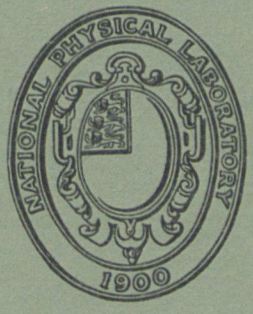
NATIONAL PHYSICAL LABORATORY

SYMPOSIUM No. 10

Mechanisation of Thought Processes

VOLUME II



LONDON: HER MAJESTY'S STATIONERY OFFICE

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

(not to be sold separately)

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 II

LONDON: HER MAJESTY'S STATIONERY OFFICE
1959

SESSION 4A

IMPLICATIONS FOR BIOLOGY

Chairman: PROF. J. Z. YOUNG, University College, London

SESSION 4A

PAPER 1

SENSORY MECHANISMS, THE REDUCTION OF REDUNDANCY, AND INTELLIGENCE

by

DR. H. B. BARLOW

BIOGRAPHICAL NOTE

HORACE BARLOW, Physiological Laboratory and King's College, Cambridge. Age 36. Studied Physiology at Cambridge and Medicine at Harvard and London (Univ. College Hosp.). Has worked on various aspects of vision, including eye-movements; spatial properties of receptive fields in the frog's retina; changes in temporal and spatial summation with level of adaptation; and thresholds as signal/noise discriminations. Worked for a year with S. W. Kuffler at Johns Hopkins on changes in retinal organisation in the cat's retina during dark adaptation. Interested mainly in the nervous organization of the visual pathways.

SENSORY MECHANISMS, THE REDUCTION OF REDUNDANCY, AND INTELLIGENCE

by

DR. H. B. BARLOW

SUMMARY

PSYCHO-PHYSICAL and physiological investigations have shown that the eye and the ear are remarkably efficient instruments: consequently the amount of information being fed into the central nervous system must be enormous. After a delay, which may vary from about 100 msec. to about 100 years, this information plays a part in determining the actions of an individual: therefore some of the incoming information is stored for long periods.

The argument is put forward that the storage and utilization of this enormous sensory inflow would be made easier if the redundancy of the incoming messages was reduced. Some physiological mechanisms which would start to do this are already known, but these appear to have arisen by evolutionary adaptation of the organism to types of redundancy which are always present in the environment of the species. Much of the sensory input is not shared by all individuals of a species (eg. stimuli provided by parents, language, and geographical locality) so a device for "learning" to reduce redundancy is required. Psychological experiments give indications of such mechanisms operating at low levels in sensory pathways, and "intelligence" may involve the capacity to do the same at high levels.

In order to exemplify the operations contemplated, a device which reduces the correlated activity of a pair of binary channels is described.

THE usual mechanistic approach to the higher nervous system begins with a consideration of the factors which can be shown to have an immediate effect on the output of the nervous system. The commonest starting point is the simple monosynaptic reflex in which a single sensory input controls a single motor output, as shown diagrammatically in *fig. 1(a)*. The next stage is to elaborate this by taking into account other sensory modalities, inhibition, internuncial neurones, and controlling neurones from elsewhere

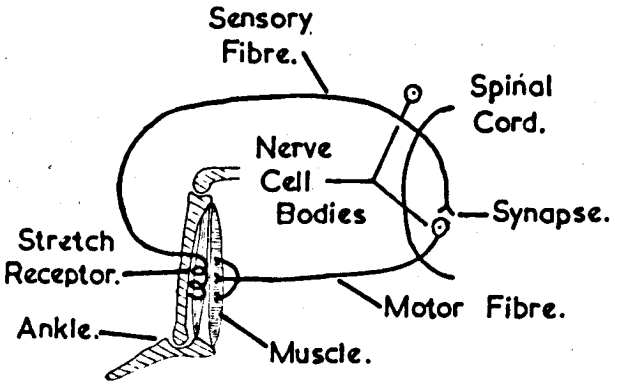


Fig. 1(a)

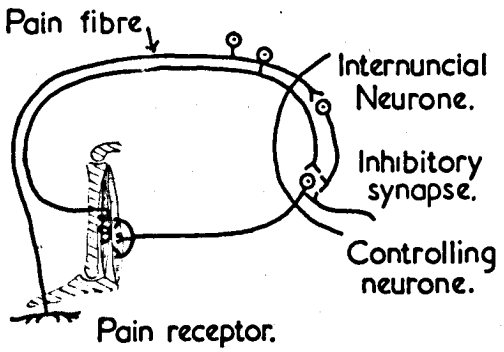


Fig. 1(b)

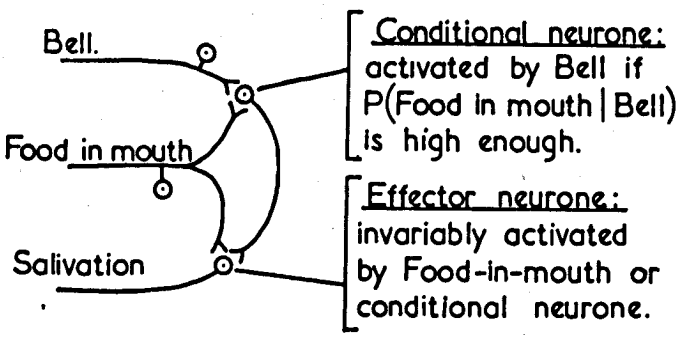


Fig. 1(c)

Fig. 1. Diagram showing approach to higher nervous function from motor (effector) side. (a) monosynaptic stretch reflex; (b) same with addition of internuncial neurones, controlling neurones from other parts of the central nervous system, and inhibition by pain endings; (c) conditioned reflex.

in the nervous system, as shown in *fig. 1(b)*. With all its trimmings this gets one to a stage of complexity perhaps comparable to that of an automatic tracking radar set, or the automatic pilot of an aeroplane. It will show none of the plasticity or adaptability to new surroundings which is characteristic of the higher nervous system, so the Pavlovian conditioned reflex is next introduced. The principle here is that if there are two sensory stimuli (Bell and food in mouth), one of which (food in mouth) always produces a response (salivation), then if they occur jointly with sufficient frequency, the one which, to begin with, did not cause a response, begins to do so (Bell alone causes salivation). This is shown diagrammatically in *fig. 1(c)*, and is perhaps the simplest type of learning behaviour that has been studied in animals, though it has not been investigated in a simple isolated preparation as the diagram might suggest. Uttley (1954, *refs. 22 and 23*) has clarified the principles of operation of such mechanisms and built conditional probability devices which show the same properties of learning and inference.

Now the simple feedback diagram in *fig. 1(a)* has a single input channel, *fig. 1(b)* and *(c)* have two inputs, and Uttley's machine has up to five inputs; but a human brain has something like 3×10^6 sensory nerve fibres leading into it. If it could be supposed that a million or so devices like that of *fig. 1(c)* would deal with the sensory inflow one would be well satisfied with the understanding gained from this approach: but this is not so. The essential operation in a conditional probability device is to measure the frequency of occurrence of *combinations* of activity in the input. Now if the number of binary inputs is increased from two to a million the number of possible combinations is increased from 2^2 to $2^{(\text{million})}$; an arrangement like that of *fig. 1(c)* takes one less far than at first sight appears. I think it follows from this consideration that conditional probability machines cannot be fed with raw sensory information, and the problem of digesting or processing the sensory information entering the brain is an important one. Furthermore, modern electrophysiological techniques are making it possible to record from nerve cells at various levels in the sensory pathways, so this is a problem which is becoming accessible to experimental investigation.

In this paper I have first tried to make rough estimates of the rate at which information flows into the human brain. It is then suggested that an essential step in organising this vast inflow is to derive signals of high relative entropy from the highly redundant sensory messages. For this something similar to the optimal codes discussed by Shannon (1949, *ref. 19*) needs to be devised for the sensory input, and the steps required to do this are considered. Finally, a modified form of such recoding is proposed, some evidence that it occurs is brought forward, and it is suggested that the idea may be extended to cover some of the processes going on in consciousness and called reasoning or intelligence.

1. THE SENSORY INFLOW

(a) Properties of Nerve Fibres

We are equipped with sensory instruments of astonishing sensitivity and versatility which supply information about the environment to the central nervous system. This information is carried along nerve fibres, and since a good deal is known about what these fibres can and cannot do, one can derive an approximate upper limit to the rate at which information enters the brain. If the simple assumptions are made that (i) the maximum frequency of impulses is 700/sec, and (ii) in 1/700th.sec a nerve can only be used to indicate the presence or absence of an impulse, then the maximum rate at which it can transmit information is 700 bits/sec. Mackay and McCulloch (1952, *ref. 16)* point out that the nerve might be used more efficiently if, instead of detecting the presence or absence of an impulse, the intervals between impulses are used to convey information. Using such pulse interval modulation, and assuming (i) accuracy of estimation of intervals of 0.05 msec, (ii) a minimum interval of 1 msec, they give the maximum capacity as 2880 bits/sec. This would require a mean frequency of 670 impulses/sec, but at a mean frequency of 50/sec, such pulse interval modulation still allows 500 bits/sec to be transmitted. These figures are actually too low, because Mackay and McCulloch incorrectly assumed that the optimum distribution of intervals was uniform instead of exponential: however, if the other assumptions are granted, they show clearly that a single nerve fibre could be used to transmit information at a rate well above 1000 bits/sec.

The total capacity of the sensory inflow appears to be above 3×10^9 bits/sec, but it is certain that nothing like the full capacity is utilised. The mean frequency of impulses must be far below the optimum; peripheral nerves appear to use pulse frequency rather than pulse interval modulation, so that there will be high serial correlations between the values of intervals; furthermore, there are generally considerable overlaps in the pick-up areas of neighbouring fibres, which are therefore bound to show correlated activity. Finally, the figure for the performance of a nerve fibre given above might be approximately true for the large diameter fibres, but those of smaller diameter, which make up a large fraction of the total number, must have a smaller capacity. It would be pure guesswork to try to allow for these factors, but one can get indications of the utilised capacity from two other sources.

(b) Sensory Ability

Jacobson (1950, 1951 *refs. 13,14*) has made estimates of the informational capacity of the ear and the eye. For the ear he calculated 50,000 bits/sec from the number of discriminable pitches (about 1450), the number of discriminable intensities at each pitch (average about 230), and the time required

to make such discriminations (1/4 sec). This does not make any allowance for masking - the observed fact that the presence of one tone interferes with the perception of other tones. Jacobson calculated that this would reduce the information capacity by a factor of about six, bringing it down to 8,000 bits/sec. Now there are 30,000 nerve fibres from the ear, so each fibre must carry an average of about 0.3 bits per sec.

For the eye he calculated from published data of central and peripheral acuity that there were 240,000 resolvable elements in the visual field (he seems to omit a factor of two in the integration, but this is perhaps compensated by the rather high figure for acuity which he uses). He supposes that each element can be discriminated at two intensities, with an average temporal resolution of 1/18 sec. These figures give 4.3×10^6 bits/sec. In the optic nerve there are just under a million fibres, so about 5 bits/sec are conveyed on the average by each fibre,

These are crude estimates. For instance, no account has been taken of colour discrimination, or of the ability to localise a sound by binaural effect and judge depth by stereoscopic vision. Nevertheless, they are probably of the right order of magnitude and they are probably good enough to justify the claim that optic nerve fibres carry much more information than those of the auditory nerve. This may be significant and will be referred to later.

These figures suggest that total sensory inflow along the three million sensory fibres is rather under 10^7 bits/sec.

(c) Communication bandwidths

The capacity of the communication channels engineers need to transmit auditory and visual signals is clearly related to the capacity of the sensory pathways. Engineers, in the interests of economy, may be expected to try to use the narrowest bandwidths which will satisfactorily load up the sense organs involved, and recipients may be expected to insist that such satisfactory loading is not too far short of normal loading.

