Mr. Newman has suggested. While there is a bias, in some systems, to associate consecutive events, a perceptron which is designed so that the values of inactive A-units lose an amount which just balances the gain of the active units (what we have called the gamma-system) may actually exhibit interference between successive stimuli, if they are sufficiently different from one another.

With regard to Mr. Newman's comments on the importance of boundary or contour detection as a means of concentrating on the important input information, I am in full agreement. In addition to the networks which Dr. Taylor has described *(ref. 1)*, an approach to this problem has been considered in the first report on the perceptron program *(ref. 2)*.

Dr. Shuey has asked about the philosophical necessity of randomness in the perceptron. I do not believe randomness to be a philosophical necessity, but rather a practical one. In analyzing a nerve net about which we have complete information — a knowledge of all of its connections, the state of every unit, etc. -- it is clearly possible, in principle, to perform an exact logical analysis of its response to every logically possible input configuration, count those cases in which the response is the desired one and those in which it is not, and thus determine exactly the "probability of correct response" for the finite environment in question. In practice, this can hardly be recommended. Biological brains and physical environments being what they are, it is unlikely that we shall ever have the complete information which is required for a procedure of complete enumeration, and, moreover, this would still not enable us to talk about brains as a class of systems, but only one particular brain having the particular connections described. To talk about brains as a class we would further have to enumerate every admissible system of connections. It seems to me that the probabilistic approach is our only recourse at this point. It offers the further advantage that the necessary conditions for a class of systems to work *in general* (even though individual models may fail) can be stated in terms of a simple set of rules of construction, or statistical parameters, rather than in terms of a detailed and unmanageable wiring diagram. If we grant that we can never, in practice, attain complete information about such systems and their environments, then the question of whether probability is philosophically essential becomes entirely academic: if we are to analyze the systems at all, we really have no choice in the matter.

Concerning the relationship of perceptron discrimination to human discrimination (Dr. Shuey's fourth question), I think it is premature to expect any close correspondence between the "psychology" of any of our brain models and that of human subjects; too many ingredients are still missing in the perceptron. One important difference between the learning curves

observed for spontaneously organizing perceptrons and those of humans, however, is worth noting: The learning curves for the perceptron are convex, whereas those for a human subject under similar circumstances would certainly be concave. That is, if a perceptron is required to distinguish horizontal from vertical bars, in a spontaneous learning experiment, it quickly learns to classify the first sixty or seventy percent of the cases, and takes longer and longer to establish the correct response for the few stimuli still unclassified. A human subject, in such a problem, once he has achieved an "insight" would undoubtedly jump to 100% accuracy immediately thereafter.

I am sure I will not be able to do full justice to Mr. Beer's entrancing remarks concerning my paper. I find that I am almost mesmerized into believing them. Perhaps, however, I can arrive at *a possible description* of what Mr. Beer says that I am trying to say. I say "a possible" because I may have missed some subtlety in his analysis which makes it more relevant than it appears; I say "description" because this has nothing to do with a disproof, but with a set of fundamental misunderstandings of my paper evidenced by his remarks.

First of all, let me express my complete agreement with Mr. Beer about the lack of rigor in the paper. Had a more rigorous analysis been possible at the outset, I would have been spared the necessity of pointing to mistakes in my own work in order to defend its logical status, as I am about to do. Mr. Beer would, perhaps, have been happier with his whole analysis had the "corollaries" been labelled "lemmas" and placed before the theorems instead of after. This would actually have been preferable, as the "corollaries" are actually independent of the "theorems", although the reverse is not true -- the corollaries were used to establish a case in point, on the basis of which the more general existence theorems might be asserted. Mr. Beer suggests that this analysis is a rationalization of behavior which I must have "observed" previously in the perceptron. This is not true. Neither of the types of behavior described in the paper were observed until after the analysis was completed, at which time a new simulation program was written. It then turned out that while we did indeed obtain the predicted behavior most of the time, it was by no means infallible. Therefore, the analysis, far from being tautologous, must actually be *wrong* if interpreted rigorously. This led to the revised analysis which I outlined this morning, and which does not represent a completely new set of chimerical perceptrons, as Mr. Beer suggests, but rather a new way of looking at the same kinds of systems that we have been discussing all along.

Now what is the substance of Mr. Beer's argument? First of all, it should be noted that Mr. Beer mistakenly takes the terms "Class C" and "Class C′" to refer to classes of *stimuli*, whereas they were intended to refer to classes of *perceptrons*. Secondly it seems to me that he has systematically inverted the roles of premises and conclusions, from the original paper -- an error for which the admittedly peculiar relationship of "theorems" and

"corollaries" is probably responsible. He then proceeds to try to deduce a set of characteristics for the A-system, and particularly a system of value-dynamics, which is consistent with one of the two assumptions that (1) the perceptron, in the Class C case, effectively "uncouples" the responses from the stimuli, so that the probability of the response R=1 is 0.5, regardless of the stimulus, *or* (2) the perceptron, in the Class C′ case, has a switching mechanism (unspecified in Beer's description) located, of all places, in the retina, which is specifically designed to evoke the response R=1 for one class of stimuli, and R=0 for the other class. It should be noted, first of all, that Mr. Beer's "Class C" perceptron does not in any way correspond to our own "Class C" system. The Class C perceptron, in the original paper, is one which tends towards a terminal condition in which either (a) every stimulus evokes the response R=1 *or* (b) every stimulus evokes the response R=0. The reason there is no information transmitted in this case is not that the responses are random, but rather that they are inevitably the same, whatever the stimulus happens to be. Actually, as our analysis shows, increasing the variance of the values to infinity does not lead to random responses, as Mr. Beer infers, but rather to the perfectly consistent responses characterizing our Class C case, *provided the indicated rules of reinforcement are observed.*

I find myself unable to recognize the source of Mr. Beer's difficulty in understanding the nature of the v-set. Each A-unit, a_i, is characterized by a value, v_i, which is assumed to be initially zero, and which changes through time in accordance with some clearly specified rules. Again, the *idée fixe* that these rules are a *consequence* of some preexisting classification system, rather than the cause for the stochastic development of a classification system in an initially random machine, must be responsible for the confusion.

I think Mr. Beer's assertion that the classification scheme must somehow be built into the retina suggests that he has missed the principle point of my paper. Let me repeat, therefore, that the C′ perceptron arrives at a classification scheme *independently* of any scheme which previously exists in the mind of the experimenter. The classification arrived at by the C′ type of perceptron does *not* depend on the experimenter, nor is it built into the retina; the experimenter is forbidden to intervene in any way which might help this machine to arrive at the "right" decisions. Uttley's models do seem to represent a possible first step in this direction, and should not have been omitted from my original review of the field. Actually, I think Taylor's system comes at least as close to being a sponaneous classification device, in this case, as Uttley's, although in Taylor's model, as I understand it, the system is actually forced to learn the "proper" associations by presenting an "unconditioned stimulus" together with the "conditioned stimulus" during learning — a procedure which is logically analogous to our "forced learning" procedures with the perceptron *(ref. 2)*.

With regard to Dr. Taylor's remarks, I don't believe either of us can claim to have invented the threshold, but I agree that in setting his threshold equal to the number of input connections to a cell, Dr. Uttley is somewhat straining the biological credibility of his system. Clearly, our neuron models are very similar, as they are both based on the same fund of established physiological knowledge. The one important difference which I see in our neuron models is in the choice of a suitable memory variable -- in Dr. Taylor's case the threshold, and in my own case, the strength or "value" of the output signal, I have already indicated my reasons for this choice of variable. I would like to refer Dr. Taylor to Appendix V of my original paper *(ref. 2)*, in which it is shown that a reinforcement of the cell as a whole is logically equivalent to a reinforcement of particular connections, for an optimally designed perceptron. I agree, however, that at present the reinforcement of specific connections seems more plausible than the reinforcement of the cell as a whole, and I am particularly grateful for the reference to the electron-microscope work on synaptic areas, which I look forward to reading.

The "vocabulary size" of a perceptron, which Dr. Taylor asks about, is in part a function of the level of reliability required in the perceptron's responses. To give some idea, however, a system of 1000 A-units should be able to distinguish all letters of the alphabet with a better-than-chance probability of being correct in each case (say, somewhere between 0.6 and 0.9, depending partly upon whether or not the letters always appear in the same position, whether they may be rotated, etc.). The reasons for building a hardware model of the perceptron are (1) the simulation program is slow and inflexible, particularly with respect to the variety of stimulus material which can be presented; (2) while it is easy to change parameters in the simulation program it is much harder to introduce new qualitative constraints, which we are now interested in studying. The hardware model, with patch boards to permit flexible interconnections of the units, will make this considerably easier. A third point is that we are interested in gaining some experience in the design of components which might ultimately be used in a practical system, rather than a purely research model.

I think Dr. McCarthy may have an important point in his distinction between description and discrimination. As he is using the term, *description* implies the statement of *relations among parts*, and this is something that the perceptron is quite incapable of doing, in its present form. I think Dr. McCarthy might be interested in the work of Grimsdale and associates at the University of Manchester *(ref. 3)*, who have also attempted to develop a digital computer program for the description of visual forms.

Dr. Spetner raises the question of whether $r_{\alpha\bar{\alpha}}$ is actually the best measure of similarity to use in the perceptron. I am sure that better measures of similarity could be constructed, if this were actually what we had in mind. However, we are not trying to construct a system which will

embody some predetermined similarity function, or equation. Rather, we are constructing a system which must (1) be composed of simple, neuron-type units, and (2) be connected in a relatively free, unconstrained fashion. I do not know whether such a system will really do what we would like it to, but this is precisely what we are trying to find out. $r_{\alpha\bar{\alpha}}$, therefore, is not a measure of similarity which we have deliberately set out to use, but rather an analytic property of this type of system. The concept of a response space which maps the "similarity relations" of the stimulus space is an interesting one, which is clearly basic to the whole field. It seems to me that this describes a desirable end result, towards which the perceptron is a tentative first step.

REFERENCES

1. TAYLOR, W. K., Electrical Simulation of Some Nervous System Functional Activities, in Information Theory (Third London Symposium), Ed. by Colin Cherry, *Butterworths, London*, (1955).
2. ROSENBLATT, F., The Perceptron - A theory of Statistical Separability in Cognitive Systems, *Cornell Aeronautical Laboratory Report* # VG-1196-G-1, January, 1958.
3. GRIMSDALE, R. L., SUMNER, F. H., TUNIS, C. J., and KILBURN, T. A System for the Automatic Recognition of Patterns, Proc. *I.E.E.*, 1959, 106, Part B.

SESSION 3

PAPER 5

LEARNING MACHINES

by

DR. A. M. ANDREW

BIOGRAPHICAL NOTE

Alexander M. Andrew, National Physical Laboratory Teddington. Age 33. Studied Mathematics and Natural Philosophy at Glasgow University, later doing research in Physiology. Spent 18 months with Warren McCulloch's group at M.I.T., working on visual pathways in the frog. Recent activities have swung away from biology, but this will be remedied.

LEARNING MACHINES

by

DR. A. M. ANDREW

SUMMARY

THE application of learning machines to process control is discussed. Three approaches to the design of learning machines are shown to have more in common than is immediately apparent. These are (1) based on the use of conditional probabilities, (2) suggested by the idea that biological learning is due to facilitation of synapses and (3) based on existing statistical theory dealing with the optimisation of operating conditions. Although the application of logical-type machines to process control involves formidable complexity, design principles are evolved here for a learning machine which deals with quantitative signal and depends for its operation on the computation of correlation coefficients.

---

1. INTRODUCTION

In this paper learning machines will be discussed with particular reference to industrial applications, but it is believed that the arguments have a bearing on the study of human and animal learning. The extent to which parallels between brains and machines may profitably be drawn has been discussed by many writers including MacKay *(ref. 1)* and Sluckin *(ref. 2)*, and it is clear that extreme caution is required. It does appear, however, that valuable biological experiments are likely to be suggested by a study of the difficulties encountered in attempting to make practical applications of learning machines as similar difficulties have presumably been met and solved at some stage of biological evolution.

1.1 *The definition of learning*

Thorpe *(ref. 3)* provisionally defines learning in animals as "that internal process which manifests itself as adaptive change in individual behaviour as a result of experience." By its inclusion of "adaptive" this definition

implies that the learning process is necessarily associated with goal-directed behaviour. The goal may be the acquisition of food when hungry, or drink when thirsty, or the satisfaction of an instinctive drive such as reproduction or care of the young. The animal's behaviour may be interpreted as an attempt to maximise some function of its sensory inflows. Wiener *(ref. 4)* uses the term "affective tone" to describe this function; Selfridge *(ref. 5)* refers to a similar function as "hedony".

In the case of a learning machine to control an industrial process the quantity which it recognises as "hedony" and tries to maximise will be a pre-arranged function of a number of quantities associated with the process. These will include measures of yield and quality of the product and the costs of running the process. The learning machine will keep modifying its policy of control by trial-and-error with the aim of increasing the amount of "hedony". This is, of course, essentially what a person does in acquiring skill in a control task.

1.2 *A type of industrial application*

Since chemical processes lend themselves to automation more readily than do processes involving the handling of solid objects, attention has been paid particularly to the possibility of using a learning machine to control a chemical process. Consequently it will be assumed in the following that the learning machine must deal with quantitative rather than logical information, but may, nevertheless, be able to make binary decisions and to change its own internal connections.

Figure 1. represents a process controlled by a learning machine. The quantities a, b and c may be temperatures, pressures, flow-rates or other variables associated with the process. The learning machine effects control by varying the quantities d and e, which may control valve settings or heating power applied to parts of the process. In order to compute the degree of goal-achievement or hedony, h, the output of the process is measured and probably tested for quality by suitable transducers. The inflow of raw material may also be monitored.

The arrangement of *fig. 1* is considerably different from an ordinary servo-mechanism employing negative feedback. In a servo-mechanism the quantity which indicates the degree of goal-achievement, namely the output error, can be used directly to control the active element. In the arrangement of *fig. 1*, on the other hand, it is unlikely that the quantity h can be used directly to control the process as there will usually be a considerable delay between the occurrence of a variation in d or e and the appearance of the resultant change in h.

In general the only way in which the process can be controlled is by letting d and e be controlled as functions of a, b and c. The quantity h can only enter into the control system in a more subtle way. The mathematical functions relating d and e to a, b and c must contain a number of

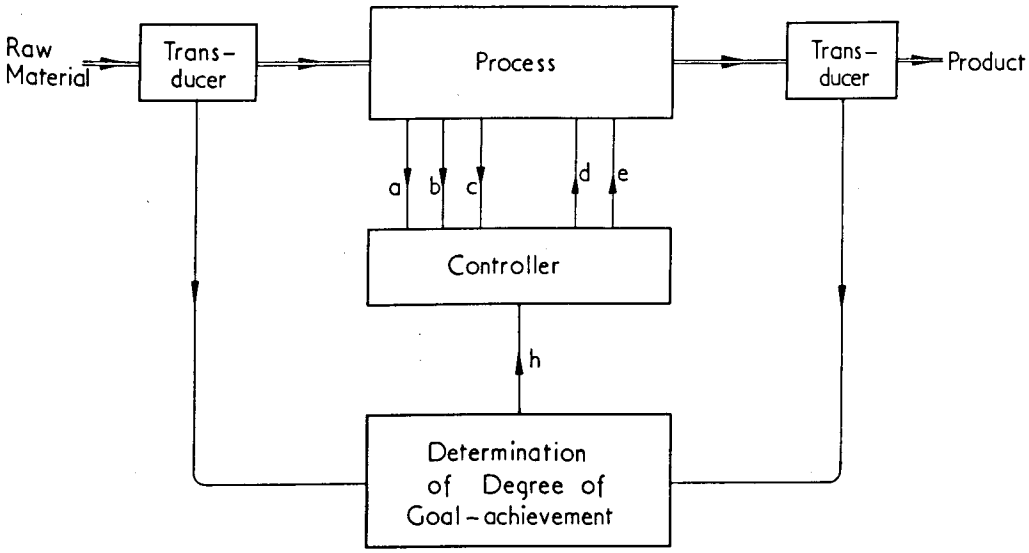


Fig. 1

parameters which can be altered in a trial-and-error fashion so as to maximise h. At any instant, therefore, the effect of a high value of h is to help set the seal of approval on the forms of the control functions which were operative before it occurred, while the effect of a low value of h is the opposite. By such trial-and-error procedures the control functions can be made to approach their optimum forms. It is perhaps worth observing at this point that the optimisation is not necessarily restricted to finding the best values of parameters in the functions; it is possible to devise optimisation procedures which can alter the function more fundamentally, as will be described later.

Although a distinction was drawn between servo-mechanisms and the type of arrangement shown in *fig. 1*, this could represent a self-optimising servo in the case where a, b and c are functions of the input and output quantities of the servo-mechanism, and h is some time-averaged monotonically-decreasing function of the servo error.

It is expected that learning machines will allow efficient automatic control of processes which are not amenable to precise mathematical analysis. The use of a learning machine is, in fact, an empirical method for approximating to the optimum control function. Other empirical methods

are possible, involving the continuous recording of all relevant quantities associated with the process and the periodic analysis of records by means of a computer. The method employing a learning machine is attractive insofar as trial-and-error adjustment can proceed continuously, and consequently the optimum control function should be approached more rapidly than by other methods. This relatively rapid approach to the optimum could be particularly valuable if the process was influenced by some fluctuating quantity which could not be monitored, but which influenced the optimum control functions.

The principle of operation of a learning machine as described in the foregoing is very similar to Ashby's principle of ultrastability, *(ref. 6)*. Three different approaches to the problem of building a learning machine will be discussed. These appear, at first sight, to have little in common but will be shown to be closely related. Then, certain of the principles which were involved in these approaches will be further elaborated to show how it is hoped to build a practical learning machine suitable for industrial use.

2. FIRST APPROACH TO THE DESIGN OF LEARNING MACHINES. THE CONDITIONAL PROBABILITY COMPUTER

The ideas of probability theory must obviously be involved in any empirical approach to process control, since the aim is to maximise the probability of the desired goal in the future. Since a certain amount of information about the process is available to the controller at any time (in the form of measures of a, b and c in *fig. 1*, for instance) the probability levels which are important are the *conditional* probabilities which are computed subject to the information being as indicated.

A computer has been built by Uttley *(refs. 7,8,9)* and Russell which is specially designed to evaluate conditional probabilities. Some further developments have been described by Russell *(ref. 10)*. This type of computer has a number of input channels which can be activated singly or in groups. The signals in the channels are necessarily in dichotomous ("yes-no") form; each channel at any instant is either activated or not activated. The channels are connected to a classification system (see Uttley *ref. 11*) of a kind which is shown in *fig. 2.* for the case of a 3-channel machine. The capital letters A, B and C have been used to represent the signals in the three channels; the convention will be adopted here of using capital letters for dichotomous quantities and small letters for continuously varying quantities.

The classification system can activate a unit for every possible group of input channels. A machine with n input channels must have $2^n - 1$ units.

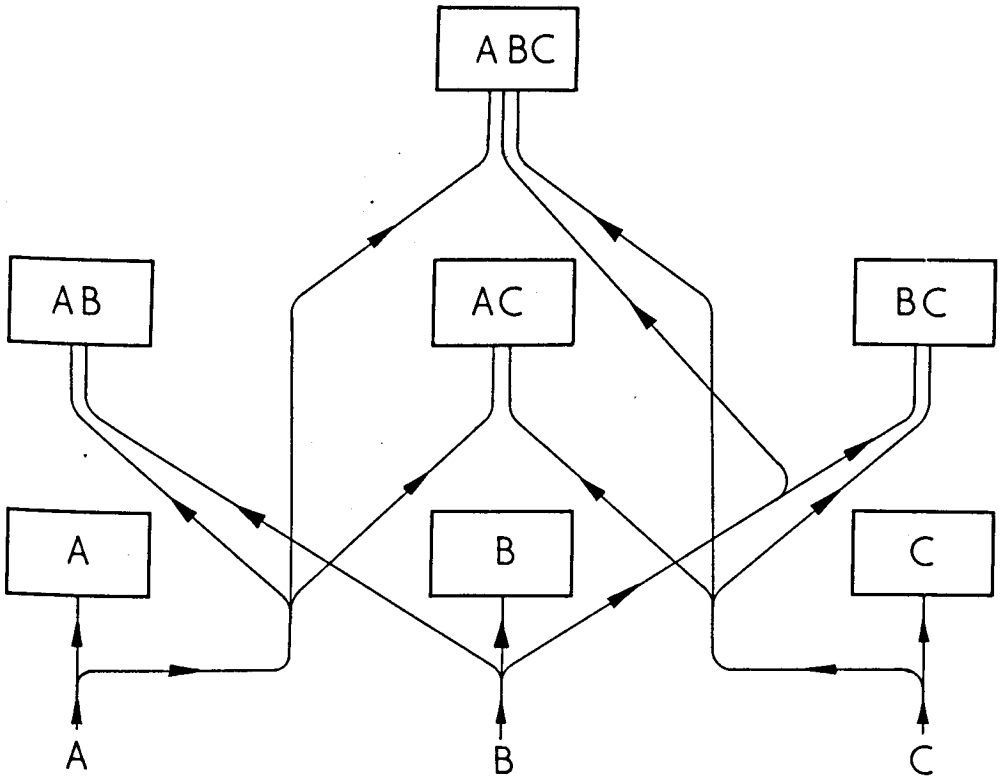


Fig. 2

The practical conditional probability computer which was built by Uttley and Russell has five input channels and consequently 31 units. This computer will be exhibited at the Symposium. In order to compute conditional probabilities, each of these units counts the number of times it is activated. In the existing computer this is done in an analogue fashion by arranging that the charge on a capacitor is altered every time the unit is activated. Then the conditional probability of activity in any particular channel, given activity in some other channel, or group of channels, can be computed as the ratio of the counts stored in a particular pair of units.

Suppose, for instance, that the A and B channels are active. Then the conditional probability of activity in C is given by:-

Probability of C, given A and B =

$$p_{AB}(C) = \frac{\text{count stored in } (ABC) \text{ unit}}{\text{count stored in } (AB) \text{ unit}} \qquad (1)$$

When the A and B inputs of the conditional probability computer are activated simultaneously, it computes and quantity $p_{AB}(C)$ and if there are D and E channels it similarly computes $p_{AB}(D)$ and $p_{AB}(E)$. The computer could have been made so as to indicate the actual values of these probabilities, but it in fact determines whether or not they exceed a predetermined threshold value. If this value is exceeded an inference is made of activity in the corresponding channel or channels C, D or E. The output signals of the computer consist, therefore, of inferences of activity in the same channels as are used to convey the input signals.

2.1 *Goal-seeking by conditional probability*

The conditional probability computer was not designed for the type of industrial application which is considered here, and it is not inherently goal-seeking. There are three main ways in which it can be modified to become goal-seeking. These are illustrated in *figs. 3,4* and *5* respectively.

In *figs. 3* and *4*, one of the input channels of the computer is connected so that it is activated when there is an indication of success from the

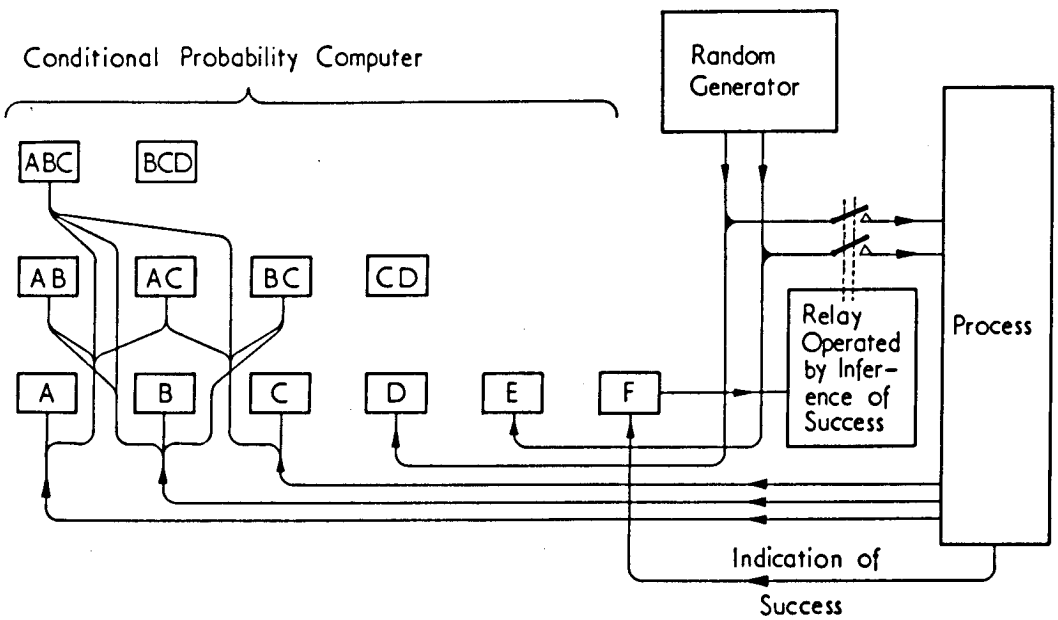


Fig.3. First Type of Goal Seeking

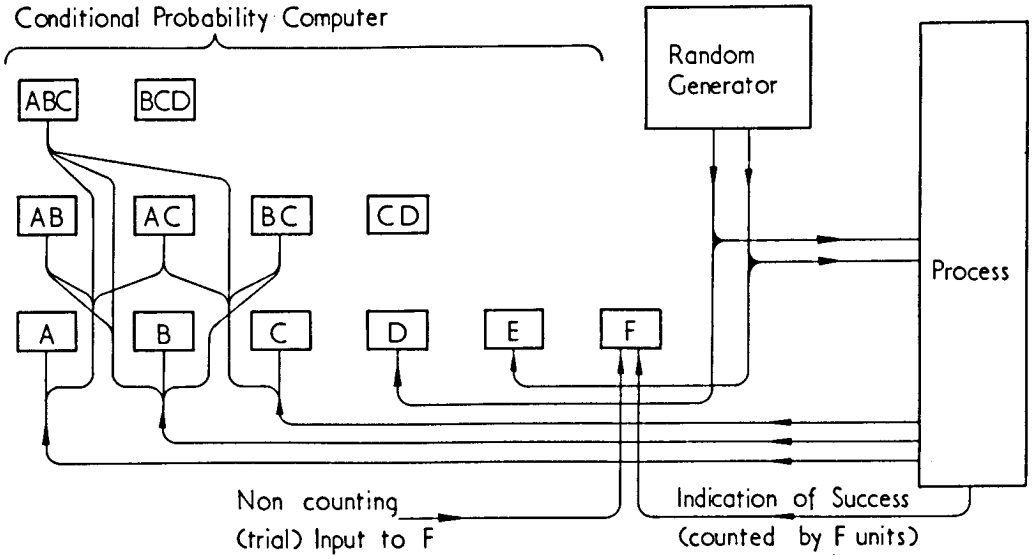


Fig.4. Second Type of Goal Seeking

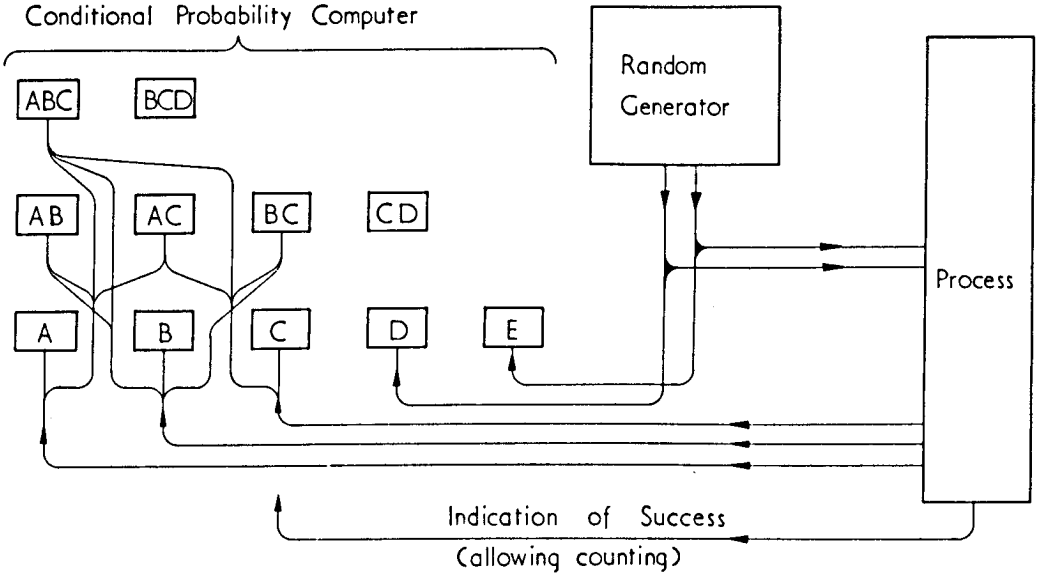


Fig.5. Third Type of Goal Seeking

process. (It is assumed in these arrangements that the indication of success is a yes-no signal, and not a continuous variable like h in *fig. 1*). In *fig. 3* the random generator applies different patterns of activity to the D and E channels, until the whole pattern of activity applied to the A, B, C, D and E channels is one which the computer associates with activity in channel F, which represents success. When F becomes active by inference, the relay closes and allows the pattern of activity presented to D and E to be applied to the process. In order that the computer can learn it is necessary that the relay closes sometimes when an inference of activity in F has not been made; it might be arranged, for instance, that if the computer does not make an inference in F within a certain time of random searching, the relay, in any case, closes.

In *fig. 4* the pattern of activity in channels D and E is actually inferred by the computer. It infers the pattern of activity in these channels which it has come to associate with the pattern existing in the A, B and C channels, and with activity (representing success) in the F channel. The input to the F channel which is used to invoke inferences is applied in such a way that it does not affect the counts on the units of the computer. If this was not done, the computer would come to associate success with all possible patterns on the A, B and C channels in combination with an absence of activity in D and E.

In *fig. 5*, the indication of success is not actually applied to a channel of the computer. Instead, it is arranged that the units of the computer are only allowed to count if there is an indication of success. By this means the computer can learn to associate favourable patterns of activity in D and E with the patterns which occur in A, B and C.

The types of goal-seeking shown in *figs. 4* and *5* (2nd and 3rd types) are essentially equivalent to one another. In *fig. 5* the indication of success allows counting in all units, while in *fig. 4* it allows counting in those units which have a connection to the F channel. But since, in *fig. 4*, it is only the units having a connection to the F channel which are important in making inferences, *figs. 4* and *5* are essentially the same.

Figure 5, however, indicates how the principle can be extended to the case where the indication of success is a continuous variable such as h in *fig. 1*. The indication of success in *fig. 5* could be such a quantity, and might even take negative as well as positive values. The amount by which the counts in the units are altered would be proportional to h.

In spite of its rather severe limitations due to having only five on-off channels to convey information to it, the existing conditional probability computer has proved remarkably versatile. When coupled to simple pieces of external apparatus it provides a number of interesting and amusing demonstrations.

3. SECOND APPROACH TO THE DESIGN OF LEARNING MACHINES. FACILITATION OF SYNAPSES

A second approach to the design of learning machines is suggested by the idea that biological learning is due to changes in the properties of synaptic connections between neurons. A learning machine might consist of a number of units capable of activating each other through some kind of linkage, where the properties of the linkages are modified as part of the learning process. There would be a variable quantity associated with each linkage and the effectiveness of the linkage in transferring the activity from one unit to another would depend on the current level of this quantity. The values of these quantities would be continuously modified as functions of the activity of the network.

MacKay *(refs. 1, 12, 13)* has discussed the possibility of building networks in which the linkages do not behave in a deterministic manner, but have a certain probability of transferring activity. In this case the probability of a transfer would be governed by the current level of the variable quantity associated with the linkage. An alternative to MacKay's scheme would be to let the occurrence or non-occurrence of a transfer depend absolutely on the current level of the quantity, according to whether it was greater or less than some threshold value.

In order that this kind of network may learn, the variable quantities associated with the linkages must be suitably modified as functions of the activity of the network. In an application where the network is required to be goal-seeking, an obvious rule for modifying these quantities would be the following:-

If a transfer of activity in a linkage is followed by an indication of goal-achievement, let the effectiveness of the linkage in transferring activity be increased. If the transfer is not followed by goal-achievement, let the effectiveness of the linkage be reduced. An extension of this rule which would probably produce faster learning would be the following:-

> If simultaneous activity in the two units which are coupled by a linkage is followed by goal-achievement, let the effectiveness of the linkage be increased and vice versa. According to this latter rule the effectiveness of the linkage is altered whether or not the simultaneous activity in the units was due to this particular linkage. (It is assumed that the linkage can transfer activity in one direction only).