Ten k.c. bandwidth at 40 d.b. signal noise ratio give a good quality auditory signal, and has a capacity of 133,000 bits/sec. This is more than ten times Jacobson's final figure for the capacity of the ear (8,000 bits/sec), and the discrepancy is presumably due to (i) the transmission of relative phases of the frequency components, which gives information not utilised by the ear - at least in the type of discrimination taken account of by Jacobson; (ii) the failure of the engineer to exploit the loss of efficiency of the ear which results from masking.

A satisfactory 400 line television picture requires three megacycle bandwidth at about 10 d.b. signal-noise ratio, and this corresponds to 1.2×10^7 bits/sec. One is much more aware that such a television picture falls short of one's normal visual signals than one is in the case of a

10 k.c. 40 d.b. auditory signal because it does not fill the visual field, and lacks detail and colour, but it is still more than double Jacobson's estimate of the eye's capacity. In this case the most notable matching errors are the failure to exploit (i) low peripheral acuity of the eye, (ii) reduced temporal and spatial resolving power in low intensity regions of the image.

Engineers seem to require 5 - 10 bits/sec channel capacity per nerve fibre to load up our sensory pathways, but the discrepancies between this figure and those obtained from direct estimates of sensory abilities can probably be attributed to poor matching.

(d) Time of storage

Not only is the input to the nervous system enormous, but some, at least, of the messages received are stored for very long periods. Most people would agree that sensory impressions can be recalled after a lapse of, say, 70 years, and sometimes a person can produce objective evidence of the accuracy of his recollections. In addition there are, of course, many sensory impressions which cannot be recalled, but which have, none the less, left their mark: we do not remember the successes and failures by which we acquired the correct usage of 'yes' and 'no', but this correct usage is often retained beyond the retiring age. If one allows for fifty years of waking life, the total sensory input is something like 10^{16} bits. Complete storage of all this information is neither likely to be possible nor, of course, is it what is needed.

(e) Fate of Sensory Information

The rest of this paper is about a suggested plan of storing and displaying this enormous sensory input, but one must first have some idea of the use that is made of the sensory information and the neural equipment which is available for dealing with it. According to Craik (1942, *ref. 4*) the sensory information is used to build up a model of the external world which provides a basis for determining what course of action is most likely to lead to the survival of the individual and his species. That is a brief answer to the first question, and it also gives the answer to another fact which might otherwise be puzzling. A man can only make decisions on the basis of sensory information at a maximum rate of about 5 to 25 bits/sec. (Hick, 1952, *ref. 11* Quastel, 1956, *ref. 18*): why, then, does he need a sensory input of 10^7 bits/sec.? Craik's answer would probably have been that the greater the sensory input the more complete and accurate the model, and hence the surer its basis for planning survival.

The question of the equipment available can also, because of our ignorance, be answered briefly. There are some 10^{10} interconnecting nerve cells in the central nervous system, and quite a large proportion of them must be

available for the task of dealing with the sensory input and building up the model. We are only beginning to determine the properties of these cells; it has been known that their long processes transmit information as all-or-none impulses for more than fifty years, but how information is stored is not yet understood. In what follows I shall be talking about *what* the nervous system does rather than *how* it does it, so our ignorance of the method of storage of information is not too serious. The problem might be discussed abstractly, but for the sake of a definite model one can think of each nerve cell having "excitation laws" which determine the conditions under which it becomes active, and suppose that these laws can be changed so that it becomes active in response to a different set of patterns of activity in the nerve cells in contact with it. The excitation laws for all the neurones would then form a store of information and the current display would consist of the pattern of nerve cells which are actually transmitting impulses down their long processes at any given moment.

With this model in mind the problem is: what should the excitation laws of the neurones be, and how should they be alterable, in order that the display of activity shall help the individual and species to survive in the situation giving rise to the current sensory input? To avoid basing the argument on uncertain preconceptions of what the brain does, one could put it in more general terms in this way. The barrage of nervous impluses reaching the nervous system seems to be unmanageably large; how should a selection of this activity be made for current display and future reference?

2. ORGANISATION OF THE SENSORY INPUT

The proposition is that the initial selection is performed according to those statistical properties of the past sensory messages which determine how much information particular impulses convey. It is supposed that the sensory messages are submitted to a succession of re-coding operations which result in reduction of redundancy and increase of relative entropy of the messages which get through. Ideally one might imagine that an optimal code is constructed, so that the output, or "display" of current input, has no redundancy, relative entropy 1, and carries all the information of the input. This ideal obviously cannot be reached, but the re-coding operations are supposed to tend towards the ideal: that is, outputs are derived from the input, which have high relative entropy and carry as much of its information as possible.

Shannon has shown that it is possible in principle to obtain near optimal coding if a sufficient number of messages of a given length have occurred to give knowledge of the statistical structure of the messages,

and if delays are permitted between input and output. Fano and Huffman (1953, *ref. 12*) have described procedures for constructing such codes. The first steps are to define what shall constitute a single message and then to measure the frequency of occurrence of all possible messages of this class. Clearly the class cannot be the whole of the sensory input to the brain up to a particular moment, for this message has only occurred once. The input must be sub-divided in time, and first consider the operation required to re-code messages of duration, say, one second. The capacity of the input channel has been shown to be about 3×10^9 bits/sec. which corresponds to $10^{(\text{thousand million})}$ possible messages per second. If one takes account of the restrictions which reduce the utilised capacity to some 10^7 bits/sec., and considers messages of one-tenth second duration, there are still some $10^{300,000}$ possible messages. It would clearly be hopeless to devote neural equipment to the counting of each possible message, for it is highly improbable that any single message will be exactly repeated and most of such equipment would be unused at death. This is, essentially, the same difficulty that was levelled against the idea that conditional probability devices could be served with unprocessed sensory data, but when one considers optimal coding there is a possible solution. Because the code is reversible, no information is lost by re-coding small sections of the sensory input independently, and such preliminary re-coding will enable the whole message to be passed down a channel of smaller capacity, and thus facilitate subsequent steps.

The idea is best illustrated by considering the order in which different types of redundancy might be encountered, and eliminated, during the successive re-coding operations. First there is the very large amount which results from the inefficient utilisation of peripheral nerve fibres. Looking only at the nerve impulses as they arrive, it would be found that impulses occurred at different mean rates in different fibres and in all of them at rates well below the optimal frequency for information transmission. This type of inefficient utilisation of a set of communication channels is a form of redundancy, but for reasons discussed later (Section 4) it may be less important to eliminate than other forms: for the moment one can consider the capacity of a nerve fibre as determined, not by maximum frequency of impulses, but by the mean frequency at which they occur.

Next, still looking only at the impulses as they reach the central nervous system, it would be found that impulses do not occur completely at random in time but tend to follow one another in sequences and bursts: the first re-coding operation might be a mechanism which reduced the serial correlations so that the same amount of information was carried by fewer impulses. In addition it would be found that certain groups of nerve fibres tended to become active at the same time. These would be fibres whose receptive fields on the sensory surface overlapped, so that this particular form of redundancy results from the anatomical properties of

fibres and sense organs, just as the serial correlations in time result from the fact the intensity of a stimulus is coded as frequency of impulses at the sense organs.

These first steps, then, would reduce the orderliness in the sensory messages which results from characteristics of the sensory apparatus. But if this orderliness can be eliminated, so can that resulting from the characteristics of the environment which is providing these stimuli. For instance, it will often happen that a stimulus covers more than a single point on the sensory surface and therefore causes activity in a group of fibres larger than those whose receptive fields overlap. Advantage could be taken of this to reduce the number of impulses required to convey information about such a stimulus. Again, a stimulus will often be moved across a sensory surface causing excitation in sequences of nerve fibres. Such repeated, ordered, sequences of activity would be a form of redundancy which could be reduced by suitable re-coding. In fact, any pattern of stimuli which represents a departure from complete randomness - such as simultaneous stimuli at different points on the sensory surface, stimuli which are maintained for long duration of time, ordered sequences or cycles of stimuli - present an opportunity of reducing the magnitude of the sensory inflow by suitable re-coding. It is clear that many of the complex features of our environment will come into this category. For instance, the stimuli which result from an animal's parents or its habitat are repeated frequently, and economies could be effected by reducing the space in the sensory representation occupied by these familiar stimuli and allowing more space for the infrequent and unexpected stimuli.

It is suggested, then, that the processing or organisation of sensory messages is carried out by devising a succession of optimal or near-optimal codes adapted to the messages which have been received. In the early stages the total inflow will be sub-divided into many small sections, presumably taking in each section the messages coming along neighbouring fibres during a short interval of time. In the later stages the coded outputs will be re-mixed, possibly with the addition of delayed inputs (as utilised by Uttley in conditional probability devices) to allow detection of movement and other ordered sequences of activity, and then will be sub-divided again into small sections. Thus in the later stages the nerve messages being re-coded may be derived from more and more remote parts of the sensory inflow and may also come from sensory stimuli more and more separated from each other in time of occurrence. It will be seen that at each stage storage of some of the sensory information is required in order to construct the optimal code, and thus the code itself forms a kind of memory.

Now the idea that our brains detect order in the environment is not new. Empiricist philosophers have talked of percepts being associated sense impressions, and of causality corresponding to invariant succession of sense impressions. Behaviourists have emphasised the importance of

association, and Gestalt psychologists talk of ordering sensation according to certain schemata (though here there seems to be some confusion as to whether the ordered schemata are derived from sensations or imposed upon them). Thus the fact that our higher centres are much concerned with the redundancy of the sensory messages has often been pointed out, but two aspects of this fact have not, I think, been so widely recognized. First, the detection of redundancy enables the sensory messages to be represented or displayed in a more compact form; and second, the reduction of redundancy is a task which can be subdivided and performed in stages. *Figure 2* shows diagrammatically how the suggested scheme of storage and display compares with more othodox representations of memory and consciousness. It will be seen that in the present scheme a large part of the storage of information occurs before the display - that is before the level of re-coding which might correspond to conscious awareness of sensory stimuli. The re-coding is supposed to continue at conscious levels, so some of the information reaching consciousness is also stored, but this would only be sufficient, first, to enable the process of building up the code to continue, and second to enable "useful" association to be made between motor acts and features of the current sensory input (e.g. between salivation and bell).

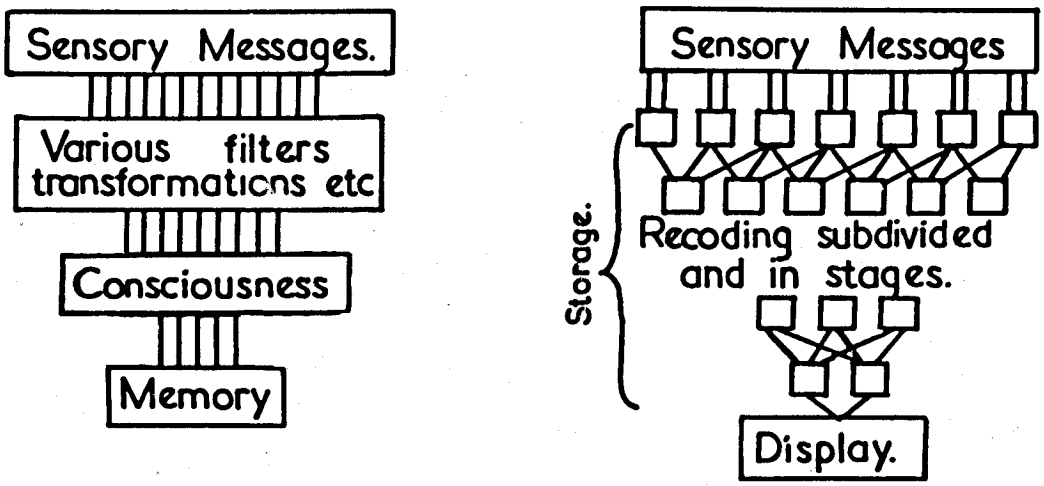


Fig. 2. Diagram contrasting memory *after* consciousness in orthodox scheme with storage *before* display in optimal coding scheme.

It seems a help to consider the processing of sensory information as optimal or near-optimal coding for two reasons. Practically, it enables the subject to be approached along the firm path of sensory physiology instead of through the shifting sands of conscious introspection and philosophy. And conceptually it shows a way in which complete mental acts, which seem appalling in their complication and perfection, may be sub-divided into a succession of much simpler operations; this is clearly a prerequisite for gaining an understanding of the physiological basis of mental function.

It is worth noting that the possibility of sub-division rests on Shannon's proof of the possibility of near-optimal coding; if the early transformations of the sensory information were not reversible, redundant features which are detected later might be lost: and if the earlier transformations did not increase the relative entropy of the messages, they would not facilitate the detection of higher order redundancy.

3. DESIRABILITY OF OPTIMAL CODING

In the last section an outline scheme for dealing with the enormous sensory inflow was suggested. In this section some reasons for the desirability of optimal coding are put forward. It will be argued that it is desirable on the grounds of accessibility, stability, and economy, and because it requires storage of information sufficient to form a model of the animal's environment. Of course, such arguments for its desirability are not sufficient reasons for believing that it actually occurs.

(a) Accessibility.

Optimal coding will improve the accessibility of information in two ways. First, the capacity of the display required for the current sensory input will be decreased. This simplifies the task of finding useful associations just as reducing the size of a haystack simplifies the task of finding needles. The second way is less obvious. In messages of high relative entropy, the probability of a given message occurring is close to the product of the probabilities of the individual signs which make it up. Now a dog feeds once or twice a day, and when looking for sensory correlates of salivation it would not be worthwhile to search among combinations of individual signs whose probability of joint occurrence was so low that they would be expected only, say, once a week, nor amongst those whose probability was so high that they would be expected, say, once an hour. If the input to a conditional probability device is known to be of high relative entropy, great economies of design are possible.

(b) Stability

It is sometimes argued that redundancy is a good thing because it protects a message from noise. There may well be random effects inside the nervous system against which the storage and display of sensory information needs protection, but the redundancy of the internal representation which would achieve this is not in general the same as the redundancy which occurs in the sensory input. When driving at night the internal representation of a pedestrian crossing the road requires as much protection as the representation of the blinding glare from an oncoming car, but in the incoming sensory message the former may be represented by a barely significant disturbance in the pattern of nerve impulses, the latter by high frequency volleys of impulses in many fibres. Stability of storage and display require, at least, a re-adjustment of the redundancy of the sensory messages.

(c) Economy in transmission and storage.

Sensory information has to be transmitted from place to place in the central nervous system and the reduction of redundancy before this is done would enable the number of internal connecting fibres to be reduced. An example where the economy so effected seems to be particularly desirable is the connection between the eye and the brain. It would clearly interfere with the mobility of the eye if the optic nerve was very much larger than it is, and according to Jacobsen's estimates it would have to be fifteen times larger if the nerve fibres were utilised as inefficiently as they are in the ear. The attainment of this 15-fold economy may, as Jacobsen suggests, be the main function of the nervous layer of the retina which links receptors to optic nerve fibres. Squids and octopuses form an interesting comparison, for they have eyes which are comparable optically to those of vertebrates, but their retina is much simpler with no synaptic layer - the optic nerve comes direct from the receptor cells. It is bulky, containing a vast number of fibres, and seems likely to be a factor restricting the mobility of their eyes.

The same argument might be applied to storage of information, since it is clearly more economical to store messages after their redundancy has been reduced. Here, however, there is a complication. The devising of a redundancy-reducing code requires storage of certain properties of the sensory message, and it has not been shown that more capacity would be saved by storing messages after re-coding than would be utilised in devising the code. The condition that this should be so depends upon the number of the times that the code, once devised, is subsequently utilised, but a discussion of this point cannot go far without knowing what parts of the sensory inflow are in fact stored: the argument of the next section is that the coder itself stores sufficient information to form a working model of

the animal's environment, and therefore represents a large fraction of the total storage the animal needs.

(d) Modelling the Environment.

Craik suggested that sensory information was used to form a model of the animal's environment. By a model one does not mean a simple copy of those aspects of it which have given rise to sensory stimuli: it must also mimic the structure of the environment, so that an operation performed on the model will give the same result as the analogous operation performed on the environment. When the schoolboy turns his model engine round, he receives visual stimuli similar to those he would have received if a real engine had been turned round in front of him. The model imposes restrictions on the sensory stimuli which are received in certain situations, these restrictions being the same as those inherent in the properties of the object modelled. Now it is precisely these restrictions - the departures from complete randomness of the sensory input - which the coder utilises to increase the relative entropy of the signals. The particular code adopted is related to the particular restrictions of past sensory inputs and is therefore, in a sense, a model of the animal's environment. In the example above, the model was static, but the restrictions must often be dynamic; sets of sensory stimuli frequently follow one another in a repeated sequence, and such repeated sequences will also be reflected in the particular code adopted. Thus the code contains a working model of the environment.

If the code stores sufficient information to form a model of the environment, its potential use in aiding survival is not confined to the provision of a more compact display of the sensary input. But to make full, predictive, use of these potentialities some additional facility for getting at this stored information seems to be needed. To return to the earlier example, what facility do we have for turning round the model engine in our brain so that we can look at the other aspect?

4. MODIFIED RE-CODING

So far the type of optimal coding envisaged has been that described by Shannon, Fano, and Huffman, in which the output is the smallest number of binary signals capable of carrying the information of the input. At first sight this seems to be what is needed in the nervous system, for nerve fibres transmit all or nothing impulses and thus seem to use a binary system. However, it has already been pointed out that the mean frequency of impulses is well below the optimal for information transmission even in peripheral nerve fibres, and there is some evidence which suggests that the

mean frequency is even lower in the more central neurones (Galambos, 1954, *ref. 6*). Furthermore if the Shannon type of re-coding was occurring, one would expect to find the sensory pathways becoming more and more compact as the sensory information was coded on to fewer and fewer elements. This does occur in the retina, where some 10^8 sensory elements are connected to 10^6 nerve fibres, but as one follows the optic nerve into the brain there is no evidence of further compression on to a smaller number of nerves, but rather the reverse. The striate region of the cerebral cortex which is mainly, perhaps exclusively, concerned with vision, contains some 10^8 nerve cells; in other regions of the cortex there are about 6.5×10^9 cells (Sholl 1956, *ref. 20*) many of which must be partially concerned with visual information. Galambos (1954, *ref. 6*) gives striking figures showing how the number of nerve fibres available for auditory information increases as one follows the sensory pathway from ear to cortex.

These facts do not fit in with the idea that coding in the higher nervous system compresses information into a smaller number of nerve fibres, and suggest that, if optimal coding occurs, the output is not in the form of binary signals at the optimum frequency for information transmission.

For an engineer designing a communication link, the capacity of the channel is one of the factors under his control, and he can effect economies by coding his signals so that they require a smaller capacity. In the nervous system the number of nerve fibres available for a particular task must, to a large extent, be determined genetically. One may expect evolutionary adaptation to have performed part of the engineer's job in selecting suitable codes for the sensory signals, but such inherited codes obviously cannot be adapted to the redundancy of sensory input which is peculiar to each individual. Now although the number of nerve cells available is probably determined genetically, the number of impulses in the nerve cells is not, and some of the advantages of optimal coding would apply if the incoming information were coded - not onto the smallest possible number of nerve fibres each working at its optimal mean frequency - but into the smallest possible number of impulses in a relatively fixed number of nerves. This type of coding can be epitomised as *economy of impulses:* the nervous system will tend to code sensory messages so that they are represented, on the average, by the smallest number of impulses in the nerve cells available. There is an important difference between this type of re-coding and the Shannon - Fano - Huffman type; the latter does not distinguish between redundancy caused by non-optimal frequency of utilisation of the individual signs of the input message, and that which is caused by correlation between signs. If impulses rather than nerve fibres are economised, mean impulse frequencies of the output will be as low as the rate of inflow of information permits, and will thus possess maximum redundancy of the first type and minimum redundancy of the second type.

A reversible coding device is described in the appendix which decreases the frequency of occurrence of a pair of binary output signs by getting rid of some of the redundancy caused by correlations between a pair of binary input signals.

5. EVIDENCE

So far some grounds for believing that the optimal coding of sensory information would be desirable have been given, arguing from the enormous quantity of information pouring in and from rather vague ideas about what the brain does with it. In this section some of the evidence in favour of the view that it does actually occur is sketched, but this is intended to show the kind of consequences of optimal coding which may be found experimentally, and is neither a claim that it has been proved to occur, nor a critical review of the evidence for and against it. The evidence comes from a number of sources.

(a) Introspection of sense impressions

This is a notoriously unreliable way of obtaining scientifically valid evidence, but it is immediately accessible to everyone, so it comes first. If the hypothesis is correct, the sensory messages reaching consciousness will have been partially re-coded, and will therefore have higher relative entropy and lower redundancy than the raw sensory messages. This seems to me likely to be true of the furniture of my own consciousness, and others may feel it is true also: if, however, somebody did not agree I don't think I could persuade him by verbal arguments. More objective evidence can be obtained by looking at some messages which do *not* reach consciousness but which are known to be impressed on the sense organs. Examples of this are the shadows of the blood vessels which run on the retina in front of the sensitive elements; the fact that if distorting or inverting spectacles are worn, after some days one ceases to be aware of the distortion or inversion; adaptation to the curious tone quality imposed on all sounds by the average domestic wireless set and so on. In all of these examples there are features of the sensory messages which are constantly repeated and are therefore redundant; a code which increased the relative entropy of the messages might be expected to reduce their prominence, and the fact that we cease to be conscious of them suggests that this re-coding does take place before sensory messages reach consciousness.

An experimental approach to this problem may be possible through the investigation of threshold sensations. These are perhaps the simplest elements of our consciousness, and according to the hypothesis they should tend to possess the highest possible relative entropy of a binary signal

after the physical limitations of the stimulus and of the sense organs have been taken into account, and they should show a tendency to retain this property in a great diversity of stimulus conditions.

(b) From sensory neuro-physiology.

During the past thirty years various types of relation have been observed between an applied physical stimulus and the resulting pattern of nerve impulses. Physiologists have perhaps got used to these transformations and no longer think of them as something requiring further interpretation, but possibly they can be looked upon as examples of the principle of economy of impulses: the relation between the physical stimulus and the occurrence of impulses is such that the number of impulses used to convey information about the stimulus is lower than it would be with other, more straightforward, relations.

(i) Adaptation. When a sustained physical stimulus is applied to a sense organ the nerve fibre often responds with a brief burst of impulses which rapidly decreases in frequency and is not sustained for the duration of the physical stimulus. In the left half of *fig. 3* comparison of the trains of impulses shows the economy brought about by adaptation. But it can, of course, only be thought of as an economy when compared to a non-adapting ending, and even then only when the physical stimuli naturally applied to the sense organ are frequently of a long-sustained type. Nevertheless, where it occurs, adaptation would lead to economy of impulses and Adrian (1928, *ref. 1*), suggested that its function might be to prevent an excessive number of impulses reaching the nervous system.

(ii) Inversion. It can be seen from the right half of *fig. 3* that an adapting nerve fibre fails to signal the end of a sustained stimulus. This defect could be remedied by having one which discharged as shown in the bottom line, and such nerve fibres are found. In the eye of the scallop (Pecten) Hartline (1938, *ref. 9*) showed that one group of fibres discharged when a light was switched on and another group of fibres discharged at 'off'. A similar, but rather more complex, situation is found in the vertebrate eye (Hartline 1938, *ref. 8*; Granit, 1947, *ref. 7*). This arrangement might be thought of as making good some of the loss of information caused by adaptation.

(iii) Lateral inhibition. Adaptation increases the relative entropy of the nerve message by preventing many impulses being used to signal a physical stimulus which is constant in time. It is clear that physical stimuli will often be applied to many neighbouring receptors simultaneously, so there is a place for a spatial analogue of adaptation. The best worked out example of this occurs in the lateral eye of Limulus, where the arrangement shown in *fig. 4* has been deduced by Hartline and his co-workers *(ref. 10)*.