Other possible rules for varying the effectiveness of the linkages might be devised. It is sometimes suggested that in the nervous system those synapses which are frequently activated become more effective. This would provide a mechanism whereby the system would learn with practice, but it is difficult to relate the principle to goal-directed activity. In any case, recent work by Professor Delisle Burns at McGill University indicates that repeated activation of a synapse has the effect of reducing its effectiveness rather than increasing it.

It has been assumed in the foregoing that a linkage either transfers activity or does not, but a network could contain units in which activity is produced by a summation of effects of a number of linkages, although any one linkage would not produce activity by itself. Such units must certainly be incorporated in any network intended to serve as a model for the nervous system. In the present treatment, however, it will be assumed for simplicity that the transfer of activity by a linkage is an all-or-nothing process.

3.1 *Conditional probability computed by linkages*

Although the conditional probability computer in its present form does not operate by altering the properties of the linkages, a computer could be designed to work in this way and to show equivalent behaviour. In the computer as previously described the storage of information is in the units themselves which count the number of times they are activated non-inferentially.

The storage can equally well be in the linkages but it is necessary in that case to have a larger number of storage locations. For a computer having n input channels, it is necessary to have about 2^n locations if the storage is in the units and $n/2$ times as many if it is in the linkages.

However in the various modifications of the computer which make it goal-seeking, many of the storage locations can be eliminated in either case. It can be shown that, for a goal-seeking computer, the number of storage locations is much the same whichever way the computer is organised.

For a non-goal-seeking conditional probability computer having storage in the linkages, every unit corresponding to a group of channels must be capable of being activated inferentially through a linkage from every unit corresponding to the same group of channels with one channel eliminated. For instance, the unit corresponding to *ACD* must be connected through linkages to the units *AC*, *CD* and *AD*. The unit corresponding to *AB* must be connected through linkages to the units *A* and *B*.

Figure 6 shows two alternative circuits which could be incorporated in the linkages to compute conditional probabilities.

For the purpose of explanation it is assumed that one of these circuits is incorporated in the linkage which can activate the *AB* unit from the *A* unit. In the circuit of *fig. 6* (a), the switch S_1 is normally in the position shown, but goes over to the right for a short time whenever the unit *A* becomes active due to activity in the *A* channel - not by inference. If the *B* unit (and hence the *AB* unit) are similarly active at the same time, the switch S_2 is in its upward position and connects the capacitor of value kC to the supply voltage V_0. On the other hand, if the *B* unit is not active, the switch S_2 connects the capacitor of value kC to earth, and it is discharged. It is clear that if event *A* is *always* accompanied by event *B*,

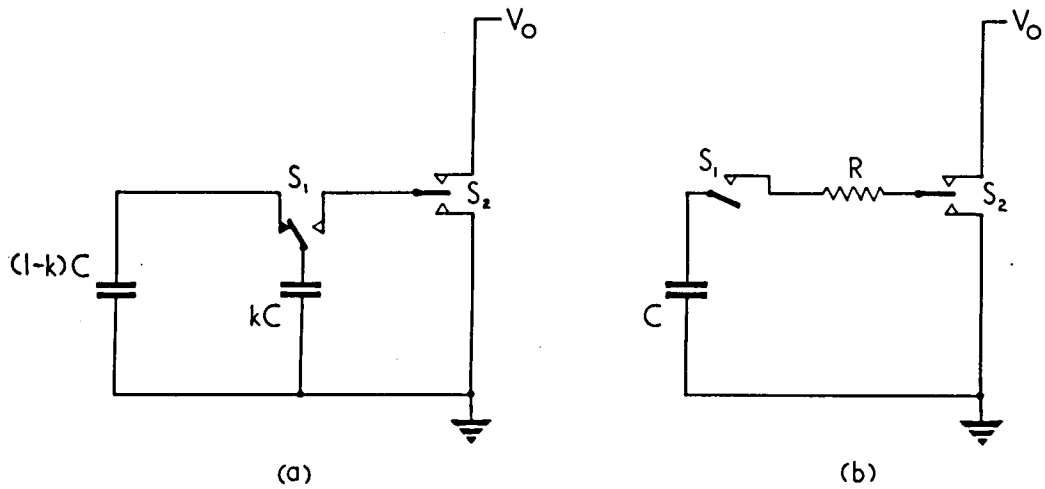


Fig.6

the voltage on the capacitor approaches V_0; if the event A is *never* accompanied by event B, the voltage approaches zero. It can be shown that if V_n is the voltage on the capacitors after the nth occurrence of activity in A,

$$\frac{Vn}{Vo} = (1-r)\left\{ \delta_n + r\delta_{n-1} + r^2\delta_{n-2} + \ldots\ldots \right\} \quad (2)$$

where $\delta_i = 1$ or 0 according to whether or not B was active on the ith occurrence of activity in A.

For the circuit of *fig.6(a)*

$$r = 1-k \quad (3)$$

and normally $k \ll 1$

The expression on the right hand side of Eq. 2 is a reasonable one to take as a "running value" of the conditional probability of B given A. Hence, $\frac{V_n}{V_0}$ as produced by the circuit of *fig.6(a)* is a suitable quantity to take as such a "running value". The behaviour of the circuit *fig.6(b)* is similar. It is required that the switch S_1 in this case closes for a definite time t_1 every time A occurs. Then if the time constant of R and

C is equal to t_2, it can be shown that the circuit of *fig. 6(b)* also conforms to Eq. 2, but in this case

$$r = \exp\left(-\frac{t_1}{t_2}\right) \tag{4}$$

In order that a computer incorporating one of these circuits may make inferences, it is only necessary that the voltage V across the capacitor to the left of the circuit is applied to a comparator which determines whether or not V/V_0 is greater than the chosen threshold value of conditional probability. If V/V_0 is greater than the threshold, the linkage produces inferences. In the case considered above, for instance, if V/V_0 is greater than the threshold, and if input A becomes active, the linkage will energise unit AB by inference (assuming input B is not active and hence unit AB is not activated directly). When unit AB is activated by inference, the computer indicates an inference of activity in channel B.

Circuits other than those in *fig. 6* can be devised to compute conditional probabilities. For an industrial application the storage element would probably not be a capacitor since this would not provide long term storage. The use of magnetic cores in an analogue fashion, as has been done by Pittman *(ref. 14)* is a possibility.

3.2 *Equivalence of approaches 1 and 2*

Since it is possible to design a conditional probability computer which depends for its operation on the modification of linkages, it is considered that approaches 1 and 2 to the design of learning machines are essentially equivalent. *Figure 7* shows a type of learning machine which could have been evolved through either approach. It is essentially a rearrangement of the 3rd type of goal-seeking shown in *fig. 5*, for the case where the conditional probability computer depends on the modification of linkages. (It does, however, have a classification system for the A, B and C inputs which differs from that in *fig. 2*. Instead of an AB unit, there is an $A.B.\sim C$ unit, active when the input pattern is "A and B and *not* C". There is no special merit in the system of *fig. 2* in the case where the data in channels A, B and C is reliable, so that it is never necessary to make inferences of A, B or C).

The random generator in *fig. 7* can either be connected to the D and E channels, corresponding to the random generator in *fig. 5*, or it can be applied to the linkages themselves making them behave like the probabilistic connections discussed by MacKay.

4. LIMITATIONS OF LOGICAL-TYPE DEVICES FOR PROCESS CONTROL

The learning machines considered in the last two sections dealt with yes-no information; the input and output channels were either active or

inactive. A computer controlling a process, however, must utilise information coming from transducers measuring temperatures, pressures, flows, etc. and it is unlikely that satisfactory control could be effected by a system which treated these measures as yes-no quantities. (The rather clumsy terms "yes-no" and "logical-type" are used here in preference to "binary" in order to avoid possible confusion with pulse-code modulation).

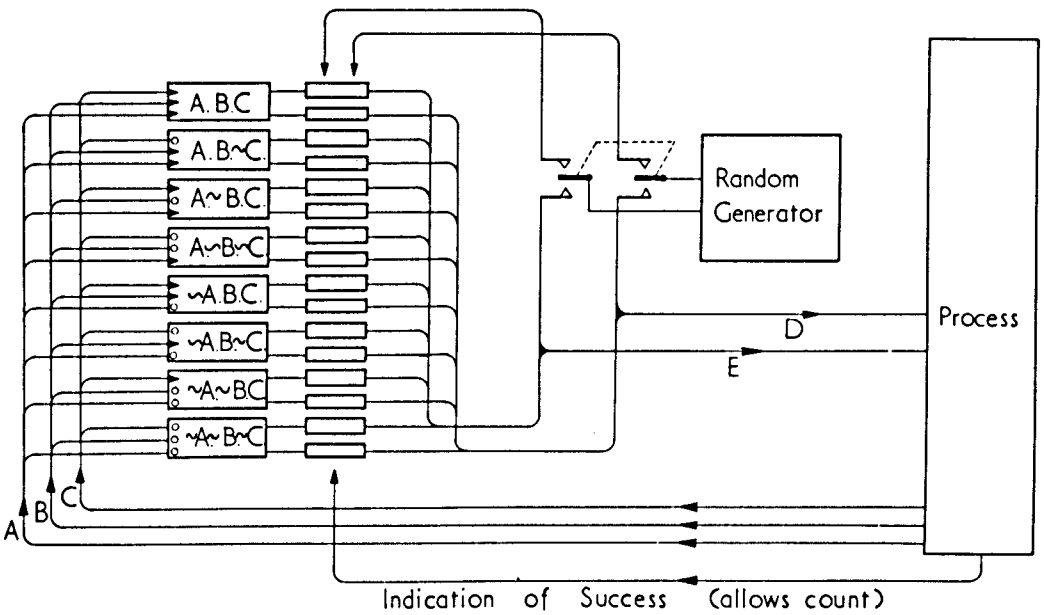


Fig.7. Modification of Third Type of Goal Seeking

A possible way of extending logical-type machines to take account of quantitative information would be to code the measured quantities in binary form and then to use several of the yes-no channels to convey the information from each transducer. In this way it is possible, in principle, to arrange for a logical-type machine to deal with any sort of input information.

Unfortunately the complexity of the learning machine which is required becomes prohibitive. It would be reasonable to represent each measured quantity by five binary digits. To control a dynamic process, the learning machine must take account of past as well as present values. In order that the machine may take account of values at two-time-displacements as well as the current value, it is necessary that each measured quantity connects with 15 yes-no input channels. Hence, the total number of yes-no channels by which information is fed to the learning machine must be 15 times the number of transducers, and is, therefore, prohibitively large.

4.1 *Achieving simplification*

There are several ways in which some simplification might be achieved. It is possible that in certain applications some of the 15 yes-no channels corresponding to an input quantity could be eliminated. It might sometimes be sufficient to represent the current value of a quantity by five binary channels, and its derivative by two, making seven in all. Simplification in this way is undesirable, however, since information thrown away can never be regained and any reduction in the amount of information supplied to the learning machine can only restrict the range of control functions which it is able to apply.

4.2 *Use of sub-goals*

Two further ways in which some simplification can be achieved are by the use of sub-goals and by interpolation. Any application of a learning controller to a process of any complexity would involve the use of sub-goals. Parts of the process would be linked to a sub-section of the controller in the general manner represented by *fig. 1* but the quantity recognised as "hedony" and maximised by this part of the controller would not be the same as the overall goal of the complete controller. In controlling a chemical process, for instance, the overall goal would be expressed in terms of the yield and quality of the product, but a sub-goal might be to maintain a particular temperature-distribution in a fractionating column. If the sub-goals were determined in advance by the designer of the controller, the use of sub-goals would be a departure from the learning principle. In fact, however, the sub-goals could be modified in a trial-and-error fashion by that part of the controller (which may by termed the "master" controller) whose goal is the overall goal of the process.

A complicated controller might actually involve a hierarchy of learning machines of which one was the "master" or first-order controller. This would have, as its goal, the overall goal of the process, and it would set the goals for the second-order controllers, some of which would, in turn, set goals for third-order controllers and so on. The use of subsidiary goals can lead to a considerable reduction in the complexity of a learning controller.

4.3 *Use of interpolation*

The other principle which could be used to reduce complexity is interpolation. It was assumed earlier that five yes-no channels are needed to convey the signal from a transducer, or fifteen to convey the current value as well as the values at two time-displacements. The system of interpolation which is proposed would let each signal be represented by a smaller number of yes-no channels, which might be two to represent a single signal, or six when time-displaced values are represented. Then the two yes-no signals can define four levels within the range of variation of the measured quantity. Suppose, for example, that the measured quantity is a temperature which can vary between 40 and 100 degrees. Then 40 degrees would be represented by 0,0 in the yes-no channels, 60 degrees by 0,1 in them, 80 degrees by 1,0 and 100 degrees by 1,1. At any instant in time it will generally be found that the temperature is not exactly 40, 60, 80 or 100 degrees. The learning machine must decide what action it would take for the nearest of these values of temperature above and below the given value. For example, if the temperature was 63 degrees, the controller would work out control actions which would be appropriate if (a) the value was 60 degrees and (b) the value was 80 degrees. Then a weighted average of (a) and (b) would be taken, with more weight given to (a) than to (b) since 63 is closer to 60 than to 80. This weighted average would be the control action exercised by the controller.

In the general case, where a number of variables are represented by the activity in two yes-no channels, a more complicated form of interpolation must be used. Let n be the number of variables, including time-displaced versions. Instead of determining what the control action would be in each of two situations, it is now necessary to determine what would be the control action in each of 2^n situations. Then a weighted average is taken of the control actions indicated for each of the 2^n situations. The weighting given to any indication would be governed by the degree of approximation of the corresponding situation to the true situation.

When interpolation is used, the process by which the learning machine adjusts the storage in its units or linkages must be modified to take account of values of the variables which are not exactly represented by the yes-no channels. For example, consider again a variable taking the value of 63 degrees, where 60 degrees is represented by 0,1 and 80 degrees by 1,0.

Then the storage locations which can be modified as a result of the activity involving this measure will include those which would have been modified had the value been 60 degrees as well as those which would have been modified if the value had been 80 degrees. The former group will be modified more than the latter, since 63 is closer to 60 than to 80.

By the use of sub-goals and interpolation, a considerable saving in complexity is possible in the application of a logical-type learning machine to process control. However, the necessary complexity is still high and it appears that a more elegant solution can be reached by departing from the logical mode of operation.

4.4 *Time taken to learn*

The reduction in complexity which is achieved by the use of sub-goals and interpolation is valuable insofar as it makes the learning machine more nearly feasible from the point of view of cost and size; it is also valuable in reducing the time taken for the machine to learn. A logical-type learning machine having, say, 30 yes-no inputs would have to learn how to behave in each of the 2^{30} situations which might arise (except for those which are physically impossible). Since 2^{30} is approximately 10^9 and since each situation must arise a number of times before the machine learns how to deal with it, it is clear that the learning process will take a long time. Even if unlimited resources were available for the construction of a learning machine, therefore, the complexity would still have to be kept within bounds in order that learning would not take too long.

4.5 *Numerical estimates of complexity*

It is instructive to calculate the number of units which are needed in a learning machine of the logical type having various numbers of yes-no input channels. In *Table 1*, let n be the number of input channels. The values shown are multiples of 6 or 15, since these are the numbers of channels required to represent one variable in the cases of operation with and without interpolation.

The second column in *Table 1* shows 2^n. This is the minimum possible number of storage locations in a learning machine of the type shown in *fig. 7* having n yes-no input channels. Actually, the number of locations could only be as small as 2^n in the case where the controller had only one yes-no output channel through which to control the process.

Since the number of storage locations tends to be large, it is desirable to consider a cheap form of storage such as magnetic tape. The quantities held in the various storage locations could be represented by binary digits recorded on tape. This would require at least 10 digits per location (see below). The number of digits is therefore 10×2^n and is shown in column 3 of *Table 1*. Column 4 shows the length of magnetic tape which would be needed assuming a packing density of 100 digits per inch.

TABLE 1

Note: 1 light year = 5.878×10^{12} miles

No. of channels n	Min. no. of locations 2^n	No. of binary digits 10×2^n	Length of magnetic tape
6	64	640	6 inches
12	4,096	40,960	34 feet
15	32,768	327,680	260 feet
24	1.6×10^7	1.6×10^8	25 miles
30	10^9	10^{10}	1,600 miles
42	4×10^{12}	4×10^{13}	6,300,000 miles
60	10^{18}	10^{19}	0.27 light-year
72	4×10^{21}	4×10^{22}	1,074 light-years
84	1.6×10^{25}	1.6×10^{26}	4,295,000 light-years

It is clear that the logical type of learning machine becomes prohibitively complex when the number of yes-no channels is more than about 20, and even for 12 channels the complexity is considerable. It is important to remember that the number of storage locations as shown in column 2 is a minimum based on the assumption that the output is by a single yes-no channel. In general the number will be considerably greater, and the figures in columns 3 and 4 would be multiplied correspondingly.

4.6 *Retrospective indication of goal-achievement*

A source of complication which has not hitherto been mentioned is the fact that it will not generally be possible to assume any particular value for the time-lag between a control action by the controller (see *fig. 1*) and the appearance of an effect in the indication of "hedony". For most processes, in fact, the effect of a control action at any instant will be spread over a period of time, and the nature of the effect may be different at different times within this period. A learning machine used to control such a process must have several quantities stored in each "storage location". In the learning machine of *fig. 7*, for instance, each of the linkages would store a number of quantities which might be (a) a quantity relating to the probability of success immediately following a control action involving the linkage, (b) a quantity relating to the probability of success 30 mins. after such an action and (c) a quantity relating to the probability of success 60 minutes after the action.

The values of 30 and 60 minutes for the lags have, of course, been chosen arbitrarily; in a practical application the values would be chosen

with regard to the time-constants of the process. It would be advantageous to store many more than three quantities and the ideal would be to store a continuous function of time. Here as elsewhere, a compromise is necessary to keep complexity within bounds.

Returning to the learning machine of *fig. 7*, the way in which learning would take place would be as follows.

A suitable memory, such as a continuous loop of magnetic tape, would be incorporated. This would contain, at any instant, a record of the input and output signals of the controller over the previous 60 minutes of operation. When a "success" signal was produced by the process, the quantity described under (a) would be modified in those linkages which had been active immediately previously, and simultaneously the quantity described under (b) would be modified in those linkages which according to the record had been active 30 mins. previously and similarly for the quantities described under (c).

The "decision" of a linkage whether or not to give an output at any instant when its input is activated must depend on a function of the three quantities listed under (a), (b) and (c). The function must be such that the linkage would become active in the case where one of the three quantities indicated that activity in the linkage was usually followed by good results in 0, 30 or 60 minutes; and the other two of the quantities made no significant indication. On the other hand the function must also be such that the linkage *would not* become active if one of the quantities indicated that activity in the linkage was consistently followed by lack of success after some interval. This latter requirement would hold, even when the other stored quantities indicated that good results were usually obtained after other time intervals.

It was because of the need to retain several quantities in each storage location that the number of binary digits to represent a storage location in digital form was taken to be as many as ten (see *Table 1*). If only one quantity was stored three or four binary digits would be sufficient.

It is clear from the foregoing that a logical-type learning machine suitable for process control is likely to be of formidable complexity. There is, however, no reason for restricting learning machines to the logical type of operation. In fact, it will be shown that machines which deal with quantitative information (by means other than converting it to binary form) will probably prove more suitable for application to process control.

Logical-type machines have been discussed at length for three reasons. These are (a) because much of the previous discussion of learning machines by other writers has been devoted to machines of this sort and (b) because it is not impossible that logical-type machines may find practical applications even though they are apparently less suitable for chemical process

control and (c) because quantitative learning machines can be regarded as a development from the logical type. The following section will be concerned with (c). It will be shown how an attempt to reduce the complexity involved in building a practical logical-type machine can lead to a scheme for a quantitative learning machine.

5. EXTENSION TO QUANTITATIVE OPERATION

Two types of classification system have been mentioned, namely the type shown in *fig. 2*, in which a given pattern of activity in the input channels can activate a number of units, and the type used in *fig. 7*, in which each unit corresponds uniquely to a particular pattern of activity. A possible way of achieving some simplification would be to omit some of the higher order units in a classification system of the type of *fig. 2*. Suppose this is carried to the limit where there is no classification at all but only direct connections to the input channels. (That is to say, all the units in *fig. 2* except those in the lowest row have been eliminated).

Suppose further that the input channels are connected through linkages to the output channels in the way that the units are connected in *fig. 7*. In this the outputs of the different linkages which affect a particular output channel are combined in an "or" fashion. In the simplified learning machine which is now proposed it is obvious that if the outputs of the linkages were combined in an "or" fashion the machine would not be able to make full use of its input information, since, as soon as the output channel was activated by activity in one input channel, it would become independent of all the other input channels. If the outputs of the linkages are summated instead of being combined in an "or" fashion, the machine can make better use of its input information. For a machine using summation, the quantity or quantities stored in the linkages would determine the magnitude of the contribution of the linkage to the summation. Thus the linkages would have a different function from that considered previously, which was to determine from the stored quantities whether or not activity should be transmitted at all. For the machine using summation, an obvious extension is to let the inputs to the units be quantitative instead of yes-no signals, and to have the output signals of the units proportional to the inputs. The ratio between the output and input signals of a linkage would be a function of the quantities stored in the linkage, which are modified by the learning process. The controlling action of a learning machine, as now proposed, would be represented by a set of equations of the form

$$d = La + Mb + Nc \qquad (5)$$

where d is an output signal, and a, b, and c are three input signals of the controller, as shown in *fig. 1*. L, M and N depend on the quantities stored in

the appropriate linkages of the learning machine. In practice a "constant term" K would also be included. K would also be modified by the learning process. Equation (5) becomes

$$d = K + La + Mb + Nc \tag{6}$$

Equation (5), of course, simply expresses d as a linear function of a, b, and c. The aim of the learning process is to adjust the values of K,L,M, and N so as to make this linear function approximate as closely as possible to the optimum control function for the process.

5.1 *The use of correlation*

The way in which the parameters K,L,M and N could be modified is by making experimental variations in their values while the process is running, and correlating these variations with the variations in the degree of goal-achievement or "hedony" represented by h in *fig. 1*. If there is a significant positive correlation between the variations of one of the parameters, say M, and the variations in h, the learning machine must make an increase in the value of M. Conversely, if significant negative correlation is observed, the machine must reduce the value of M.

5.2 *Correlation as an extension of the conditional probability principle*

This process of self-adjustment by correlation will be discussed in more detail in the next two sections. It is interesting to note that the computations carried out by the linkages in *fig. 7* are essentially correlations. The "transfer function" of each linkage (corresponding to M above) can take the value 0 or 1 in *fig. 7*. Counting in the units is allowed, only when a "success" signal is present. This means that the dichotomous "transfer function" is multiplied by the "success" indication, which is a dichotomous hedony signal, and a time-weighted average of the product is computed in the linkage. The time-weighted average of the product is essentially a correlation between transfer function and hedony. Strictly speaking, the quantity computed is not the normalised correlation coefficient defined by

$$r = \frac{\Sigma XY}{\sqrt{(\Sigma X^2 . \Sigma Y^2)}} \tag{7}$$

since in this definition the quantities X and Y must consist of fluctuations about mean values. The quantities "hedony" and "transfer function", which may be represented in their dichotomous form by H and M, do not necessarily consist of fluctuations about means.

However, the quantity computed is of the form

$$r = \Sigma H.M.W(t) \tag{8}$$

where $W(t)$ is a time-weighting function, Equation 8 essentially represents a correlation (see Licklider, *ref. 15* for a discussion of running correlation functions). Hence the use of correlation in a quantitative learning machine is a direct extension of the principles devised for logical-type learning machines.

6. THIRD APPROACH TO THE DESIGN OF LEARNING MACHINES. EVOLUTIONARY OPERATION

Any attempt to apply learning principles in industry should certainly utilise the large amount of statistical theory which has been discussed by Box under the heading of Evolutionary Operation *(ref. 16)*. The theory is described in greater detail in Chapter 11 of the book edited by Davies *(ref. 17)* and in other publications by Box and his colleagues. *(refs. 18, 19, 20, 21* and others*)*.

Box and his co-workers are concerned with the determination of optimum conditions for a process where the variables are quantities such as temperature, pressure, time of reaction, proportions of constituents and so on. *Figures 8* and *9* show two ways in which the yield (or "hedony") of a process may be plotted as a function of these variables. Box *et al* describe empirical procedures which can be used (without

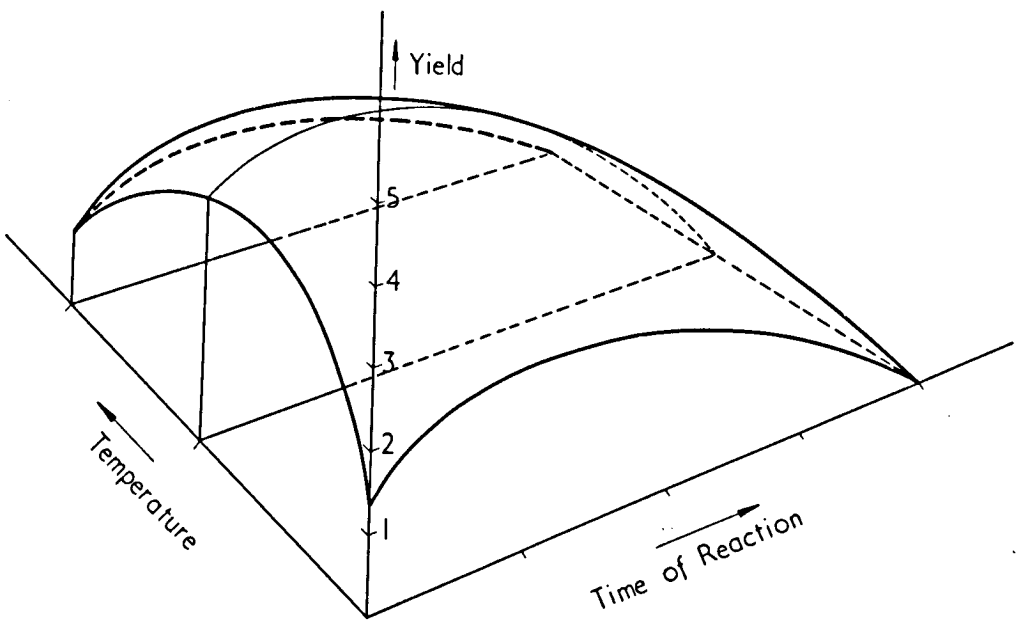


Fig. 8. Yield Surface

exploring all possible combinations of values of the variables) to arrive at those values which bring the yield to its maximum.

There seems to be no reason why the same principles should not be applied to the optimisation of a control function. Any control function may be represented to any desired accuracy by a set of polynomial equations with a sufficient number of terms. There would be one such equation for each channel through which the controller was able to affect the process. The polynomial equations would be of the same form as Equation 6, with the addition of higher order terms:-

$$d = K + La + Mb + Nc + Pa^2 + Qb^2 + Rc^2 + Sab + Tbc + Vca + Va^2b + Wabc + \ldots\ldots \quad (9)$$

where the optimisation procedure is applied to find the best values of K,L,M etc.

For simplicity, only the first order terms would be considered in a first approximation, and higher order terms would be introduced if the first approximation proved inadequate. So long as only the first order terms are considered, Equation 9 is identical with Equation 6, and the procedure for optimisation which can be derived from Box's theory is essentially that discussed in the previous section.

The application of Box's theory should make it possible to decide how, if at all, the variations in the parameters K,L,M etc. should be related to one another. One thing which is certain is that the different parameters should vary at the same time, rather than remaining constant while one at a time is varied. The realisation that factors should vary simultaneously is fundamental to the work of Box and is discussed by Fisher *(ref. 22)*.

Box arrives at procedures for conducting a set of experiments in the neighbourhood of a particular set of operating conditions, and for determining therefrom the way in which the conditions should be altered so as to approach the optimum. In *fig. 9* a particular set of operating conditions would be represented by a point P in the plane of the diagram. Suppose tests are made at each of two levels of each of the two variables. Then tests could be made under four different sets of operating conditions, which would be represented in the figure by the corners of a rectangle drawn round or near P. However, in order to determine the directions in which the operating conditions should be modified in order to approach the optimum, it is only strictly necessary to perform experiments at three of these points. Similarly, for three variables which are each tested at two levels, it is only necessary to conduct five out of the eight possible experiments. *(See ref. 17* p.507*)*.

A learning machine seeking to optimise its control function by adjustment of a number of variables might advantageously be made to relate the experimental fluctuations in the variables. If the variables were three in number it would be made to test five out of the eight possible combinations

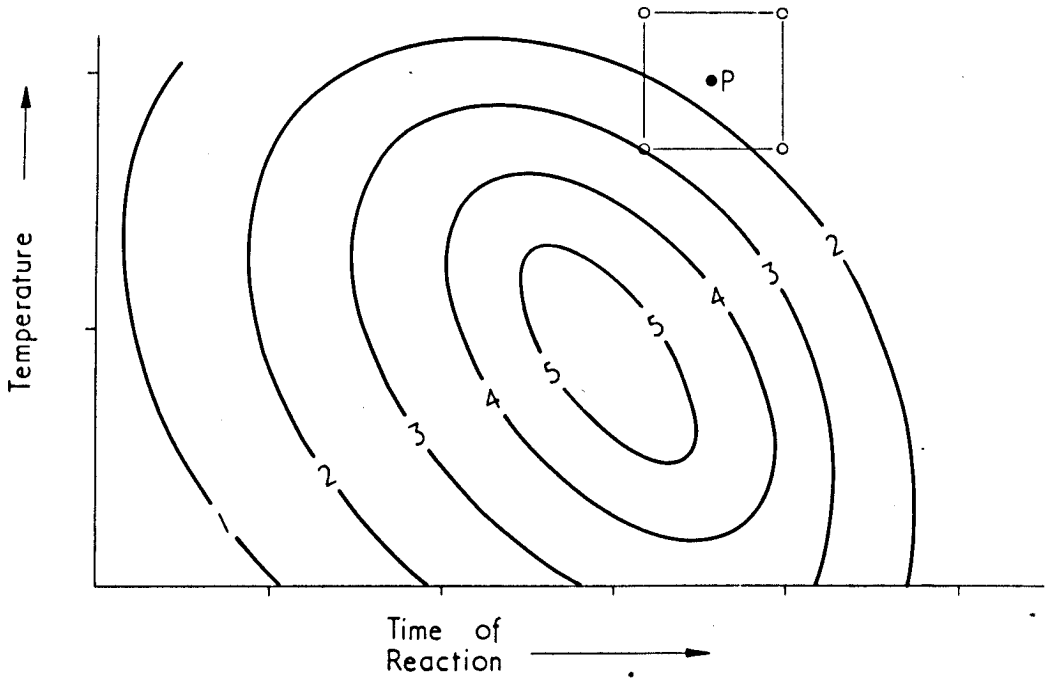


Fig. 9. Yield Contours

of signs of the fluctuations. This system would probably produce a more rapid approach to the optimum than completely random fluctuations would. This is a point which requires further investigation however. It is possible that a planned pattern of fluctuations is only advantageous when the errors of the individual observations of "hedony" are low; this is unlikely to be true in practical applications of learning machines to process control. Nevertheless it is clear that the statistical theory due to Box provides the basis for planning the experimental strategy of an industrial learning machine.

6.1 *Some other possibilities.*

A learning machine for process control could be made to operate by means other than by optimising control functions of the form of Equation 9. Other forms of empirical approximation could be used. It is also possible to devise learning procedures in which the aim is to represent the process as well as possible by mathematical expressions. From the mathematical model of the process it is straightforward to derive optimum control functions.

Kalman *(ref. 23)* has built a learning machine which controls a process by first deriving a mathematical model. The type of process he considers is one having one input channel and one output channel, but since it is a dynamic process, the controller must take account of signal levels over a range of time. Kalman represents the process by the equation

$$C_k + b_1 C_{k-1} + \ldots\ldots b_n C_{k-n} = a_0 m_k + a_1 m_{k-1} + \ldots\ldots a_q m_{k-q} \ldots \quad (10)$$

where the a_i and b_i are arbitrary constants and the b_0 has been set arbitrarily to unity. The m_i represent instantaneous values of the control effort applied by the controller to the process, at times kT, $(k-1)T$..... $(k-q)T$, where T is a fixed interval of time. The C_i represent the response of the process at times $kT, (k-1)T \ldots\ldots (k-n)T$.

Wilby and Woodcock *(ref. 24)* are building a self-optimising filter, which can adjust itself to simulate a process. This filter therefore lends itself to incorporation in a controller which, like that of Kalman, makes use of a mathematical model of the process. The type of controller proposed here does not incorporate a mathematical model of the process. The model is an intermediate step between the observations which the machine makes of the process and the control function it applies. In general, it is in the interests of simplicity to eliminate this intermediate step. There is one big advantage in having the mathematical model, however, since the learning involved in forming the model is still useful even if the goal of the learning machine is changed. In some applications of learning machines, therefore, machines incorporating a mathematical model of the process will be preferred.