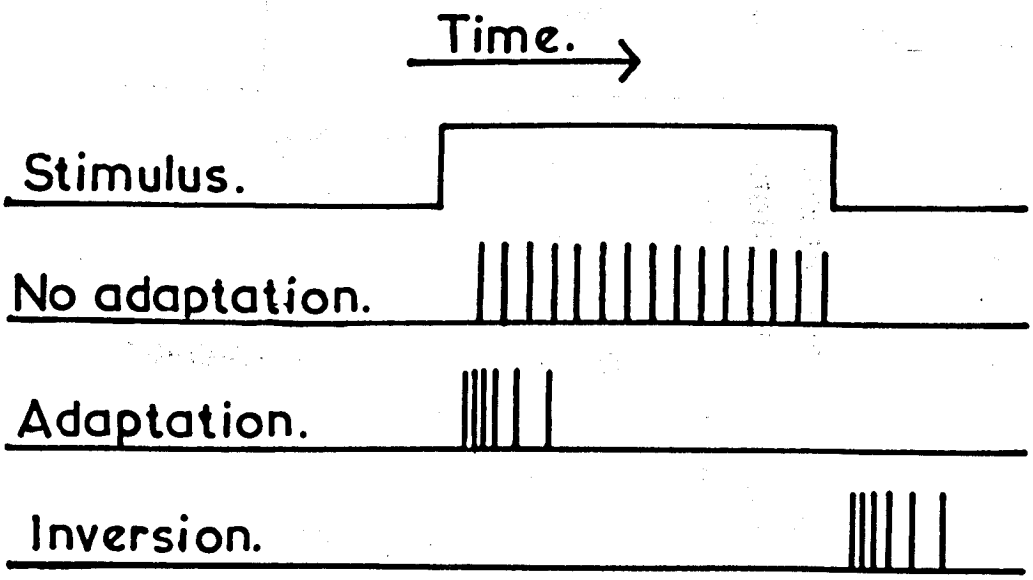


Fig.3. Diagram showing that adaptation leads to economy of impulses when a physical stimulus is of long duration, and that inversion replaces information lost by adaptation.

Apparently each receptor in the array exerts an influence, graded according to the number of impulses it is itself producing, which reduces the number of impulses given by neighbouring receptor units. It will be seen that the effect is to decrease the number of impulses coming from a uniformly illuminated area, while the number coming from the borders of the area are relatively unaffected. A similar situation exists in the frog (Barlow 1953, *ref. 2*) and cat retina (Kuffler, 1953, *ref. 15*) and it has also been described in the auditory (Galambos 1944, *ref. 5*) and tactile (Mountcastle, 1957 *ref. 17*; Amassian, 1958, *ref. 21*) pathways.

One feature of lateral inhibition in the mammalian retina is of special interest: it is found when the retina is adapted to a uniform background light, but is absent after complete dark adaptation (Barlow, FitzHugh, and Kuffler, 1957, *ref. 3*). Now it is only when the uniform background is present that the correlated discharge of neighbouring receptors will tend to occur, so it looks as though lateral inhibition is not an invariant feature of the retinal organisation, but develops in the conditions where it can increase the relative entropy of the optic nerve signals. Perhaps

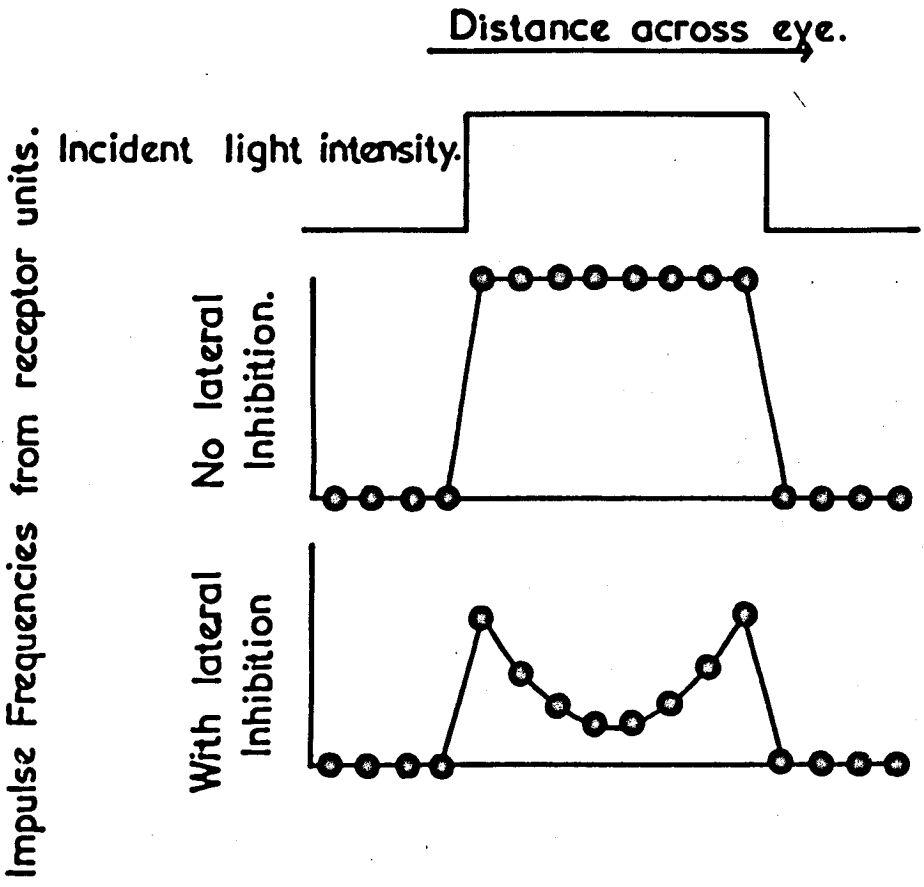


Fig.4. Diagram showing that lateral inhibition leads to economy of impulses in a uniformly illuminated area.

it is a simple example of "learnt" re-coding adapted to the redundancy which is present.

Adaptation, inversion and lateral inhibition may thus be devices used in the peripheral parts of sensory pathways to obtain signals of higher relative entropy. It is now possible to record the activity of more centrally placed neurones, but the nervous system has outwitted the physiologist who has so far been unable to determine the function of the cells whose nervous responses he has recorded. The model described in the appendix does a simple re-coding operation on two binary inputs, but it would be a difficult task to relate the output to past and present inputs without some hint about the purpose of the device. The reason, then, for putting forward the

optimal re-coding hypothesis is the hope that it may be better matched to the subtlety of the nervous system than the simpler hypotheses at present entertained in physiology.

6. INTELLIGENCE

This word was added to the title in an incautious moment, but there are reasons justifying its inclusion. If it is accepted that the large size of the sensory inflow precludes its direct utilisation in the control of learnt motor actions, then the mechanism which organises this information must play an important part in the production of intelligent behaviour. In addition, when one considers the two main operations required for optimal coding there is a striking parallel with the two types of reasoning which underlie intelligence.

The outputs of a code can be thought of as logical statements about the input, and, if the code is reversible, these logical statements, taken together, are sufficient to determine the exact input. Forming these statements and ensuring that they fulfil this condition are straightforward problems of deductive logic. If the code is optimal, the output statements must be chosen so that they fulfil the additional condition that, on the average, they are the smallest possible number which suffices to determine the input (for the type of modified optimal code suggested in Section 4, the additional condition is that a fixed number of possible statements are chosen for the output in such a way that the smallest number, on the average, are asserted as true). The fulfilling of these additional conditions is not exactly inductive reasoning, but it is closely related to it, for both depend on counting frequencies of occurrence of events. Having been presented with 1000 white swans and no black ones, the relevant parts of a code would say "henceforth regard all swans as white unless told otherwise". Inductively one would say "all swans are white". The tools of logical reasoning appear to be the same as those needed for optimal coding, so perhaps they can be regarded as the verbal expression of rules for handling facts which our nervous system uses constantly and automatically to reduce the barrage of sensory impulses to useable proportions.

Finally it should be made clear that the transformations of sensory messages taking place in the nervous system must, in fact, fall a long way short of true optimal coding: information must be lost, and the final "display" must still contain redundancy. However, the fact that the image cast on the retina is not always sharp does not mean that the focussing of light by the eye is unimportant, and the suggestion is that optimal coding

plays a part in the organisation of sensory information comparable with image formation in the working of the eye. However, even if this conjecture is correct, the means by which it is achieved, and such matters as the classes of redundancy which are easily and naturally utilised, and the classes which are not, remain largely undetermined.

APPENDIX

(In collaboration with P. E. K. Donaldson)

Object of device. To code reversibly and without delay a pair of binary inputs (A and B) onto a pair of binary outputs (X and Y) so that the redundancy of the output due to correlations is less than the same type of redundancy in the input.

Principle used. The information carried by the inputs will, in general, be less than the capacity of the input channels *first* because of redundancy due to correlations between them ($P(AB) \neq P(A).P(B)$): *second* because the frequency of signals in the individual channels is not optimum ($P(A) \neq \frac{1}{2}$ and $P(B) \neq \frac{1}{2}$). The principle used is to increase the redundancy of the second type, and so decrease that of the first type. A pair of outputs are sought which are reversibly related to the inputs, and one of which occurs with probability further from the optimum ($\frac{1}{2}$) than one, or both, of the inputs. The outputs carry the same information as the inputs, so that if such a pair can be found, the redundancy due to correlation between them must be less than is present in the inputs.

Possible Codes. There are four possible input states ($AB, A\overline{B}, \overline{A}B$, and $\overline{AB}$), and four possible output states ($XY, X\overline{Y}, \overline{X}Y$, and $\overline{XY}$). If the code is reversible these must be related to each other in a one-to-one manner, which can be done in 24 ways. Now since X corresponds to a pair of output states ($XY + X\overline{Y}$), the condition for activity in X must be the occurrence of either of a pair of the possible input states, and likewise for Y. There are six such pairs:- $AB + A\overline{B} \equiv A$, $\overline{A}B + \overline{AB} \equiv \overline{A}$, $AB + \overline{A}B \equiv B$, $A\overline{B} + \overline{AB} \equiv \overline{B}$, $AB + \overline{AB} =$ (A and B the same), and $A\overline{B} + \overline{A}B =$ (A and B different). In addition, for reversibility, the two pairs chosen must have a common member, for if this was not so X would always be active when Y was not active, and vice versa.

After a little cogitation it will be found that there are 24 possible codes, which fall into 3 groups each containing 8 codes, the groups differing from each other in the respect which interests us, namely the division of redundancy between correlation-type and non-optimal-frequency-type. One group does not differ from the input in this respect. The other two groups do differ, and they are made up of those 16 codes for which one or other of

the outputs corresponds to either $AB + \overline{A}\overline{B}$ (A and B alike) or $A\overline{B} + \overline{A}B$ (A and B different).

Condition for success, then, is that either $P(AB + \overline{A}\overline{B})$, or $P(A\overline{B} + \overline{A}B)$, should differ from $\frac{1}{2}$ by more than one or other or both of $P(A)$ and $P(B)$. This is not, of course, the same as the condition that A and B are correlated, so the recoding does not always reduce correlation redundancy when this is present. Successful recoding occurs for the smallest departures from zero correlation when either $P(A)$ or $P(B)$ is close to $\frac{1}{2}$.

Method. The device is made up of 6 similar units each of which compares two probabilities and operates a relay according to which is greater (see circuit diagram, *fig. 5*).

(a) Probabilities are measured by charging a leaky condenser when A (or B etc.) = 1; hence they are weighted for recent events, the weights decreasing exponentially with lapse of time. These time weighted probabilities are called $P'(A)$, $P'(AB + \overline{A}\overline{B})$, etc.

(b) $P'(A)$ is compared with $P'(\overline{A})$, $P'(B)$ with $P'(\overline{B})$, and $P'(AB + \overline{A}\overline{B})$ with $P'(A\overline{B} + \overline{A}B)$. In each case a signal corresponding to the smaller of the pair is selected. Call these signals K, L, M.