7. FURTHER TRICKS WITH CORRELATION

7.1 *Amplitude of the fluctuations*

As already described, the values of the parameters K, L, M ..... are adjusted in value by making small fluctuations in their values, and correlating these with variations in the "hedony" or degree of goal-achievement, h. The obvious way of deciding the magnitude of the fluctuations is to start with very small ones and then to increase their magnitude until the value obtained for the correlation is statistically significant.

If the parameter is already at a value which gives a maximum of h, there may be no significant correlation even when the fluctuations are large. It is desirable that the machine does not continue these large fluctuations indefinitely, since they must affect the process adversely.

These large fluctuations can be avoided by computing two distinct correlations for each parameter. These are (a) the correlation of the parameter itself with hedony, as already mentioned, and (b) the correlation of

the *modulus* of the fluctuations of the parameter with hedony. The magnitude of the fluctuations should increase until (a) becomes statistically significant or (b) becomes significant with negative sign.

7.2 *Self-reorganisation*

There are further correlations which may profitably be computed in the machine. Since the correlation between a parameter and the hedony appears as a continuous or "running" value, it is possible to compute a correlation between this correlation and some other variable. We had in Equation 6:-

$$d = K + La + Mb + Nc$$

Let r_{Lh} be the running value of the correlation between L and h when L fluctuates. Then if a significant correlation exists between r_{Lh} and a, the conclusion can be drawn by the machine that the equation for d could profitably include a term in a^2. Similarly, if there is a significant correlation between r_{Lh} and b, the equation could profitably include a term in ab, and so on.

With the addition of these terms, the equation becomes

$$d = K + La + Mb + Nc + Pa^2 + Qab \qquad (11)$$

If now there is a significant correlation between r_{Qh} and c, for instance, the machine may add a term in abc, and so on.

A learning machine may therefore incorporate self-reorganisation. It incorporates a fixed number of units, and may decide by the above procedure how these should be allocated to the computation of higher-order terms. It is also possible to devise criteria which the machine may apply in order to decide whether an already-allocated unit is serving a useful purpose. If it is not, it can be re-allocated. (One possible criterion to decide whether a unit is functioning usefully is to examine whether the fluctuations in its parameter, subject to the conditions discussed earlier, become large enough to involve changes in sign of the parameter in the course of the fluctuations. If they do, the unit is serving no useful purpose).

The development of the circuitry needed to allocate units to different jobs presents interesting problems. Circuits involving relays and uniselectors can be used in much the same way as they are used by telephone engineers in allocating telephone channels to subscribers. A type of static circuit which may be suitable has recently been described by Chapman and Freed *(ref. 25)* as a possible model for neuronal connections in the spinal cord.

7.3 *Composite goals*

In general, a process must be controlled so as to achieve a composite goal, or in other words to work towards a number of goals which must be

balanced against one another in importance. Many human and animal activities similarly involve multiple goals; for example a man driving a car from *A* to *B* has the goal of avoiding an accident, but also has the goal of getting from *A* to *B* as quickly as possible. The two goals conflict to some extent and the driver's conduct is determined by some sort of estimate of their relative importance.

In an industrial application, the main goals will usually be (a) to maximise the quantity of the product, (b) to keep the quality within specification and (c) to minimise wear and tear of the equipment. It will not usually be difficult to devise a function of these goals which may appropriately be taken as the "hedony", since all industry has one main aim which is the maximisation of profits.

In this paper it has hitherto been assumed that the learning machine has no indication of the degree of goal-achievement except the indication of the value of "hedony". It has been assumed that the learning process depends entirely on correlations with this function. However, if no account is taken of the different components of the hedony, information which would appear to be significant is being ignored. It seems clear that the learning process would approach the optimum function more rapidly if the machine was kept informed, not only of its degree of goal-achievent, but also of the precise way in which it has failed to achieve the goal.

It therefore appears that the machine should make use of correlations with the quantities which contribute to the hedony function, but the writer has not yet arrived at a scheme for doing this. This is one of the many aspects of the development of learning machines which require further consideration.

8. LEARNING BY WATCHING

A learning machine applied to an industrial process could not be coupled to the process while in a completely naive state. If its initial control actions were entirely random the consequences for the process might be disastrous. In many cases it would be possible to pre-set the controller so that it controls the process in a safe and reasonable fashion before any learning occurs. Then the learning process can operate to bring the control function nearer to the optimum.

As an alternative to pre-setting the controller manually, it might be preferable in some cases to couple the learning machine to the process while the process is under manual control. In this way the machine could, in effect, learn by watching the human operator. After a sufficient period of learning by watching the machine could be put in control.

A learning machine may be made to learn by watching in either of two ways. It may simply learn to imitate the control policy of the human

operator. In this case the machine takes no account of the degree of goal-achievement of the process during the period of watching. The other possibility is that the machine observes the variations which occur in the operator's policy of control, and tries to correlate them with the variations in degree of goal-achievement. Then, when the machine is put in control its policy of control can be more consistently favourable than was the policy of the operator.

8.1. *Imitative learning*

The simpler form of learning by watching can be achieved by a straightforward adaptation of existing statistical theory. A linear approximation (of the form of Eq. 6) to the control policy used by an operator in controlling d as a function of a, b and c is given by the equation given in statistical texts for the regression of d on a, b and c. (see for example, Weatherburn, *ref. 25)*.

The equation is:-

$$\begin{vmatrix} \frac{d}{\sigma_d} & \frac{a}{\sigma_a} & \frac{b}{\sigma_c} & \frac{c}{\sigma_c} \\ r_{da} & 1 & r_{ba} & r_{ca} \\ r_{db} & r_{ab} & 1 & r_{cb} \\ r_{dc} & r_{ac} & r_{bc} & 1 \end{vmatrix} = o \qquad (12)$$

where σ_i is the standard deviation of i, and r_{ij} is the correlation coefficient of i and j.

Since, to control a dynamic process, the measured variables (a, b, c in Eq. 12) must include past as well as present values, the order of determinant required in the regression equation will generally be inconveniently high. It is unlikely that a learning machine would incorporate the means of expanding the determinant, but it would certainly incorporate the means of computing the correlation coefficients and standard deviations. The conversion of the determinantal equation into the form of Eq. 6 would be done once and for all when the learning machine was changed from learning by watching to controlling. The computation involved in the conversion would probably be carried out manually or by a digital computer.

8.2 *Learning by watching with regard to goal-achievement*

For the more advanced form of learning by watching it is necessary that the regression equation be continuously computed and expanded into the form of Eqn. 6. Then the fluctuations of the coefficients in the resulting continuously-varying equation would be correlated with the fluctuations in goal-achievement. The mean values of these coefficients would also be

computed. The values of coefficients which would be tried by the machine when it was put in control would be these mean values plus modifications. The signs of the modifications would be determined by the computed values of correlation between the goal-achievement and the fluctuations of the coefficients during the period of learning by watching.

9. SOME PRACTICAL DIFFICULTIES

General principles have been discussed for learning machines for industrial applications, and it is hoped that these will lead to practical forms. Readers who have followed the discussion this far will have realised, or at least suspected, that many practical difficulties have been omitted from the discussion. One of these is the difficulty of devising suitable functions to be taken as "running values" of correlation etc.

Equation 2 gives a suitable function to represent a running value of conditional probability. There is the difficulty here, however, that the time-constant of the averaging process (controlled by the value of r in Equation 2) must be arbitrarily chosen. The use of different values of r will affect the "temperament" of the machine.

The difficulty is greater when a more complicated function must be computed as a running value. In some cases it will be necessary to utilise running values of certain variables in order to compute the value of another variable. In this case a number of different time-constants are involved, and their ratios must be suitably chosen.

In computing a running value of correlation coefficient, the quantities which are multiplied must be fluctuations about mean values, and then the product is smoothed with a certain time-constant. The mean values are themselves running values obtained by smoothing the original signals with a certain time-constant. In what has been written here it has been assumed throughout that the correlation coefficients are normalised so as to lie between -1 and $+1$. The need for normalisation complicates the issue still further.

In fact, the computation of running values is too involved a subject to be discussed fully here. It will form the subject of a separate publication.

10 CONCLUSIONS

The writer hopes to apply the principles of learning by correlation to an industrial application. The general principles have been discussed here but there are many aspects which require further consideration. There are also many avenues of speculation which have been left unexplored.

Although reference has been made to several other workers, the present paper has not reviewed anything like all of the literature on learning machines. The emphasis has been on quantitative rather than logical-type machines because it appears that the quantitative type will find application more readily. However, it is very likely that logical-type machines will also find applications. Among the logical-type learning machines which have not been discussed are the computer programmes demonstrated by Oettinger *(ref. 27)* and the model of animal learning described by Deutsch *(ref. 28)*. Another type of computer programme with learning properties has recently been described by Friedberg. *(ref. 29)*.

Although little has been said about human and animal learning, it is felt that the principles discussed here are highly significant for psychologists. Their significance does not arise from any suggestion that the learning machines are models for animal learning unless in an extremely general way. The usefulness of the present discussion to psychology is likely to be in suggesting hypotheses and experiments rather than explanations.

I am indebted to the Director of the National Physical Laboratory for permission to publish this paper, to Dr. Uttley and to my other colleagues for valuable discussions relating to the subject matter. The responsibility for the opinions expressed, is, however, entirely my own.

REFERENCES

1. MacKAY, D. M. Towards an information-flow model of human behaviour. *Brit. J. Psychol.*, 1956, **47**, 30.
2. SLUCKIN, W. Minds and Machines. *Penguin, London* (1954).
3. THORPE, W. H. The concepts of learning and their relation to those of instinct. *Symposium of the Society for Experimental Biology*, 1950, no. 4, p. 387.
   also Learning and Instinct in Animals. *Methuen, London*, (1956).
4. WIENER, N. Cybernetics, p. 150 *Wiley, New York*, (1948).
5. SELFRIDGE, O. G. Pattern recognition and learning. Paper no. 33 in Information Theory, ed. E. C. Cherry, *Butterworth, London*, (1956).
6. ASHBY, W. R. Design for a Brain. *Chapman and Hall, London*, (1952).
7. UTTLEY, A. M. The conditional probability of signals in the nervous system. *R.R.E. Memo, no. 1109*, 1955. Radar Research Establishment, Malvern, Worcs.
8. UTTLEY, A. M. A theory of the mechanism of learning based on the computation of conditional probabilities. *Proc. of the 1st International Congress of Cybernetics, Namur*, (1956).
9. UTTLEY, A. M. The design of conditional probability computers. *Information and Control*. In the press.

10. RUSSELL, G. Interim report on learning machines and adaptive control systems. *R.R.E. Memo. no. 1369.* 1957.
11. UTTLEY, A. M. The classification of signals in the nervous system. *E.E.G. Clin. Neurophysiol.*, 1954, **6**, 479.
12. MacKAY, D. M. The epistemological problem for automata. Automata Studies, p.235 ed. by C. E. Shannon and J. McCarthy, *Princeton*, (1955).
13. MacKAY, D. M. Mindlike behaviour in artefacts. *Brit. J. Philosophy of Science*, 1951, **2**, 105.
14. PITTMAN, G. F. A high-accuracy static time delay device using transistors. *A.I.E.E. Trans. Pt.1*, March 1955, p. 54.
see also Schaeffer, D. H. Magnetic core event counter for earth satellite memory. *Elect. Eng.*, 1958, **77**, 52.
15. LICKLIDER, J. C. R. A duplex theory of pitch perception. *Experientia*, 1951, **7**, 128.
see also SAYERS, B. M. and CHERRY, E. C. Mechanism of binaural fusion in the hearing of speech. *J.Acous. Soc. Amer.*, 1957, **29**, 973.
16. BOX, G. E. P. Evolutionary operation: a method for increasing industrial productivity. *Applied Statistics*, 1957, **4**, 81.
17. DAVIES, O. L., ed. Design and Analysis of Industrial Experiments. *Oliver and Boyd, Edinburgh*, (1956).
18. BOX, G. E. P. and WILSON, K. B. On the experimental attainment of optimum conditions. *J. Roy. Statistical Soc.*, 1951, **13**, series B, p.1.
19. BOX, G. E. P. The exploration and exploitation of response surfaces: some general considerations and examples. *Biometrics*, 1954, **10**, 16.
20. BOX, G. E. P. and YOULE, P. V. The exploration and exploitation of response surfaces: an example of the link between the fitted surface and the basic mechanism of the system. *Biometrics*, 1955, **11**, 287.
21. BOX, G. E. P. and COUTIE, G. A. Application of digital computers in the exploration of functional relationships. *Proc. I.E.E.* 1956, **103**, pt. B, supp. 1.
22. FISHER, R. A. The Design of Experiments. *Oliver and Boyd, Edinburgh*, 3rd. Edition, (1942) p.88.
23. KALMAN, R. E. Design of self-optimising control system. *Trans. A.S.M.E.*, (Feb. 1958,) p.468.
24. WILBY, W. P. L. and WOODCOCK, R. Self optimising filter. *Imperial College, London.* First progress report, July, 1957, second April, 1958.
25. CHAPMAN, B. L. M. and FREED, G. A. An electronic analogue of nervous activity. Unpublished paper.
26. WEATHERBURN, C. E. A First Course in Mathematical Statistics. p.247, *Cambridge University Press*, (1952).
27. OETTINGER, A. G. Programming a digital computer to learn. *Phil. Mag.*, 1952, **43**, 1243.
28. DEUTSCH, J. A. A new type of behaviour theory. *Brit. J. Psychol.*, 1953, **44**, 304.
29. FRIEDBERG, R. M. A learning machine. *I.B.M. Journal*, 1958, **2**, 2.

APPENDIX TO THE PAPER BY DR. A. M. ANDREW ϕ

One of the main objects of the paper is to point out the need for some form of interpolation in any learning machine applied to process control. Two forms of interpolation are discussed; one under the heading "interpolation" in section 4.3, and one in terms of polynomial functions in part 5. Without some form of interpolation the learning machine cannot exploit the continuity of its environment.

The principles discussed in the latter part of the paper lead to a number of interesting possibilities not mentioned in the paper. One of these is the possibility of making machines which will "learn to learn". The design of a learning machine involves the choice of values for a number of parameters, including time-constants of smoothing for the running values of statistical quantities, and threshold levels at which these quantities produce a change in the control policy of the machine. A person designing the machine would need to choose these values rather arbitrarily, since there is no simple way of determining the values which will produce most rapid learning. In fact, the design of a learning machine is precisely the kind of intractible mathematical problem which it is hoped to by-pass by the use of learning machines. Eventually, therefore, it is likely that machines will be devised in which the learning process itself will be subject to adaptive modifications.

Two of the references require further comment. In referring to Wilby and Woodcock *(ref. 24)*, the writer should have realised that these workers are under the direction of Professor Gabor, and are putting into practical form the self-optimising filter proposed by Prof. Gabor *(ref. 1)*.

A much abbreviated form of the paper by Chapman and Freed *(ref. 25)* has now been published *(ref. 2)*.

REFERENCES

1. GABOR, D. Communication Theory and Cybernetics. *Trans. I.R.E.*, 1954, **CT-1**, No. 4.
2. CHAPMAN, B. L. M. and FREED, G. A. Random channel selectors. *Automation Progress*, 1958, **3**, 430.

ϕ Added after the Symposium.

DISCUSSION ON THE PAPER BY DR. A. M. ANDREW

MR. C. STRACHEY: I find myself in some difficulty at the prospect of having to cram into five minutes all the comments I want to make on Dr. Andrew's paper. First of all, I think it is fair to say that there are two or perhaps three types of learning programme (or machine) which can be envisaged. There is the sort of self-organising type, such as the perceptron, which you hope is ultimately going to organise itself into a suitable configuration. Then there is the type which might be called the 'thinking' machine, which is sophisticated to start with and has some of its concepts already built in, as with the geometrical problem-solving machine. I find both these very interesting indeed, but neither of them at the moment show signs of having any practical application. There is a third type of learning machine, which is really not so much a learning machine as an optimum seeking machine, which is specialised, in that you decide beforehand what sort of thing you want it to learn in considerable detail. Machines of this sort have probably less claim to be called learning machines, but more claim to be called useful. They have been simulated on digital machines for a number of years, and are generally regarded as being uninteresting from the point of view of programming.

I think Dr. Andrew has confused the issue considerably by trying to apply to a practical problem a machine of the self-organising perceptron type. I do not think this is an appropriate thing to do. The first two-thirds of the paper refers to logical type conditional probability computers, and finally comes to the conclusion, with which I shall agree, that they are not much good for the type of thing he is proposing for them; the tables of the size of storage required in any practical problem certainly makes it difficult to imagine constructing a machine of this sort.

I have some further evidence which I should like to put before you to reinforce this point of view; it is not concerned with the amount of equipment but with the way in which it functions. The last third of the paper is concerned with a totally different proposal, which I think might probably stand more chance of success. But it seems to me there is nothing in this part of the paper which is other than purely speculative. It seems to me a great pity that, when you are talking about learning machines, you should start to describe them as applied to complicated physical problems such as controlling a chemical works before you have

made some practical investigation, or simulated investigation, on digital computers which you can make so very easily for these systems. I think in any practical control problem you have to make use of all the specialised information you can get hold of, though I'm afraid I have not time to expand that at the moment.

I just want to say that, after reading this paper, I was stimulated - perhaps I should say pricked - into writing a programme for a digital machine which simulates on a slightly larger scale, the machine described in figure 7. This is supposed to be a learning machine of the same type as the conditional probability computer. The machine I envisaged has two decimal digit inputs and two decimal digit outputs: you can set it to add, subtract or multiply or to give the least significant digit of the sum or the product. What you ask it to do depends on the indication of success which is given by a separate part of the routine, which is called the teacher. The way it worked was such that if it got a bad answer the probability of that one was depressed, and if it got a right answer it went on getting it right. Now the remarkable thing about this particular routine is that it behaves considerably worse than you would expect; it behaves worse than at random. At first it was only looking for the least significant digit in the sum, and before it got all the answers correct there were 649 wrong ones. On the average it had tried six and a half times for each successful answer, which is worse than if you went straight through, say, from 1 to 9. This is not unreasonable because obviously sometimes you try the same wrong numbers by chance several times.

DR. L. M. SPETNER: In this paper Dr. Andrew has indicated that it seemed quite hopeless to use the idea of logical design or digital techniques in his machine because of the large number of boxes that would be required to handle each one of his functions. The situation might, however, be little more optimistic than this, in that one might take advantage of a self-organizing type of device, such as the perceptron, in doing this job. Here the self-organization might consist, not of reinforcing the values of the association cells, but perhaps adjusting the thresholds. Adjustment of the thresholds seems to be a reasonable thing to do when one is trying to effect a logical network. The reason for this is that one may consider a threshold device with many binary inputs as a generalized logical network. For example, a threshold device with two inputs that can have the value of 0 or 1 and has a threshold of 1 is an either-or device, and so on. So that if one has N inputs into such a unit then there are N different thresholds that may be used. This allows the device to be any one of the N different symetrical logical functions. There is of course a sticky problem as to how one tells whether he should raise or lower any particular threshold to effect some given response.

This might be done by nutating the threshold during the computation of the responses, and from this nutation determine which way to change the threshold in order to improve the response.

I hope that these comments will stimulate further thought along these lines.

DR. A. M. ANDREW (in reply): I disagree with Mr. Strachey in his assumption that the third type of learning machine he mentions - the optimum-seeking machine - must necessarily be such that the sort of thing it is to learn is predetermined in considerable detail. Section 7.2 of the paper is intended to indicate how the principles of these machines can be extended to incorporate self-reorganisation. The use of sub-goals which are subject to adaptive modification, as mentioned in Section 4.2, introduces still further flexibility of operation.

I apologise for not adequately reviewing previous work in self-optimising systems. A large amount has been published, including a review of the field by Aseltine, Mancini and Sarture (*ref. 1*), and details of the Westinghouse OPCON by Burt and Van Nice (*ref. 2*). The work of Draper and Li (*ref. 3*) is particularly noteworthy.

It is difficult to comment on Mr. Strachey's experiment in programming without having further details. In any case, the problem to which he has applied the computer is rather different from the process-control type of application which I have been considering.

I agree that computer simulation of learning machines can be extremely valuable. One of my colleagues is already using a digital computer in this way, and I am starting to prepare a programme simulating a learning machine which applies the control function shown as Equation 6 of the paper. The construction of such a programme becomes a rather more complicated matter than Mr. Strachey suggests if it is designed to make a worthwhile contribution to the derivation of design principles for learning machines.

The kind of logical network suggested by Dr. Spetner does not, so far as I can see, allow interpolation any more readily than do other logical-type devices. Without this feature the machine must take a long time to learn to perform any task involving continuous variables, even if its complexity can be kept within limits of feasibility.

REFERENCES

1. ASELTINE, J. A., MANCINI, A. R., and SARTURE, C. W. A Survey of Adaptive Control Systems. *I.R.E. Trans*, 1958, vol. PGAC-6, 102.
2. BURT, D. A. and VAN NICE, R. I. Optimizing Control Systems for the Process Industries. *Technical Publication, Westinghouse Electric Corporation*, Feb. 1959.
3. DRAPER, C. S. and LI, Y. T. Principles of Optimalizing Control Systems and an Application to the Internal Combustion Engine. *A.S.M.E. Publication*, Sept. 1951.

SESSION 3

PAPER 6

PANDEMONIUM: A PARADIGM FOR LEARNING

by

DR. O. G. SELFRIDGE

BIOGRAPHICAL NOTE

Oliver G. Selfridge was born in London 10 May 1926. He studied at the Massachusetts Institute of Technology from 1942-1945, returning post-graduately from 1946-1950. After 2 years at Signal Corps Laboratories at Fort Monmouth, New Jersey, he joined Lincoln Laboratories in Group 34, Communication Techniques, of which he is now Group Leader.

PANDEMONIUM: A PARADIGM FOR LEARNING

by

DR. O. G. SELFRIDGE

INTRODUCTION

WE are proposing here a model of a process which we claim can adaptively improve itself to handle certain pattern recognition problems which cannot be adequately specified in advance. Such problems are usual when trying to build a machine to imitate any one of a very large class of human data processing techniques. A speech typewriter is a good example of something that very many people have been trying unsuccessfully to build for some time.

We do not suggest that we have proposed a model which can learn to typewrite from merely hearing speech. Pandemonium does not, however, seem on paper to have the same kinds of inherent restrictions or inflexibility that many previous proposals have had. The basic motif behind our model is the notion of parallel processing. This is suggested on two grounds: first, it is often easier to handle data in a parallel manner, and, indeed, it is usually the more "natural" manner to handle it in; and, secondly, it is easier to modify an assembly of quasi-independent modules than a machine all of whose parts interact immediately and in a complex way.

We are not going to apologize for a frequent use of anthropomorphic or biomorphic terminology. They seem to be useful words to describe our notions.

What we are describing is a process, or, rather, a model of a process. We shall not describe all the reasons that led to its particular formulation, but we shall give some reasons for hoping that it does in fact possess the flexibility and adaptability that we ascribe to it.

THE PROBLEM ENVIRONMENT FOR LEARNING

Pandemonium is a model which we hope can learn to recognize patterns which have not been specified. We mean that in the following sense: we present to the model examples of patterns taken from some set of them, each time informing the model which pattern we had just presented. Then, after some time the model guesses correctly which pattern has just been presented before

we inform it. For a large class of pattern recognition ensembles there has never existed any adequate written or statable description of the distinctions between the patterns. The only requirement we can place on our model is that we want it to behave in the same way that men observably behave in. In an absolute sense this is a very unsatisfactory definition of any task, but it may be apparent that it is the way in which most tasks are defined for most men. Lucky is he whose job can be exactly specified in words without any ambiguity or necessary inferences. The example we shall illustrate in some detail is translating from manually keyed Morse code into, say, typewritten messages. Now it is true that when one learns Morse code one learns that a dash should be exactly three times the length of a dot and so on, but it turns out that this is really mostly irrelevant. What matters is only what the vast army of people who use Morse code and with whom one is going to have to communicate understand and practise when they use it. It turns out that this is nearly always very different from school book Morse.

In the same way the only adequate definition of the pattern of a spoken word, or one hand-written, must be in terms of the consensus of the people who are using it.

We use the term pattern recognition in a broad sense to include not only that data processing by which images are assigned to one or another pattern in some set of patterns, but also the processes by which the patterns and data processing are developed by the organism or machine; we generally call this latter "learning".

PANDEMONIUM, IDEALIZED AND PRACTICAL

We first construct an idealized pandemonium *(fig. 1)*. Each possible pattern of the set, represented by a demon in a box, computes his similarity with the image simultaneously on view to all of them, and gives an output depending monotonically on that similarity. The decision demon on top makes a choice of that pattern belonging to the demon whose output was the largest.*

---

* This is an exact correlate of a communications system wherein given a received message $M(T)$ and a number of possible transmitted messages $M_i(T)$, that M_i is chosen, that is, deemed to have been transmitted, which minimizes $\int |M(T)-M_i(T)|^2 dT$. (Such a procedure is optimum under certain conditions). This integral is, as it were, the square of a distance in a signal phase space - *fig.2* - and thus that transmitted message is selected that is most similar to the received one.

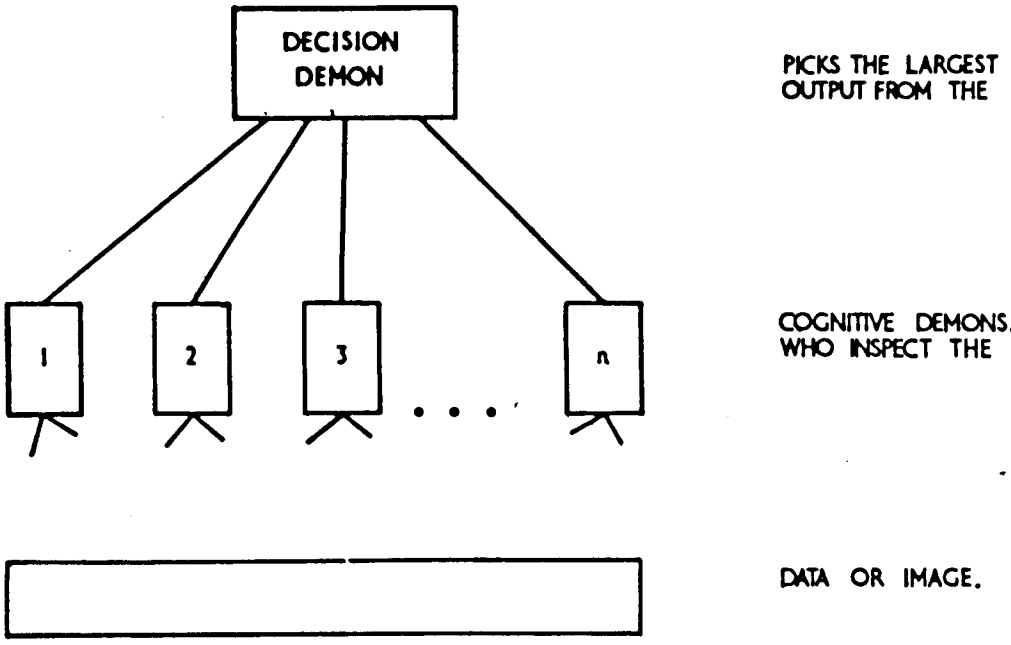


Fig. 1. Idealized Pandemonium.

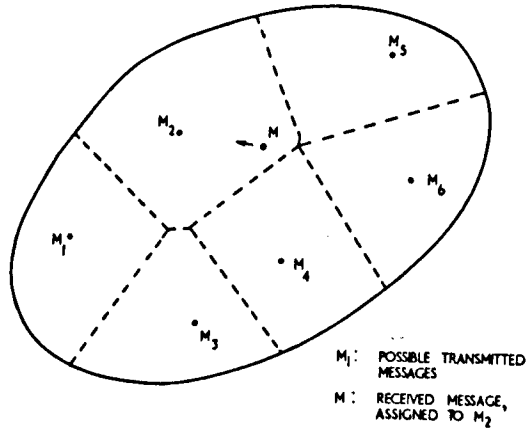


Fig. 2. Signal Space.

Each demon may, for example, be assigned one letter of the alphabet, so that the task of the A-demon is to shout as loud of the amount of 'A-ness' that he sees in the image.* Now it will usually happen that with a reasonable collection of categories - like the letters of the alphabet - the computations performed by each of these ideal cognitive demons will be largely the same. In many instances a pattern is nearly equivalent to some logical function of a set of features, each of which is individually common to perhaps several patterns and whose absence is also common to several other patterns.†

We therefor amend our idealized Pandemonium. The amended version - *fig. 3* - has some profound advantages, chief among which is its susceptibility to that kind of adaptive self-improvement that I call learning.

The difference between *fig. 1* and *fig. 3* is that the common parts of the computations that each cognitive demon carries out in *fig. 1* have in *fig. 3* been assigned instead to a host of subdemons. At this stage the organization has four levels. At the bottom the data demons serve merely to store and pass on the data. At the next level the computational demons or subdemons perform certain more or less complicated computations on the data and pass the results of these up to the next level, the cognitive demons who weigh the evidence, as it were. Each cognitive demon computes a shriek, and from all the shrieks the highest level demon of all, the decision demon, merely selects the loudest.

THE CONCEPTION OF PANDEMONIUM

We cannot ab initio know the ideal construction of our Pandemonium. We try to assure that it contains the seeds of self-improvement. Of the four levels in *fig. 3*, all but the third, the subdemons, which compute, are specified by the task. For the third level, therefore, we collect a large number of possible useful functions, eliminating a priori only those which could not conceivably be relevant, and make a reasonable selection of the others, being bound by economy and space. We then guess reasonable weights for them. The behaviour at this point may even be acceptably good, but usually it must be improved by means of the adaptive changes we are about to discuss.

---

* It is possible also to phrase it so that the A-demon is computing the distance in some phase of the image from some ideal A; it seems to me unnecessarily platonic to postulate the existence of 'ideal' representatives of patterns, and, indeed, there are often good reasons for not doing so.

† See, for example, Jerome Bruner, "A study of Thinking",

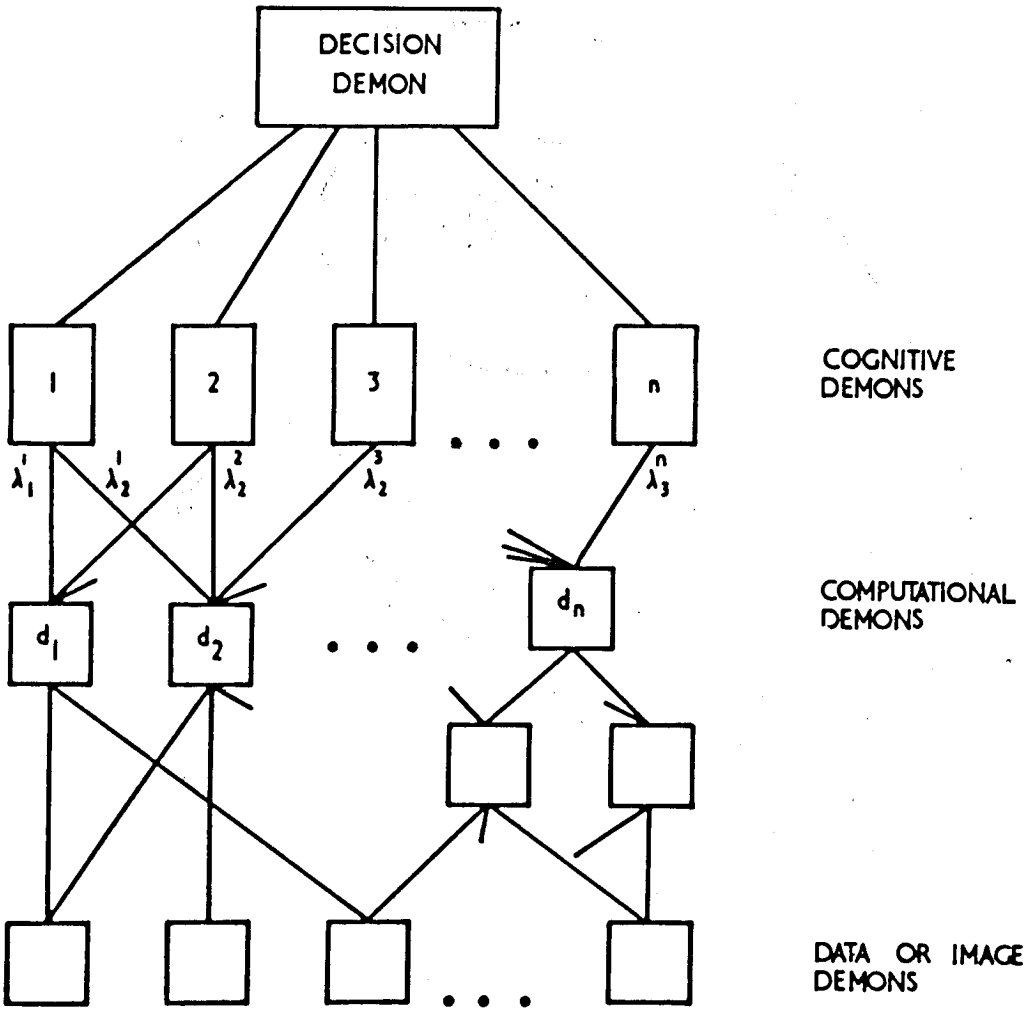


Fig.3. Amended Pandemonium

The Evolution of Pandemonium

There are several kinds of adaptive changes which we will discuss for our γ Pandemonium. They are all essentially very similar, but they may be programmed and discussed separately. The first may be called "Feature Weighting".