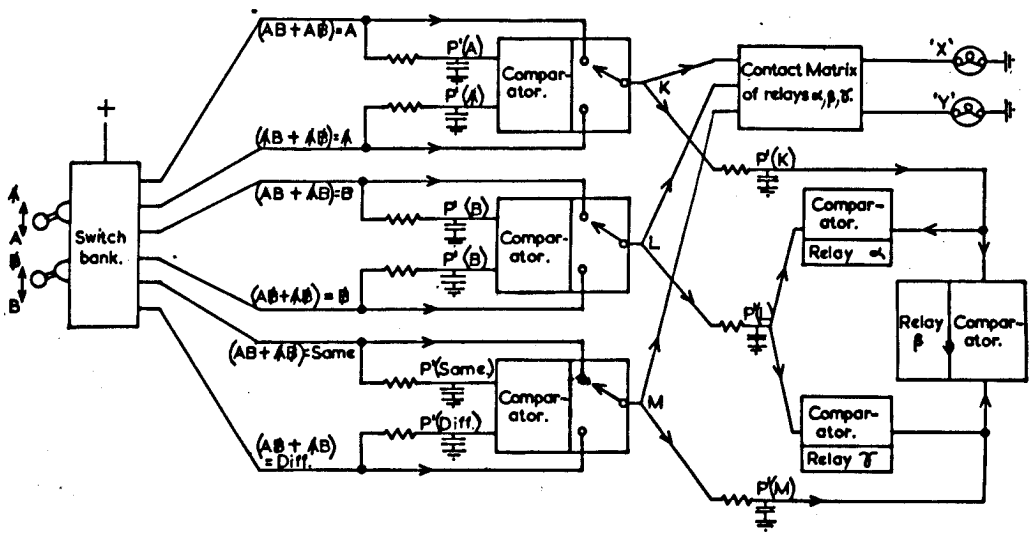


Fig.5. Circuit diagram of recoding device.

(c) $P'(K)$ is compared with $P'(L)$, $P'(L)$ with $P'(M)$, and $P'(M)$ with $P'(K)$. Switching is performed according to the result of these comparisons so that

$$X \equiv \text{smallest of } K,\ L.\ M.$$

$$Y \equiv \text{next smallest of } K,\ L,\ M.$$

Result. The result of these operations is more specific than the original objective in that one particular code is chosen from a group of 8, any one of which would have met the requirements. The added specificity results from the fact that we have chosen outputs which occur *least* frequently, not *most* frequently, and have arranged that $P(X)$ shall be less than $P(Y)$.

Note that if there is any logical relation in the inputs (e.g. $\bar{A}B = 0$), then the outputs become mutually exclusive ($P(XY) \equiv 0$). If there is a double relation (e.g. $\bar{A}\bar{B} = 0$ and $\bar{A}B \equiv 0$), then only one output channel operates ($P(X) = 0$). The device might be roughly described as one which determines inductively what logical relations, if any, are obeyed by its input. If two such relations are found, one output channel is not used; if one is found, the two outputs become mutually exclusive; if none is found, but there is statistical correlation between the inputs, it will sometimes find outputs which are less correlated.

REFERENCES

1. ADRIAN, E. D. The Basis of Sensation. *Christophers, London,* (1928).
2. BARLOW, H. B. Summation and Inhibition in the frog's retina. *J.Physiol.*, 1953, **119**, 68.
3. BARLOW, H. B. FITZHUGH, R, and KUFFLER, S. W. Change of organisation in the receptive fields of the cat's retina during dark adaptation. *J.Physiol.*, 1957, **137**, 338.
4. CRAIK, K. J. W. The Nature of Explanation. *University Press, Cambridge,* (1943).
5. GALAMBOS, R. Inhibition of activity in single auditory nerve fibres by acoustic stimulation. *J.Neurophysiol.*, 1944, **7**, 287.
6. GALAMBOS, R. Neural mechanisms of audition. *Physiol. Rev.*, 1954, **34**, 497.
7. GRANIT, R. The Sensory Mechanisms of the retina. *Oxford University Press, Oxford,* (1947).
8. HARTLINE, H. K. The response of single optic nerve fibres of the vertebrate eye to illumination of the retina. *Amer.J.Physiol.*, 1938, **121**, 400.
9. HARTLINE, H. K. The discharge of impulses in the optic nerve of Pecten in response to illumination of the eye. *J.Cell.Comp.Physiol.*, 1938, **11**, 465.

10. HARTLINE, H. K. and RATLIFF, F. Inhibitory interaction of receptor units in the eye of limulus. *J.Gen.Physiol.*, 1957, **40**, 357.
11. HICK, W. E. Why the human operator? *Trans.Soc.Inst.Technol.*, 1952, **4**, 67.
12. HUFFMAN, D. A. A method of construction of minimum redundancy codes. Symposium on Communication Theory pp.102-110. Editor W. Jackson. *Butterworths, London*, (1953).
13. JACOBSEN, H. The informational capacity of the Human Ear. *Science*, 1950, **112**, 143.
14. JACOBSEN, H. The informational capacity of the Human Eye. *Science*, 1951, **113**, 292.
15. KUFFLER, S. W. Discharge patterns and functional organisation of mammalian retina. *J.Neurophysiol.*, 1953, **16**, 37.
16. MACKAY, D. M. and McCULLOCH, W. S. The limiting information capacity of a neuronal link. *Bul.Math.Biophys.*, 1952, **14**, 127.
17. MOUNTCASTLE, V. B. Modality and topographic properties of single neurones of the cat's somatic sensory cortex. *J.Neurophysiol.*, 1957, **20**, 408.
18. QUASTLER, H. Studies of human channel capacity. Third London Symposium on Information Theory. pp.361-371. Editor C. Cherry *Butterworths, London*, (1956).
19. SHANNON, C. E. and WEAVER, W. The mathematical theory of communication. *The University of Illinois Press, Urbana*, (1949).
20. SHOLL, D. A. The organization of the cerebral cortex. *Methuen, London*, (1956).
21. TOWE, A. L. and AMASSIAN, V. E. Patterns of activity in single cortical units following stimulation of the digits in monkeys. *J.Neurophysiol.*, 1958, **21**, 292.
22. UTTLEY, A. M. The conditional probability of signals in the nervous system. *R. R. E. Memorandum 1109*, (1954). Ministry of Supply, Gt.Malvern.
23. UTTLEY, A. M. The classification of signals in the nervous system. *E.E.G. Clin. Neurophysiol.*, 1954, **6**, 479.

DISCUSSION ON THE PAPER BY DR. H. B. BARLOW

DR. I. C. WHITFIELD: This is a very interesting hypothesis about the way the apparently large information input into sensory systems is handled. I think I am possibly one of the people referred to who does not agree with the estimates, which Dr. Barlow quotes, of what the informational intake is. The assumptions involved in obtaining such estimates appear to me to be unsound. There certainly is however a very large *potential* informational input, and I think we would agree that this must be handled in some way in the system, either by selecting the information which is going to be used, or else by re-packing it as Dr. Barlow suggests, with the reduction of redundancy.

The central problem seems to be in this question of unreliability in the units of the system. My own feeling is that there may be in the sensory pathway a selection of information at any given time, so that the available channels, effectively reduced in number through unreliability, can be utilized to concentrate on some particular aspect (which may well vary from time to time) of the stimulus situation. The smaller amount of information thus handled could be represented as a statistical picture in the response and one would have the observed result that any particular neurone does not respond in the same way to the same stimulus applied on successive occasions.

On the other hand, it may be, as I think Dr. Barlow suggests, that the differences are due to a continuous rearrangement of the coding, in which case we would have to look at a particular unit in terms of its relations to a large number of other units from instance to instance. I would like to have his view on what he thinks about this optimal coding system in relation to the apparent unreliability of operation of a particular unit, because it does seem to me that his hypothesis would argue a much more precise ability of the neurone to produce a particular pattern of impulses, than is required by the statistical hypothesis where the requirement may be no more than to fire or not to fire.

MR. J. T. ALLANSON: I think that Jacobson must have been one of the people to whom Dr. Selfridge referred when he said that whenever people reach the right conclusion, they have invariable done so from wrong premises. 8,000 bits/sec. may be the right sort of figure for the information capacity of the auditory nervous system, but, as far as I can see, the experiments which show that a man can discriminate so many frequencies

or so many amplitudes are not ones in which a man has been faced with one out of 1428 frequencies and asked which one of them it is. Leaving that on one side, because it is very difficult to find alternative ways of getting an accurate figure, the second point that I would like to make is in relation to the interpretation which Dr. Barlow puts on the data, in his section on the evidence from sensory neuro-physiology. It seemed to me, for instance, that Dr. Uttley could make a perfectly strong case in relation to his machine, that all the evidence here cited is evidence for economy not in terms of impulses, but in terms of the number of active units. If one postulates, for instance, that it is desirable to convert signals from a temporal pattern into a spacial one, regardless of the existence or not of an Uttley machine, then 'on' and 'off' units have essentially identical functions when thought of in spacial terms. One particular small group of fibres are firing or another small group of fibres are firing, as is shown in figure 4 on lateral inhibition, and one can indicate uniform activity over quite a wide region, simply by having those fibres active which mark the boundaries between activity and inactivity, and all the others inactive. Thus, one could regard this as evidence for quite a different type of economy, and I would like to ask whether Dr. Barlow can think of any sort of experiment which would enable one to discriminate between his hypothesis and some alternative one. It seems to me to be this is what we really require in this situation if any theories of this kind are to have value.

MR. J. WILSON: One of the suggestions in Dr. Barlow's paper was that coding mechanisms reduce orderliness in the sensory input by removing any departure from complete randomness. The coded message would be determined by the input, but it would look random and therefore be more efficiently coded.

I want to suggest an alternative way of looking at the same process. From my point of view, the sensory input is apparently not orderly but random, and the coding mechanisms are required to convert this apparently random signal into an orderly and understandable message. If the mechanisms have become adjusted to the statistics of the input signals, they can introduce this order by identifying familiar patterns in the input and giving as output a signal representing the comings and goings of these patterns. The output is objectively more random than the sensory input, but it is subjectively more orderly because its components have an established meaning in terms of past experience.

For Dr. Barlow, the detection of redundancy enables the sensory messages to be represented in more compact form; the suggestion here is that this compression of the message could alternatively be regarded as an insertion of meaning into an apparently random signal. Perhaps Dr. Barlow will comment on this.

DR. JOHN BROWN: Dr. Barlow suggests that there is recoding to reduce redundancy in the input. This seems highly probable but it also seems highly probable the recoding not merely compresses but also loses an enormous proportion of the information present in the sensory inflow. His suggestion that a signal is accepted if it has redundancy seems to be important and true. For example, if you look at a loaf of bread, the contour has quite a lot of redundancy but the microstructure of its surface has very little: what you tend to see is the outline and not the microstructure.

DR. A. M. UTTLEY: There is just one point in this paper which worries me greatly. Dr. Barlow has referred to the various papers at this symposium which were embarrassed by the same thing - that the number of units required for discrimination is impossibly large - 2^n is too large. None of these systems are embarrassed at all by the number of impulses arriving, and Dr. Barlow's proposal provides an economy in the number of impulses, but none at all in the number of channels. In the nervous system too, impulse frequencies are low compared with the capabilities of nerve fibres. It is in units and not in impulses that the nervous system must be economical.

DR. D. M. MACKAY: The reduction in the number of impulses might have another reason. If perception is a progressive internal matching-process *(ref. 1)* then what Dr. Barlow has described may represent the stage-by-stage reduction of the input to a null-detecting system. If so, it would not be so much the order or disorder in the impulses, but simply their presence or absence as an indication of mismatch, that would be operationally significant.

I would suggest that in order to make sense of perception we need not so much the quantitative notion of redundancy, as the qualitative idea of *redundant features*, also mentioned by Dr. Barlow. Perception is concerned solely with features that are in some sense redundant, to which the perceptive process can develop a matching response.

May I say in passing how good it was to hear mention of Craik *(ref. 2)* who was one of the most stimulating thinkers on brain-modelling. His book should certainly be read by all who are entering this field.

REFERENCES

1. MACKAY, D. M., In search of basic symbols. Trans. 8th Conf. on Cybernetics. *Jas. Macey Jr. Foundation, New York* (1951) p198. See also *Brit. J. Psychol.* 1956, **47**, 30.
2. CRAIK, K. J. W., The Nature of Explanation. *University Press, Cambridge*, (1943).

DR. W. K. TAYLOR: There is just one point which I would like to raise about the transmission of information along nerve fibres. I do not think that it can be usefully separated from the information that can pass through the other elements in the nervous system. In other words we must also know how much can pass from a fibre through a synapse to a cell and how much can pass from a fibre to a muscle and be recovered as overt behaviour. To give a simple example, if we have the system shown in *fig. 1.*

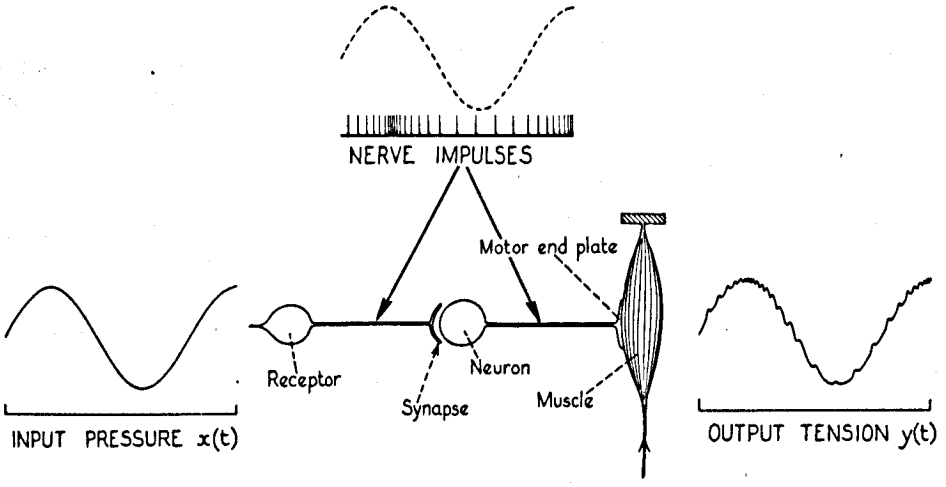


Fig. 1

supplied by an input $x(t)$ consisting of a sine wave superimposed on a steady level the output $y(t)$ is a tension that resembles the input. The intermediate nerve impulses have a high rate at the peaks of the input and a low rate at the troughs. The impulses are demodulated by the muscle, each impulse developing a tension of the form te^{-dt}. The overall channel capacity is limited by the effective bandwidth since if the imput frequency is increased the variations in tension get smaller. A second limitation is the modulation noise which is superimposed on the output sinusoidal component in an unpredictable way so that the input cannot be determined by looking at the output; we can only look at the fundamental and use the formula for continuous information transmission which depends on bandwidth and signal/noise ratio.

DR. M. L. MINSKY: I will try to make some remarks about what can be done to reduce N. I fear that I cannot say these things with sufficient precision in the limited time but I will outline the idea briefly. The

model that Dr. MacKay has in mind is very important here--this is the notion of a machine matching the input against some concept of hypothesis inside the machine and then saying 'yes' or 'no' according to the match. (Cf. MacKay *ref. 1*). It seems clear, introspectively, that our higher thinking is concerned with relatively small numbers of highly significant symbolic expressions, symbols, or the like, and that the large N of the neural point events in input stimuli is not handled directly in our abstract thinking.

Now we have been talking about the input information and it seems clear, again introspectively, that the huge quantity of receptor information must be reduced by the brain nets between the sensory organs and the higher centers--yet information relative to particular hypotheses is retained. How can a net do this--i.e., compute the appropriate functions of the input. Now when we talk about 'information' we are all indebted to things C. E. Shannon has said clarifying the notions of channels, sources, and capacities. I want to call attention to another piece of work by that same Shannon which may be at least as important to us as the channel capacity idea — this is the work on the size of a relay-net required to realize an arbitary n-variable Boolean function (*ref. 2*). It turns out (by use of methods similar to those in the proof of the channel capacity theorem) that 'almost' all Boolean functions of n variables require numbers of the absurdly large order of $2^n/n$ contacts in a realizing relay network and that the number of functions which can be realized with smaller nets is insignificant numerically. But these relatively few, peculiar, functions probably contain almost all the 'interesting' or meaningful' ones. They are essentially the functions which are made up of simpler functions by composition and other relatively simple connections. (See also McCarthy and Ashby, *refs. 3 & 4*). On the mysterious but compelling hypothesis that these are indeed the important functions, we can expect to find, I think, that the significant nets will be composed, in some sense, of layers, representing the levels of composition.

One kind of layer or processing level would be the kind that distinguishes between active and passive areas, e.g. handling functions of accommodation and gradient detection. This alone would yield a hugh reduction in N for pictorial information. But perhaps the most important

REFERENCES

1. MACKAY, D.M. The Epistemological Problem for Automata. Automata Studies, *Princeton* (1955).
2. SHANNON, C. E. Synthesis of two-terminal switching networks. *Bell System Technical Journal*, 1949, **28**, 59.
3. McCARTY, J. Automata Studies, *Princeton* (1955) p.177
4. ASHBY, W.R. " " " " p.232

kind of 'layer', from this point of view, would be those responsible for what we call 'attention'. The higher brain centers may have an idea about some part of the stimulus and so they send down a signal to a 'horizontal' net which says 'just let through the signal from such and such a region'. Thus we can turn our mental eye on some particular small part of the stimulus, and perhaps match it as Dr. MacKay has proposed. You see that then N is greatly reduced, not by any clever processing net which performs some heroic redundancy-elimination feat, but by only treating those parts of the signal which have some a priori interest with respect to the state of mind of the rest of the system.

DR. A. M. ANDREW: I would like to ask Dr. Barlow what sort of time scale he has in mind for the adaptive changes in the type of coding. If the changes are reasonably rapid, and if the coding is to be reversible, it is necessary to inform the central nervous system of the changes in code. In *fig. 1* the sense organ is represented by the box marked "coder". The requirement of reversibility implies that another device to decode and recover the original signal is possible in principle, and in *fig. 1* there is an auxiliary channel to signal the changes in type of coding.

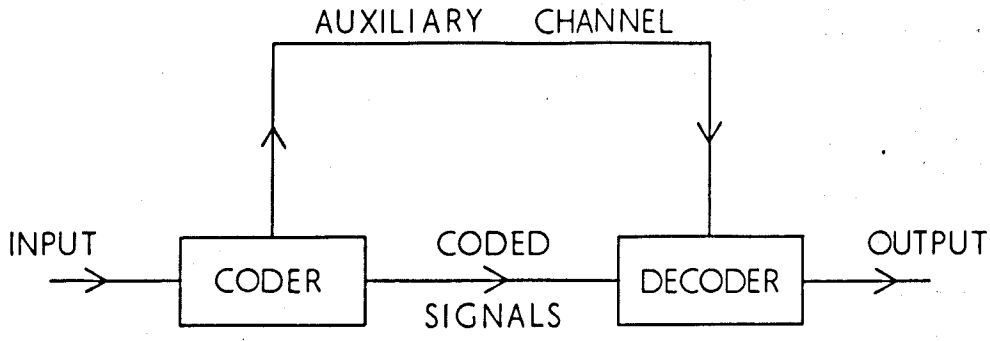


Fig. 1

It is possible to dispense with the auxiliary channel; as in *fig. 2*, if the changes in code are such that they could have been determined by statistical information collected from the *output* signals of the coder, as represented by the broken lines in *fig. 2*. These signals are available to the decoder, so it can produce corresponding changes in its system of decoding.

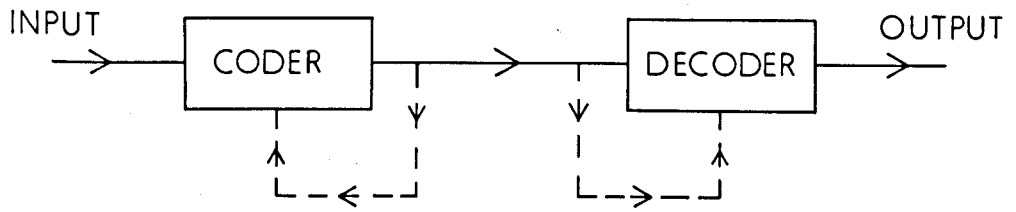


Fig. 2

A coding device operating in the above fashion could operate so as to remove redundancy completely, provided the redundancy remains unchanged in type and amount for a sufficient time. In that case, however, the operation of both coder and decoder would depend on quantities which were integrated over all past experience, and consequently the decoder could get out of step with the coder if there was a temporary break in transmission between them. A more robust kind of system could be made by arranging that the coder does not entirely eliminate the redundancy, but allows its coded signals to retain some redundancy of the same kind as was present in the original signals. In this way it is possible to have a coder and decoder which will automatically come back into step following a disturbance.

These considerations give a neat explanation of the behaviour commonly shown by sense organs, illustrated in *fig. 3*. In this θ represents some input quantity affecting a sense organ (e.g. the angle between two bones of a joint, as in Boyd and Roberts, *ref. 1*) and f the discharge frequency of the organ, plotted on the same time scale. The form of the upper curve can be explained by supposing that during AB the form of coding is that produced by a sustained low value of θ. When θ changes to a higher value, the change in f is large at first, since the form of coding is not altered until f has been at a higher value for sufficiently long. Then as the new type of coding comes to be applied, the value of f falls away, but never quite comes back to the value during AB. Following the change at C the inverse effects are seen.

REFERENCE

1. BOYD, I. A. and ROBERTS, T. D. M. Proprioceptive discharges from stretch-receptors in the knee-joint of the cat. *J. Physiol.*, 1953, **122**, 38.

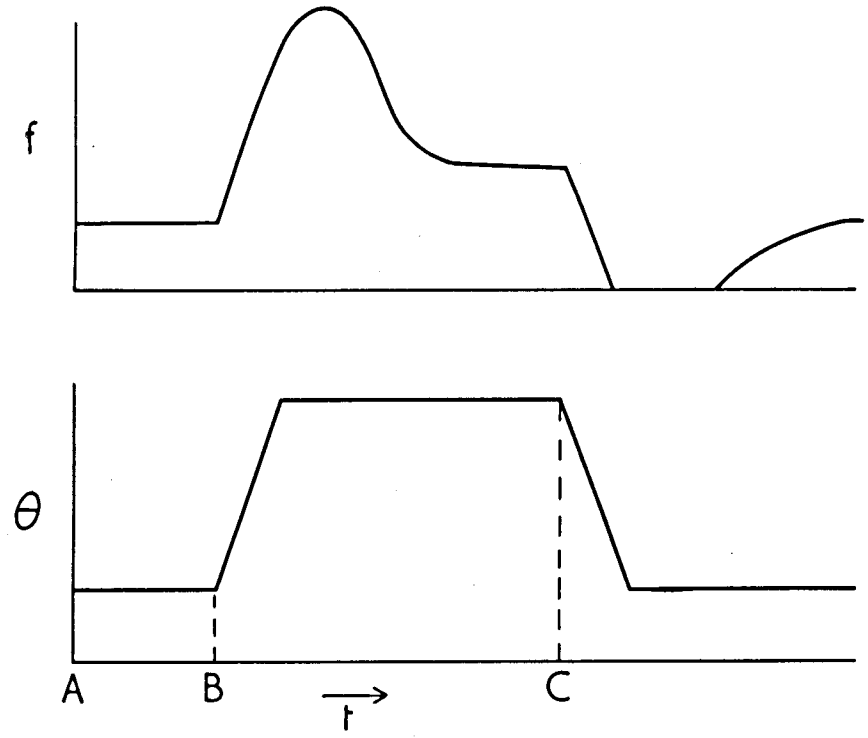


Fig. 3

The theory is fairly trivial when applied to a simple sense-organ, but I think it can be extended to other kinds of redundancy, and in particular to redundancy between channels.