Although we have not yet specified what the cognitive demons compute, the sole task at present is to add a weighted sum of the outputs of all the computational demons or subdemons; the weightings will of course differ for

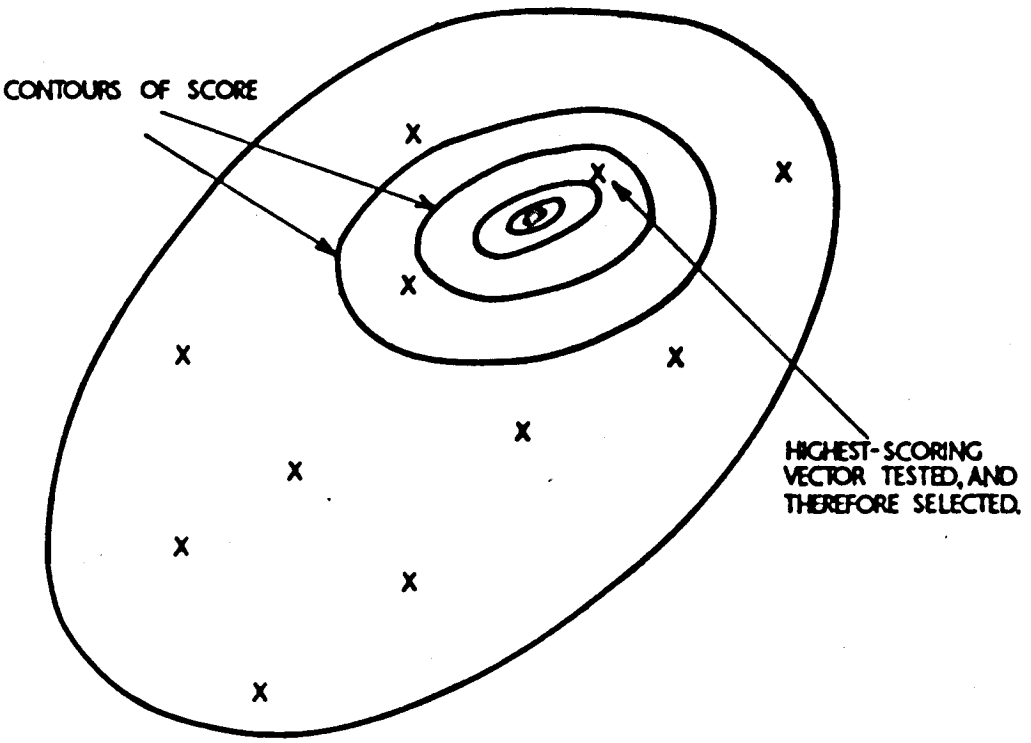


Fig.4. First hill climbing technique: pick vectors at random (points in the space), score them, and select the one that scores highest.

the different cognitive demons, but the weightings will be the only difference between them. Feature weighting consists of altering the weights assigned to the subdemons by the cognitive demons so as to maximize the score of the entire γ Pandemonium. How then can we do this?

The Score

What we mean by the score here is how well the machine is doing the task we want it to do. This presumes that we are monitoring the machine and telling it when it makes an error and so on, and for the rest of the discussion we shall be assuming that we have available some such running score. Now at some point we shall be very interested in having the machine run without that kind of direct supervision, and the question naturally arises whether the machine can meaningfully monitor its own performance. We answer that question tentatively yes, but delay discussing it till a later section.

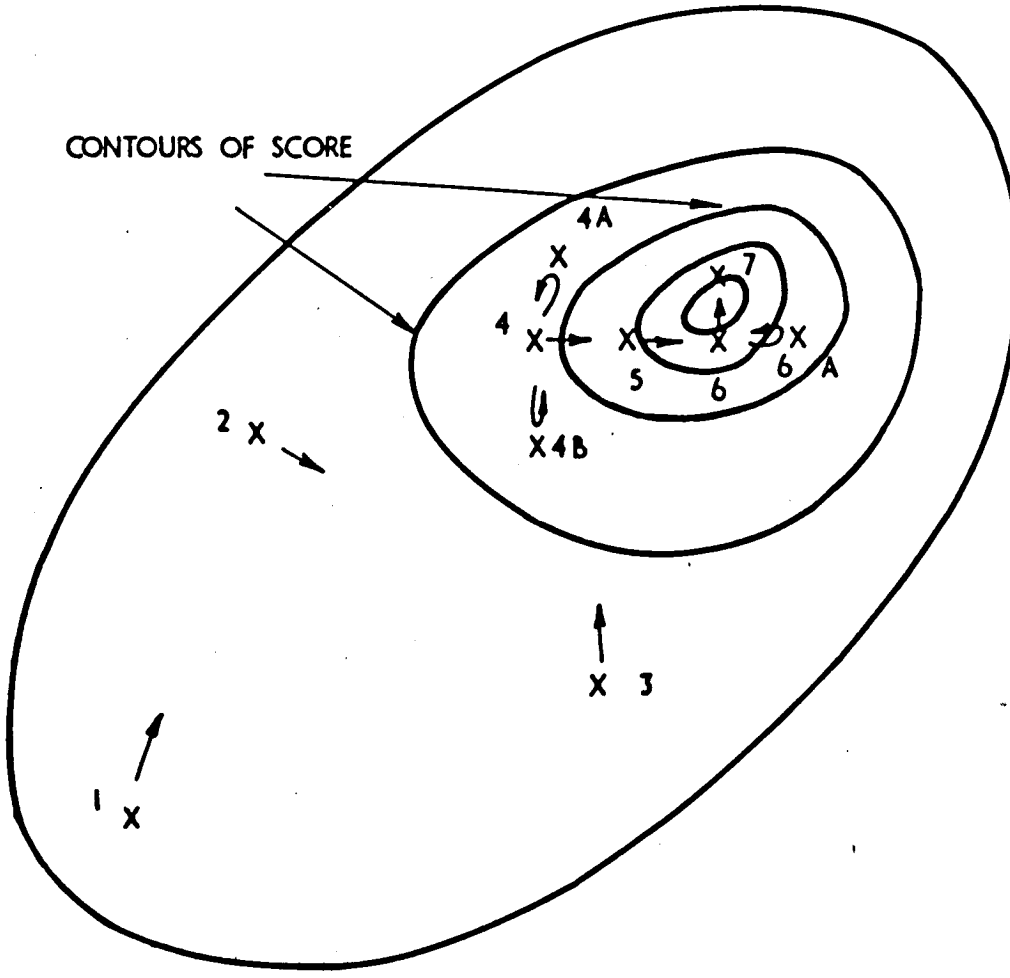


Fig.5. Second hill-climbing technique: pick vectors until one of them (Number 4) outscores the previous ones. Then take short random steps, retracing any that decrease the score.

Feature Weighting and Hill-Climbing

The output of any cognitive demon is

$$d_i = \Sigma \lambda^{i}_{j} d_j$$

so that the complete set of weights for all the cognitive demons over all the subdemons is a vector:

$$\Lambda = \left\{ \Lambda^i \right\} = \left\{ \lambda^1_1, \lambda^1_2 .. \lambda^2_1, \ldots \lambda^n_m \right\}$$

Now for some (unknown) set of weights Λ, the behaviour of this whole Pandemonium is optimum, and the problem of feature weighting is to find

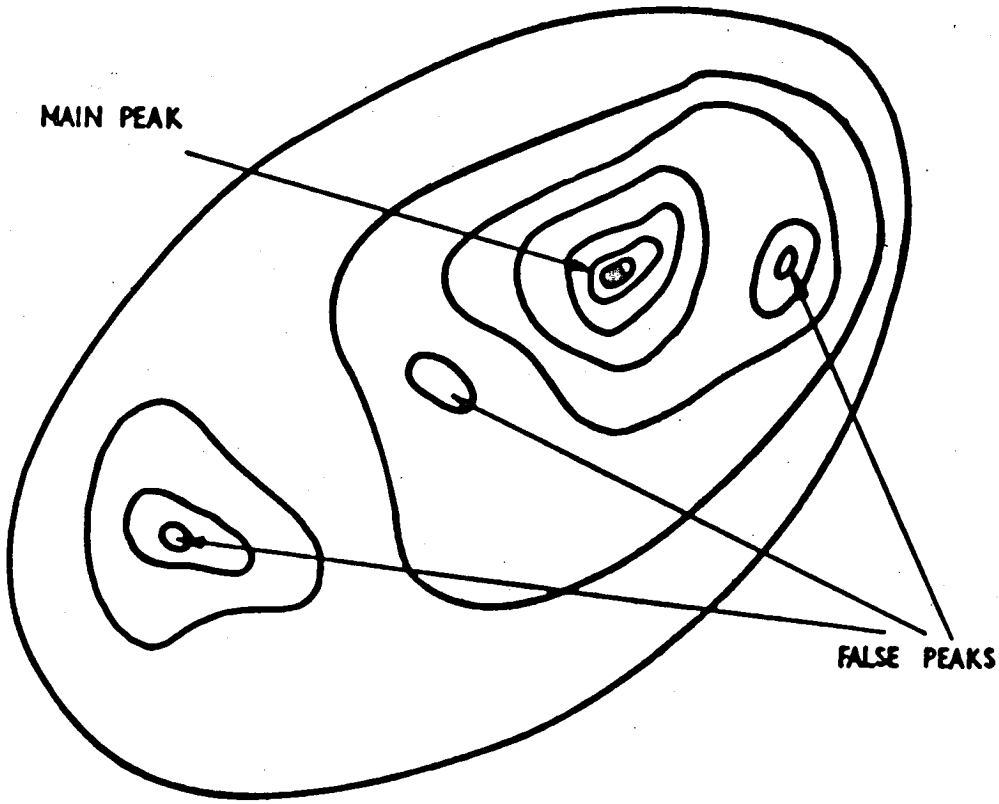


Fig.6. General space showing false peaks. One of the false peaks is quite isolated from the main or true peak.

that set. This may be described as a hill-climbing problem. We have a space (of Λ) and a function on the space (the score), which we may consider an altitude, and which we wish to maximize by a proper search through Λ. One possible technique is to select weighting vectors at random, score them, and finally to select the vector that scored highest (see *fig. 4*). It will usually, however, turn out to be profitable to take advantage of the continuity properties of the space, which usually exist in some sense, in the following way: select vectors at random until you find one that scores perceptibly more than the others. From this point take small random steps in all directions (that is, add small random vectors) until you find a direction that improves your score. When you find such a step, take it and repeat the process. This is illustrated in *fig. 5*, and is the case of a blind man trying to climb a hill. There may be, of course, many false peaks on which one may find oneself trapped in such a procedure (*fig. 6*).

The problem of false peaks in searching techniques is an old and familiar one. In general, one may hope that in spaces of very high dimensionality the interdependence of the components and the score is so great as to make very unlikely the existence of false peaks completely isolated from the main or true peak. It must be realized, however, that this is a purely experimental question that has to be answered separately for every hill-climbing situation. It does turn out in hill-climbing situations that the choice of starting point is often very important. The main peak may be very prominent, but unless it has wide-spread foot-hills it may take a very long time before we ever begin to gain altitude.

This may be described as one of the problems of training, namely, to encourage the machine or organism to get enough on the foot-hills so that small changes in his parameters will produce noticeable improvement in his altitude or score. One can describe learning situations where most of the difficulty of the task lies in finding any way of improving one's score, such as learning to ride a unicycle, where it takes longer to stay on for a second than it does to improve that one second to a minute; and others where it is easy to do a little well and very hard to do very well, such as learning to play chess. It is also true that often the main peak is a plateau rather than an isolated spike. That is to say, optimal behaviour of the mechanism, once reached, may be rather insensitive to the change of some of the parameters.

Subdemon Selection

The second kind of adaptive change that we wish to incorporate into our Pandemonium is subdemon selection. At the conception of our demoniac assembly we collected somewhat arbitrarily a large number of subdemons which we guessed would be useful and assigned them weights also arbitrarily. The first adaptive change, feature weighting, optimized these weights, but we have no assurance at all that the particular subdemons we selected are good ones. Subdemon selection generates new subdemons for trial and eliminates inefficient ones, that, is, ones that do not much help improve the score.

We propose to do this initially by two different techniques, which may be called "mutated fission" and "conjugation". The first point to note is that it is possible to assign a worth to each of the subdemons. It may be done in several ways, and we may, for example, write the worth W_i of the ith demon

$$W_i = \sum_j |\lambda_i^j|,$$

so that the worthy demons are those whose outputs are likely to affect most strongly the decisions made.

We assume that feature weighting has already run so long that the behaviour of the machine has been approximately optimized, and that scores and worths of machine and its demons have been obtained. First we eliminate those subdemons with low worths. Next we generate new subdemons by mutating the survivors and reweighting the assembly. At present we plan to pick one subdemon and alter some of his parameters more or less at random. This will usually require that we reduce the subdemon himself to some canonical form so that the random changes we insert do not have the effect of rewriting the program so that it will not run or so that it will enter a closed loop without any hope of getting out of it.*

Besides mutated fission, we are proposing another method of subdemon improvement called "conjugation". Our purpose here is two-fold: first to provide a logical variety in the functions computed by the subdemons, and, secondly, to provide length and complexity in them.

What we do is this: given two 'useful' subdemons, we generate a new subdemon which is the continuous analogue of one of the ten nontrivial binary two-variable functions on them. For example, the product of two subdemon outputs, corresponding to the logical product, would suggest the simultaneous presence of two features. The ten non-trivial such functions are listed in *Table 1*.

$A.B$	$A \vee B$
$A.\sim B$	$A \vee \sim B$
$\sim A.\sim B$	$\sim A \vee \sim B$
$\sim A.B$	$\sim A \vee \sim B$
$A.B \vee \sim A.\sim B$	$A.\sim B \vee \sim A.B$

Table 1. Non-trivial binary functions on two variables.

* We are at present running our Pandemonium on an IBM 704. The analogues for kinds of simulation are obvious.

Control Adaptation

The first two levels of adaptation are directly concerned with immediate improvement of behaviour and the score. We should also like to improve the entire organization, and in the same way. We shall deal with this point somewhat cursorily, being reluctant to specify things too far in advance of experiment. In principle, we propose that the control operationsshould themselves be handled by demons subject to changes like feature weighting and subdemon selection. It is obviously a little more difficult and perhaps impossible here to define the usefulness or worth of a particular demon. It is also clear that it will sometimes take much longer to check the usefulness of some change in some control demon - for example, in one of those which control the mutations in subdemon selection. Furthermore, at this level, some of the demons, presumably, will be in a position to change themselves, for otherwise we should need another level of possible change, and so on. This raises the possibility of irreversible changes, and it is not obvious that *all* parts of the machine should be subject to adaptive change. But these are largely heuristic questions.

The Evolutionary Process

The adaptive changes mentioned above will tend, we hope, to promote a kind of evolution in our Pandemonium. The scheme sketched is really a natural selection on the processing demons. If they serve a useful function they survive, and perhaps are even the source for other subdemons who are themselves judged on their merits.

It is perfectly reasonable to conceive of this taking place on a broader scale - and in fact it takes place almost inevitably. Therefore, instead of having but one Pandemonium we might have some crowd of them, all fairly similarly constructed, and employ natural selection on the crowd of them. Eliminate the relatively poor and encourage the rest to generate new machines in their own images.

Unsupervised Operation

So far all of the operation of the machine has been on the basis of constant monitoring by a person who is telling the machine when it makes an error. A very valid question is whether the machine can form any independent opinion of its own on how well it is doing. I suggest that it can in the following way: one criterion of correct decisions will be that they are fairly unequivocal, that is, that there is one and only one cognitive demon whose output far outshines the rest. Some running average of the degree to which this is so would presumably somewhat reflect the score of the machine. Note that it would be vital that the machine be trained first to do well enough before it is left to its own resources and supervision.