MR. D. W. DAVIES (written contribution): The coding device described by Dr. Barlow can be extended, in principle quite easily, to any number of channels.

Given n channels, we first require a decoding tree with 2^n outputs A_i. Each of these outputs is associated with a probability P_i. Also we require recoding network with 2^n inputs B_i and n final outputs. A reversible (i.e. non-information-losing) coder will be constructed by connecting the 2^n outputs A_i to the 2^n inputs B_i in a one-one correspondence.

To optimise this coder for a given set of probabilities P_i we proceed thus: list the 2^n output terminals A_i in order of increasing probability and list the 2^n input terminals B_i in order of decreasing 'weight', where the weight is the number of 'ones' in the output combination on the n final outputs. Associate corresponding members of the two lists.

There are $(2^n)!$ different reversible codes, and most of these, for n not small, are impossible to produce without almost the full decoder-recoder arrangement. However, for any given set of probabilities there are many equally good coders, because the 2^n input terminals B_i have only $n + 1$ possible weights, so that the list in order of decreasing weight is not unique. In any case, exact optimization is not important. In consequence reasonably economical coders might be possible.

It might be supposed that these more complex optimum coders could be constructed by joining together sets of smaller optimum coders. In at least one case this can be disproved. The optimum coders for two channels cannot be joined together to make any required optimum coder for three channels.

All the reversible coders for two channels are linear, that is to say the outputs are sums of the inputs, modulo 2, with a possible constant term ($+_2$ 1). Any network of these coders will also be linear.

There are 8! reversible codes for 3 channels. For a given set of probabilities p_i, the permutation of those terminals B_i which have equal weights gives 1 ! 3 ! 3 ! 1 ! = 36 different codes, all optimum. There are therefore 8 !/36 = 1120 classes of codes such that all the members of a given class are optimum together.

There are 168 non-singular 3 x 3 matrices composed of 0s and 1s, addition being modulo 2. These, allowing for presence or absence of a constant term ($+_2$ 1) in each of 3 equations, give 168 x 8 = 1344 reversible codes. The six different codes obtained from one of these by permuting the output variables all belong to the same class, because they have equal weights for any given input combination. Therefore, linear reversible codes are found in at most 1344/6 = 224 classes. For all the other classes there is no linear optimum code, and hence none can be constructed from assemblies of reversible two channel coders.

An interesting question is: Are there reasonably economical codes in each class? By economical, we mean less equipment than a full decoder-coder needed to produce them.

DR. H. B. BARLOW (in reply): To Dr. Whitfield. The figures I quoted in the section on Sensory Inflow are estimates of the *capacity* of the sense organs, and I agree that it is misleading to regard them as estimates of the actual rate of flow of information into the central nervous system.

I have not thought much about the reliability of the components. They seem to be much more reliable when your physiological preparation is in really first class condition, and I suspect that their unreliability is usually overestimated. I agree that one needs some redundancy of the internal representation of sensory information to protect it from disastrous changes as a result of whatever degree of unreliability is present,

but the type of redundancy required for this is not necessarily the same as that in the incoming messages, so you will certainly need to juggle around with redundancy, even if you don't eliminate it altogether.

Unreliability of the components, or intrinsic noise, is sometimes lumped together with unreliability in the messages themselves, or extrinsic noise, but it is important to distinguish them, both because the methods of dealing with them are different, and because we are uncertain about intrinsic noise whilst we know that extrinsic noise must be a constant problem in sensory discriminations.

To Mr. Allanson. I think I shall leave the point about the capacity of the input, except for remarking that Jacobsen approached the problem as if the eye and the ear were simple physical instruments. This is probably justifiable if you are interested only in the peripheral organs, but it would of course be wrong to assume that the whole auditory and visual pathways had the same capacity.

With regard to the types of economy, I would say that economy of *active* units is economy of impulses, not of units. Because you have eliminated the impulses from all but a small group of units, you cannot conclude that the inactive units themselves could be eliminated, for they may be required for the response to a different sensory stimulus. I may have misunderstood the question, but I would say that, if you succeed in confining the activity to a small number of units, that is a good example of economy of impulses, and it is exactly what I had in mind with regard to adaptation, the existence of 'on and off' units, and lateral inhibition.

The final question was whether there is a crucial test of a hypothesis like this, and of course I have been trying to think of one. I think one could devise experiments to see the conditions under which one gets lateral inhibition, and the conditions under which it fails to develop in the eye, to see whether they fitted in or not. Lateral inhibition in the vertebrate eye is not constantly present--it is only present after light adaptation. Light adaptation means that in the recent past you have activated together a large number of the retinal units; you put correlated activity on to those units, and the result is the development of lateral inhibition. This all fits in very nicely with the hypothesis, but that is hardly surprising because I knew the phenomena before I put forward the hypothesis. It would be interesting to see what happened if the eye was light-adapted by a "noisy", non-uniform field, so that activity in neighbouring regions was less correlated, and by other means define the conditions for the appearance of lateral inhibition more precisely.

Another possible point for attack would be the thalamic and cortical sensory neurones which people are beginning to record from successfully.

One can tickle the skin, flash lights, make noises, and so on, and thus determine the sensory stimuli to which activity in a particular neurone corresponds. Now if the hypothesis is correct, it should be possible to change the correspondence between the two by changing the redundancy of the previous sensory stimuli. There are, however, great technical difficulties in getting physiological preparations which are stable for long enough periods, are free from the effects of anaesthetics, and so on.

To Mr. Wilson. I think I agree with what Mr. Wilson has said. It appears paradoxical to talk of "eliminating redundancy" and producing a more "random" output, when we are all agreed that the redundancy of the input is crucially important, and that what one wants to do is to bring order to the incoming sensory messages. But as Mr. Wilson points out, objectively random corresponds to subjectively ordered so the paradox is resolved. However, to talk of the compression of messages as "the insertion of meaning into an apparently random signal", though illuminating in one sense, might be misleading in another: it rather suggests that the meanings being inserted are imposed from outside, whereas in fact they are derived from previous sensory messages.

These thoughts on the significance of decreased orderliness (objective) of sensory messages were partly started by observing the gross irregularity of impulse discharges recorded from more central points in sensory pathways. When you are recording from peripheral nerve fibres you usually get trains of impulses which are rather regular, and indeed in the early days such regularity was one of the principle means of making sure that one was recording from a single unit. When you record from more central pathways this regularity is often missing. For instance, if you are recording from a ganglion cell in the retina, which is the third link of the chain of excitable cells connecting photosensitive substance to the brain, you might find a mean frequency of impulses of 40/sec; on measuring up the intervals, which average 25 msec, they might vary between 3 and 300 msec. There is an enormous scatter of impulse intervals, and the discharge looks very irregular. Before jumping to the conclusion that some noise has been added, I think one should consider the possibility that some skilful coding has gone on in the retina, so that the output looks more irregular, but simply because the order has been detected and removed.

To Dr. Brown. I agree that a lot of information must be lost in coding. I feel sure that this is so, and for this reason it may be preferable to talk of "increasing the relative entropy of the signals", rather than "optimum coding", since the former does not exclude the loss of information, whereas true optimal coding does.

In the paper I suggested that such recoding was as important in the organisation of sensory information as the formation of the image was in

the function of the eye; you obviously cannot begin to understand vision until you understand how the eye focusses light. But there may be this important difference, that whereas the eye is optically quite efficient, considering the size of the pupil, the recoding operations are almost certainly grossly inefficient by absolute standards. But this does not, of course, mean that they are unimportant.

To Dr. Uttley. I think this is an important point, but I don't agree that economy of impulses is no help, so I first want to put the point in another way. One of the limits to the capacity of a nerve fibre is the maximum sustained frequency of impulses that it can carry. The mean frequency it is actually carrying defines another capacity, lower than the first, and economy of impulses reduces this utilised capacity still further. The number of possible states of the system is 2^{mn}, where m is the capacity of each of the n channels, so a reduction in the channel capacity is as effective as reduction in their number in reducing the number of units required for complete discrimination. The important question is whether one is justified in using the utilised, rather than the maximum, capacity; I do not feel certain about this, but economy of impulses will at least make each individual fibre declare its utilised capacity to the recipient apparatus.

Secondly, I want to show one immediate benefit of economy of *impulses* in economising *units* required in a conditional probability machine (see also section 3(a) of paper). If the input is uncoded, a combination of many inputs may occur frequently because they may be correlated, and the frequency of impulses may be high. After coding to economise impulses, combinations will occur with a frequency tending towards the product of the probabilities of the individual inputs, so for combinations of many inputs it becomes exceedingly low. A legitimate economy in designing a conditional probability machine to be fed by such an input would be to omit units responding to simultaneous activity in many inputs, and units for such high order combinations form the vast majority of all units required for complete classification.

If there is sufficient redundancy in the input, economy of impulses might convert a set of n inputs in which all combinations of activity *could* occur into a set of n outputs which *do* occur only singly, one at a time; the outputs have been made mutually exclusive. As a result the number of units required for complete discrimination is reduced from 2^n to n --- but of course we have not counted up the units required for recoding.

To Dr. Mackay: I think I have been a bit slow in appreciating that Dr. Mackay's internal matching process is closely related to the reduction

of redundancy. The development of lateral inhibition could be nicely described as a matching response: first you impose uniformity on the retina; the retina learns to "match" the uniformity, and thereafter signals lack of uniformity. This is all done in the retina itself, by the way, because it occurs after the higher centres have been destroyed.

Although the idea of a matching response gives a vivid picture of how a particular redundant feature can be tackled, I feel that the reduction of redundancy, or the increase of the relative entropy of the signals, is a more fundamental description of the process we are interested in. Also it suggests how perceptual tasks can be broken down into small scale operations performed in series and parallel (though limitations may appear if it is so broken down -- see Dr. Minsky's reference to Shannon's work on switching networks, and Mr. Davies's contribution). Such small scale operations might be performed by single neurones or a small group of them, and if one could catch them at it one would be a step nearer to understanding perception.

To Dr. Taylor. Yes, one would be badly astray if one based an estimate of the informational capacity of a muscle-controlled movement upon the number of nerve fibres used and their maximum impulse - carrying capacity. I expect there are bottlenecks to information flow in the sensory pathways, analogous to the demodulation effect of a muscle which Dr. Taylor has pointed out, but they do not seem to have the same fixed character. It is as if they were deformable and moveable (though perhaps always of the same cross-sectional area): this plasticity makes it difficult to investigate them.

To Dr. Minsky. I agree that Dr. MacKay's matching responses are relevant, though I prefer a description in terms of coding to reduce the redundancy, for the reasons already given (see answer to Dr. MacKay, above).

I am not sure that the suggestion of a variable field of attention adds very much, except that it emphasises that information may be lost in parts of the sensory inflow not attended to. What are the criteria for shifting the attention? If it is to search for something matched by an internally stored concept or hypothesis, then shift of attention corresponds to the execution of a particular tactic in the search for signals of high relative entropy. If the criteria are different, then we must know more before the idea is helpful. Words and phrases, such as "attention", "mental eye", "a priori interest", and "states of mind", are not easy to correlate with experimental facts in psychology, neurology, or physiology, because they refer primarily to our subjective experience of the workings of our minds. I believe one should use an objective terminology, and if you translate Dr. Minsky's two proposals for reducing N, I think they both turn out to be methods of reducing redundancy.

The reference to Shannon's work on switching networks seems to be highly relevant to the problem of reducing redundancy in stages. May it be that we can only grasp complicated logical functions if they are separable functions? This should be susceptible to experimental testing.

To Dr. Andrew. First the time scale of adaptive changes in coding. A large part of the change in sensitivity of the eye associated with light and dark adaptation occurs within a few seconds of changing the mean level of illumination, and I would think that part, at least, represents a rapid change of code. It takes about 1/2 minute exposure to a moving field of vision to cause the appearance of reverse movement of stationary objects. To adapt to inverting spectacles takes about 10 days. To "adapt" so thoroughly to the stylistic peculiarities of, say, Vivaldi, that a work of his could be instantly discriminated from those of his contempories, might require experience that could not be attained in less than a month or two, and in the case of a musically naive person, much longer. There seems, in fact, to be a vast difference in the time scale of adaptation to simple and more complex features of sensory stimuli.