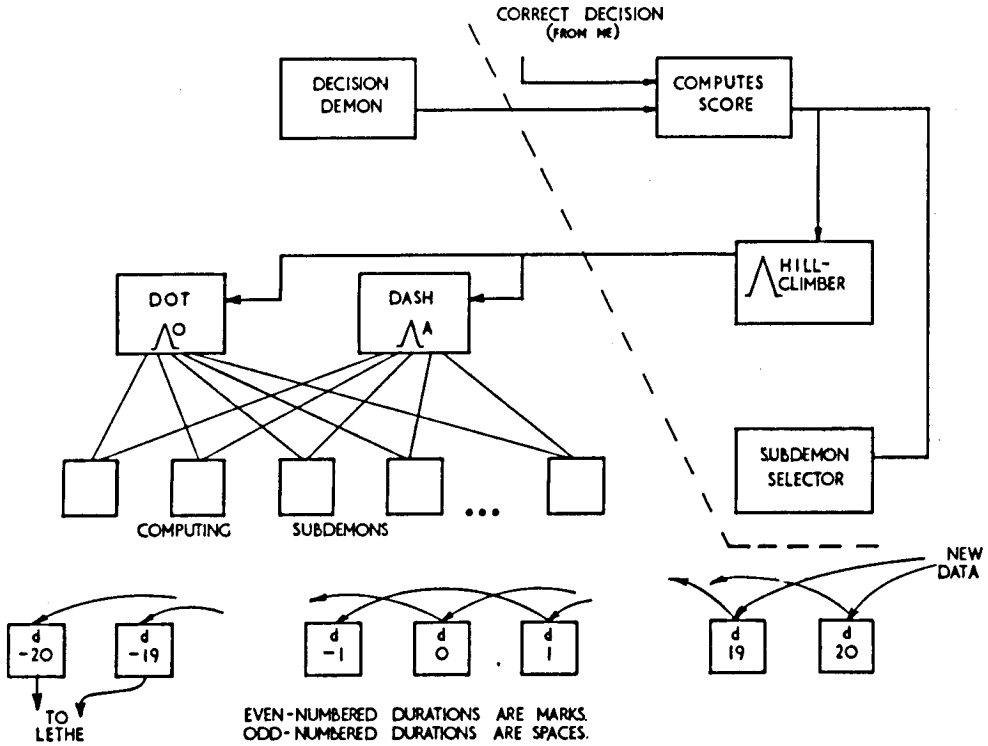


Fig.7. First morse pandemonium.

A REAL-LIFE EXAMPLE: MORSE TRANSLATION

As we mentioned before, the entire notion of Pandemonium was conceived as a practical way of automatically improving data-processing for pattern recognition. Our initial model task is to distinguish dots and dashes in manually keyed Morse code, so that our Pandemonium can be illustrated in *fig. 7*. Note that the functions and behaviour of all demons have been specified except for the computing subdemons. We shall reiterate those specifications.

(1) The decision demon's output is 'dot' or 'dash' according as the dot demon's output is greater or less than the dash demon's.

(2) The cognitive demons, dot and dash, each compute a weighted sum of the outputs of some 150 computing subdemons. Initial weights we have assigned arbitrarily, but, we hope, reasonably.

(3) The data-handling demons receive data in the form of durations, alternatively of marks and spaces, and they pass them down the line.

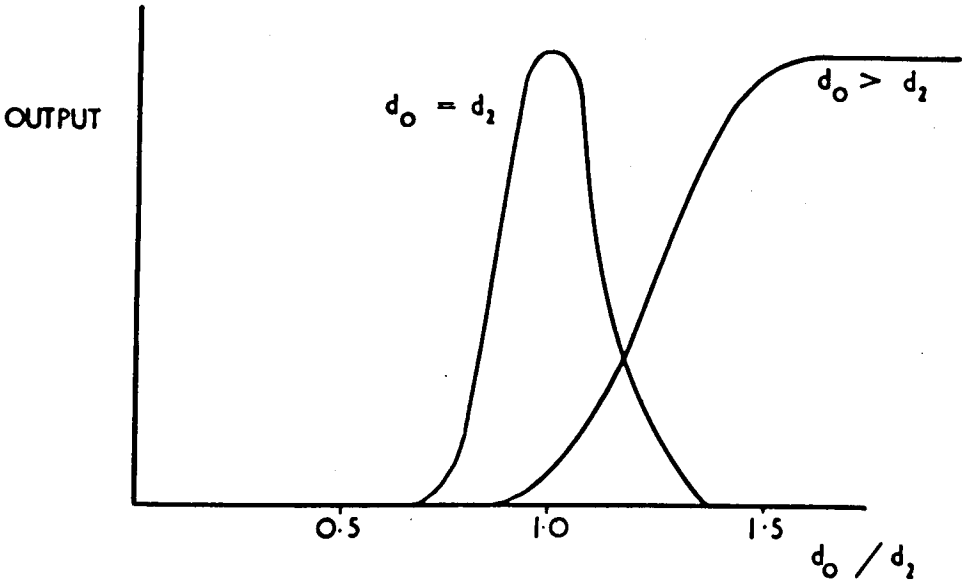


Fig.8. Operation of the subdemons $d_0 = d_2$, $d_0 > d_2$.

The computing subdemons are constructed from only a very few operational functions, which are carefully non-binary. For example the subdemons $d_0=d_2$ and $d_0>d_2$ have their outputs shown in *fig.8*. The operational functions follow:

(1) '='. This function computes the degree of equality of some set of variables (see *fig.8*).

(2) ' ', ' ', compute the degree to which some variable is less than or is greater than some other variable (see *fig.8*).

(3) '*max*', '*rmax*', compute the degree to which some variable is the largest of an arbitrary set of variables or an arbitrary set of consecutive variables.

(4) 'O_i', 'A_i', store the degree to which the ith duration *has been* identified as a dot or dash.

(5) '*Av*" computes an average of some set of variables.

(6) '*M*' is a family of tracking means. For example, it might compute

$$M^{\alpha}(c) = \alpha\, M(i-1) + (1-\alpha)$$

(7) '*Ox*', '*Ax*'. *Ox1* is the last duration identified as a dot. *Ax3* is the third last duration identified as a dash, etc.

The above is the present functional vocabulary of the computing operations for our subdemons. The subdemons themselves are built with a simple syntax. For the initial set, at conception, we merely select a set of operational functions and follow them with the numbers of some particular data demons.

CONCLUSION

What I shall present at the meeting in November will be the details of the progress of Pandemonium on the Morse code translation problem. The initial problem we have given the machine is to distinguish dots and dashes. When the behaviour of the machine has improved itself to the point where little further improvement seems to be occurring, we shall add three more cognitive demons, the symbol space, the letter space, and the word space. Presumably after some further time this new Pandemonium will settle down to some unimprovable state. Then we shall replace the senior or decision-making demon with a row of some forty or so character demons with a new decision-making demon above them, letting the new cognitive demons for the character demons use all the inferior demons, cognitive and otherwise, for their inputs. It is probably also desirable that previous decisions be available for present decisions, so that a couple of new functional operations might be added. There need be little concern about logical circularity, because we have no requirement for logical consistency, but are merely seeking agreeable Morse translation.

How much of the whole program will have been run and tested by November I cannot be sure of. At the present (July) we have had some fair testing of hill-climbing procedures.

ACKNOWLEDGEMENT

I should like to acknowledge the valuable contributions of discussions with many of my friends, including especially M. Minsky, U. Neisser, F. Frick and J. Lettvin.

DISCUSSION ON THE PAPER BY DR. O. G. SELFRIDGE

DR. J. W. BRAY: May I ask Dr. Selfridge what he thinks of this approach to his problem? Let x = duration of the last signal, which may be a dot or a dash, a signal space, letter space or word space. Let y = 2, if in fact it was a dash, 1 if dot, 0 if signal space, -1 if letter space and -2 if word space.
To form the polynomial:

$$y = A_0 + A_1x + A_2x^2 + A_3x^3 \ldots\ldots\ldots\ldots$$

take a number of observations, as he suggests, and let the machine learn the code by determining the coefficients A_0 etc. by simple curvilinear regression. The duration of previous and subsequent signals and the interpretation given to previous signals could be added as further variables on the right hand side.

MR. C. STRACHEY: At the end of the paper you promise us some further information about the latest state of the programme. Could you let us know what this is?

DR. J. McCARTHY: I would like to speak briefly about some of the advantages of the pandemonium model as an actual model of conscious behaviour. In observing a brain, one should make a distinction between that aspect of the behaviour which is available consciously, and those behaviours, no doubt equally important, but which proceed unconsciously. If one conceives of the brain as a pandemonium - a collection of demons - perhaps what is going on within the demons can be regarded as the unconscious part of thought, and what the demons are publicly shouting for each other to hear, as the conscious part of thought.

DR. D. M. MACKAY: Dr. Selfridge's 'pandemonium' has a certain family resemblance to a class of mechanism considered in some earlier papers *(ref. 1)* (though I had never suspected its demonic implications!).

In one of these, *(ref. 2)* after discussing the general principle that you penalise the unsuccessful, I pointed out that the amount of information

1. MACKAY, D. M. Mentality in Machines. *Proc. Arist. Soc. Suppt.*, 1952, 61.
2. MACKAY, D. M. The Epistemological Problem for Automata, ed. J. McCarthy and C. E. Shannan *Automata Studies, Princeton* (1955). See also *Brit. J. Psychol.*, 1956, **47**, 30 and *Advancement of Science*, 1956, 392.

per 'kick' (to use Dr. Selfridge's metaphor) is very small unless the probability of success and failure are equal. If you have a system where there are a vast number of possibilities to be eliminated by rather feckless trial and error, then of course failure occurs much more often than success. The solution I suggested was to form a kind of syndicated learning process in which at first large numbers of elements, destined eventually for independence, should be coupled together so as to reduce the diversity of response. A hand, for example, might not at first have each finger separately controllable, but could work clumsily as a whole. In that way, you can greatly decrease the amount of groping which is necessary before a successful adaptive action is found, although of course the degree of success achieved may be less. On this principle, if pursued to the limit, even the earliest trials could have a non-negligible chance of success, (though success in a very small way).

Given this easy start, simple self-organising subroutines can build up fairly quickly. As they increase their number and their success, however, the idea is that the couplings between elements should gradually be dissolved to increase the complexity of the problem. If you keep the complexity increasing step by step with the degree of development of successful internal matching sub-routines, then fully adaptive behaviour can be enormously more quickly developed than if the system starts with the full repertoire to be explored. My question is why Dr. Selfridge has not incorporated this principle in his 'pandemonium', so that each 'kick' could have something nearer to one bit of information instead of an almost negligible fraction.

DR. P. C. PRICE (written contribution): This is a very interesting and stimulating paper. I have one comment to make, and that is on the discussion of "Feature Weighting and Hill-Climbing".

I think that in this discussion the author has not brought out one important distinction between types of "hill-climbing" problems - that between determinate and stochastic problems. Whereas in a determinate problem the "hill" is defined by a single function of many variables

$$f(x) \equiv f(x_1, \ldots x_n)$$

in a stochastic problem it is defined by the mean value of a number of functions:

$$f(x) \equiv f_\alpha(x)$$

when

$$f_\alpha(x) \equiv f(x_1, \ldots x_n;\ \alpha)$$

and α is a random variable whose value depends on the particular trial made.

Now I would have thought that most of the more interesting pattern recognition problems are essentially stochastic: the objective is to make

a machine that will obtain as high a proportion as possible of correct answers to a series of questions "what is this pattern?" referring to a random series of patterns.

The author has given an excellent account of some of the difficulties of determinate hill-climbing and of the techniques for overcoming them, but stochastic hill-climbing presents additional problems, and I think this may become very apparent when a pandemonium is built to deal with a practical task. In particular, two problems which will need investigation are:

(i) how large a sample of trial values $f_\alpha(x)$ and $f_\alpha(y)$, belonging to two points x and y in the vector space being explored, will be needed to obtain a satisfactory estimate of $[f(x) - f(y)]$ for the purposes of hill-climbing?

(ii) how much longer will a particular stochastic hill-climbing procedure take than the corresponding determinate procedure, and will this ratio prove to be too large, in some practical instances, to allow the evolutionary development of the pandemonium to take place?

I should be very interested to learn if the author has considered these problems and developed any solutions. The problem of stochastic hill-climbing is of practical interest outside the context of this paper, for instance in connection with the evolutionary operation of chemical processes.

DR. O. G. SELFRIDGE (in reply[ϕ]): Since I have been working on Morse code, which I am doing in addition to working on learning, I have met probably 50 people and when I say that I am working on a machine to do manually keyed Morse code, they say "I will tell you how to do that" and they then proceed to come up with some scheme. Actually, this is the first time I have heard this particular one, whereas I have come across the others many times, so Dr. Bray should be congratulated on a new scheme for solving Morse code. Let me assure him that it will not work; we have tried it. Morse code manually received is not that simple. The context necessary apparently extends at least 10 letters on *either* side - not just one side, both sides. In fact, the real question about doing this - why I chose Morse code - was exactly that it had had this kind of interactions with each other. You cannot do it by looking at binary functions of the durations, that is, expressing durations as binary digits and then looking on binary functions of lots of them. If I take 10 letters on either side, each has three marks, you should consider the spaces as well, but leaving those out, that is three times 20, which is 60 durations, and if we specify them to one part in 32, that is 300 bits, and binary functions of 300 bits cannot be picked at random. So the question is to take some steps in the right direction and then hope the steps are right enough so that you will get enough improvements, so that the following steps will get you even more so.

[ϕ]Edited versions by the author was not available.

I maintain that there is some merit in studying self-improvement systems and if I am going to do that I am going to study systems where I know what I want, rather than more difficult problems, however attractive they may be to mathematicians. Mathematicians like to work on unsolved problems for the greater glory, and problems already solved like the prime number theorem are left to graduate students. I have a suspicion that John McCarthy might later bring up some important points about descriptions, and here I see my point about solving useful problems, because Morse code is a useful problem. The whole question is at what level you are going to deal with descriptions of your data. In most problems, especially in dealing with the amount of things which people do, binary functions are just not adequate. For one thing there is too much data, and one of the first questions you talk about in the conditioned reflex - remember that conditional probability assumes that you have already recognised the stimulus; if you know that the stimulus belongs to one of a small class of functions, it is pretty easy, as a man, but mostly in human problems you do not know this.

Dr. Mackay made some very kind remarks. I would go further back than he did, as he well knows, in crediting demonology. All of us have sinned in Adam, we have eaten of the tree of the knowledge of good and evil, and the demonological allegory is a very old one, indeed. As for his remarks about the syndicate approach, the evolution of diversative organisation, I think that is an extremely accurate and good point. One of the things a species learns is not only how to survive but to have exactly the right variants in its members. The reason horseshoe crabs have not changed much is not because they don't adapt; they can adapt, but there is no variation, so they can only adapt very little, and they are about as good as they can be; whereas there are all kinds of people.

In speed and information theory, I really consider that speed in the classic sense of computation is so completely irrelevant to this problem - the number of binary operations one does a second - that it does not interest me very much though it interests computer designers. I would rather like to see lots of operations which could conceivably be done in parallel, being done sequentially, because this is the only machine we have, and then I hope people will have machines which will work in parallel, so that when I want a machine to do twice as difficult a problem I merely build twice as big a machine, instead of letting one machine work twice as long.

Dr. Strachey asked about results, and they are roughly as I have indicated. Improvement does take place; the hill-climbing does work. The sub-demon selection in the programmes that we have did not do what we hoped, but the sub-demons were effectively thrown out, and the ones which were kept were largely concentrated around those demons which relate those

ones very close. The most effective demon by itself, which worked 90 per cent. of the time, which was very surprising, was when a mark was the maximum, or within a just noticable difference of the maximum of six demons clustered round zero from minus 3 to plus 3. I am afraid that that is as much as one can say fairly.



SURPLUS - 1
LIBRARY OF CONGRESS
DUPLICATE

© *Crown copyright 1959*

Printed and published by
HER MAJESTY'S STATIONERY OFFICE

To be purchased from
York House, Kingsway, London w.c.2
423 Oxford Street, London w.1
13A Castle Street, Edinburgh 2
109 St. Mary Street, Cardiff
39 King Street, Manchester 2
Tower Lane, Bristol 1
2 Edmund Street, Birmingham 3
80 Chichester Street, Belfast
or through any bookseller

Printed in Great Britain

S.O. Code No. 48—140*