The point about transmission of the *code*, as opposed to the messages, is an interesting one, and I like the idea that a residue of redundancy is left in the messages in order to signal what has been removed. It would be a nice explanation of the type of result exemplified in Dr. Andrew's fig. 3. But sometimes you find that the maintained change of discharge rate is in the opposite direction to the transient change. This can be observed in the cat's retina *(ref. 1)*. Also, from the fact that one gets illusions, that is, false impressions, following the cessation of adapting stimuli one suspects that the more central parts of the nervous system are not always kept adequately informed about code changes at lower levels. I don't understand this: when one stops walking, the after image of the movement of the ground should give one the impression of moving backwards, but I have never experienced this myself, nor heard it reported. Is it the absence of other sensory correlates of movement during the exposure that allows one to see the after image of movement after looking at a waterfall, when sitting in a train, or after gazing at a rotating spiral in the laboratory?

To Mr. Davies. Mr. Davies has gone further than I have in considering the limits of assemblies of two channel coders. If recoding in the nervous system is done serially and in small stages, then analogous limitations would be expected, and might be detectable as breakdowns of perception.

REFERENCE

1. KUFFLER, S. W., FITZHUGH, R. and BARLOW, H. B. *J. General Physiol.*, 1957, **40**, 683.

SESSION 4A

PAPER 2

STIMULUS ANALYSING MECHANISMS

by

DR. N. S. SUTHERLAND

BIOGRAPHICAL NOTE

Dr. Sutherland read Greats at Oxford before taking a second honours degree in the newly established school of Psychology. He was elected to a Fellowship by examination at Magdalen College, Oxford, (1954-58). At present he is engaged on a project on stimulus analysing mechanisms financed by the Nuffield Foundation at the Institute of Experimental Psychology, Oxford. His main interests in psychology are in the mechanisms which underlie perception and learning.

STIMULUS ANALYSING MECHANISMS

by

Dr. N. S. SUTHERLAND

SUMMARY

TWO distinct approaches to the problem of stimulus analysing mechanisms in organisms are outlined. The first, often adopted by engineers, is to assume that there is a very general analysing system at work for each modality which in principle is capable of categorising stimuli in all possible ways. The second approach is to assume that specific analysing mechanisms are at work and that stimuli can only be categorised in a limited number of ways by these specific mechanisms.

If the first sort of system were correct it would be impossible to make predictions about how stimuli would be categorised based merely on the system: such predictions could only be made from a knowledge of previous inputs to the system. Evidence is produced to show that it is highly probable that some stimulus analysis is performed by specific mechanisms, at least in sub-mammalian organisms where many categorisations seem to be innate. The evidence for mammals is inconclusive, but suggestions are made for how it would be possible to test for the existence of the general type of analysing mechanisms in mammals. The existence of specific analysing mechanisms would not only be economical in terms of the number of nerve cells required, but it would account for many facts of animal and human behaviour, particularly if we envisage the possibility that the same specific analysing mechanisms may actually be used in different ways of categorising stimuli in one sensory modality and that some of the analysing mechanisms may be common to more than one modality. Thus competition for specific analysing mechanisms would explain why the human being seems to function as a single information channel, it would give a rationale for recent neurological findings on the peripheral blocking of incoming stimuli, and it would account for findings on animals which suggest that animals learn not merely to attach a response to a stimulus but also learn which analysing mechanisms to switch in on a given occasion. Some of the difficulties in using behavioural evidence to set up hypotheses about specific analysing mechanisms are discussed, and the difficulty of deciding what sorts of coding mechanisms are at work in the central nervous system is discussed

with reference to a particular example of a simple discrimination. It is suggested that engineers might profitably turn their attention to the design of specific analysing mechanisms intended to account for some of the ways in which animals are known to classify stimuli.

INTRODUCTION

THIS paper will be concerned with the problem of how ideas based on engineering concepts can best be applied to help solve some of the problems which arise when we consider how patterned stimuli are classified by organisms. The paper will deal mainly with visually presented patterns.

There are two different kinds of theoretical mechanism which have been proposed to explain what organisms do. The first kind is one which attempts to give a very general explanation of what organisms do by postulating a broad model or type of connectivity, and then pointing to some very general features of the model's behaviour which are said to be shared by animals. Examples of this type of theory are those of Hebb (1949, ref.20), Uttley (1956, *refs. 63, 64*), Taylor (1956, *ref. 61*), Ashby (1952, *ref. 1*). The second kind of theory is of a more specific sort and is designed to explain a limited range of behaviour by putting forward a model whose parts are highly differentiated: examples of this kind of model are those put forward by Deutsch (1955, *ref. 10*), Dodwell (1957, *ref. 11*), and Sutherland (1957, *ref. 55*) to explain certain features of shape recognition, or by Tinbergen (1951, *ref. 62*), to explain certain features of instinctive behaviour. In the specific type of mechanism, the parts are arranged in a highly systematic way. In this sort of mechanism, information contained in the input can be selectively lost so that this sort of system is potentially more economical than general analysing mechanisms. This gain in economy is offset by a loss in flexibility. Wherever the classification of stimuli can be shown to be innate, the existence of a specific analysing mechanism is implied. Specific analysing mechanisms could, however, arise out of an initially randomly connected network as a result of learning: a small initial bias in part of the system could in theory lead to a highly specific system being developed out of an initially randomly connected network. For example, Sutherland (1957, *ref. 55*), has suggested that the tendency of the octopus to move its head up and down while viewing shapes could lead through a learning process to the development of a visual analysing system in which the vertical extents of shapes are counted at different points on the horizontal axis, while other information is lost.

In what follows the two types of theory will be discussed in detail, and an attempt made to specify their characteristics. The rather general type of system will be considered first, and it ought to be said at the

outset that I am perhaps prejudiced against this type of system. The rationale of this prejudice will be made clear later. In general it is typical of the first sort of system that it is put forward by engineers who know little about psychology, it is typical of the second sort that it is put forward by psychologists who know little engineering: Hebb is an exception to this rule.

CHARACTERISTICS OF GENERAL ANALYSING SYSTEMS

If the neural mechanisms mediating stimulus classification are highly unspecific, this sets a severe limitation on the possibility of discovering what the neural mechanisms are and working from knowledge of the mechanisms to predictions about what animals will actually do. This can be illustrated with reference to Hebb's theory of shape recognition. According to Hebb the connections from the primary visual projection area in the brain are initially random: thresholds at synapses will presumably vary in a random way from moment to moment according to such factors as how recently the post-synaptic nerve cell has fired. This means that when a given shape is first projected onto the retina, it is impossible to predict what output the system will give: initially any shape is as likely to produce a given output as any other. However, once a shape has given a certain output the chance of its giving the same output on the next occasion of presentation will be increased (provided it stimulates the same retinal cells), since it is postulated that the probability of a given synaptic connection being used increases every time it is in fact used. Moreover, given that a given shape falling on a given part of the retina excites a given cell assembly as a result of an increase in the probability of transmission at specific synapses, any shape projected onto the retina in close temporal contiguity with the initial shape should come to excite the same cell assembly through the mechanism of spatial summation in the nervous system: of the random connections which the new shape might excite the ones it will be most likely to excite are those already excited by the initial shape. Thus the two shapes will come to have a cell assembly in common - they will tend to give a common output. Since the same shape will be constantly being shifted across the retina from moment to moment, this means that the same shape projected to different parts of the retina should come to give the same output.

An interesting paradox arises. When Hebb's theory was first put forward it was hailed as showing how it might be possible to account for behaviour in terms of plausible neurophysiological mechanisms: it was thought that Hebb had demonstrated the possibility of explaining the findings of psychology in physiological terms. However, a moment's reflection shows that, if he is right, what he has really succeeded in doing is to

demonstrate the utter impossibility of giving detailed neurophysiological mechanisms for explaining psychological or behavioural findings. According to Hebb the precise circuits used in the brain for the classification of a particular shape will vary from individual to individual with chance variation in nerve connectivity determined by genetic and maturational factors, they will vary within the individual with chance variations in the threshold at synapses at times when a given shape is first seen and during the succeeding presentations, and they will vary according to the frequency and temporal order of shapes projected onto the retina when learning is occurring. This means that even if we knew the precise sequence of shapes an animal has been subjected to in its previous history, it would be impossible to translate the effects of that sequence into actual brain circuitry and then work from the brain circuitry to predictions about subsequent behaviour: the circuits will be so complex, so scattered over different parts of the brain and, above all, they will vary so much from individual to individual that trying to take the intermediate steps for translating the effects of early experience of shapes into actual brain circuitry becomes an impossibility. Different individuals will achieve the same end result in behaviour by very different neurological circuits. If we wish to make predictions about individuals we must concentrate on correlating differences in early environment with differences in later behaviour, and translating these differences in early environment into differences in brain circuitry and then working from there to predictions about subsequent behaviour becomes impossible. If Hebb's general system is right, it precludes the possibility of ever making detailed predictions about behaviour from a detailed model of the system underlying behaviour. This seems to me a most unfortunate consequence since the explanation of complex phenomena in terms of a simpler system is something which is intellectually satisfying, and which is in some ways more exciting than the working out of statistical correlations between early stimulus sequences and subsequent behaviour. This is clearly not a reason for rejecting the account given by Hebb, but it is certainly a good reason for looking to see whether it is possible that the initial system is less randomly organised than Hebb supposes.

The same sort of consideration would apply if Taylor's model for classification were correct or if Uttley's were. Taylor is in fact trying to discover whether a system of the general sort Hebb proposes would have the properties Hebb attributes to it, by actually building a model of it. This is one of the cases in which building a physical model of a system is useful, because it is impossible to predict whether or not Hebb's system will have the properties he attributes to it, and it would be impossible even if some of the variables in the system were precisely specified because the equations necessary for a solution would be too complex to solve. Thus as more and more shapes occur together on the retina it seems

possible on Hebb's system that all cells will ultimately be connected up with all other cells so that any shapes will eventually fire all cell-assemblies: Milner (1957 *ref. 34*) has recently proposed a most ingenious solution to this problem, but in order to do so he has to assume considerably more specificity in the arrangement of cortical cells than Hebb envisaged.

In the discussion of Hebb's theory so far, no account has been taken of the specific mechanisms which he alleges may be present from birth. The most important of these is the alleged tendency of animals to fixate successively the corners of figures, which in the case of rectilinear figures would result in the scanning of the contours. This mechanism would result in the equivalence of figures of different sizes since if corners are successively fixated the successive patterns at the fovea will be identical for the same shape irrespective of size and also eyemovements will occur in the same directions though they will be of different lengths. Unfortunately Hebb never bridges the gap between recognition occurring in this way and recognition occurring without eye movements and when shapes are projected to different parts of the retina, although it is quite certain that human beings are capable of recognition under both these conditions (Collier 1931, *ref. 7*). It should be noticed that in order for the eye movements to occur at all there must be a specific mechanism in the brain for reading off the position of a point relative to another in order to send the appropriate message to the eye muscles, and this is already a considerable limitation on the randomness of further connections from the primary projection area. If such a specific mechanism exists then logically it might play a part in shape recognition without the eye movements occurring. Since this is a limitation on the operation of a general mechanism which could according to training discriminate any shape from any other shape, and since evidence about the role of eye movements in shape recognition is almost entirely lacking, it will not be further discussed here. The possibility of using the way information is coded to determine some primitive response (such as fixation) to throw light on possible coding mechanisms in use to perform more complex functions (such as shape discrimination) will, however, be further discussed below.

Uttley (1956 *refs. 63 & 64*) has proposed a system of classification based on the principles of set theory. Input units are connected to output units in such a way that each possible combination of input units is connected to one output unit: whether or not a given combination of input units is firing can then be detected by reading off whether the appropriate output unit is firing. With some limitations, the same result can be achieved with an initially random connection of input and output units. Such a system will not mediate generalisation where the generalisation involved is not from a given input pattern to a pattern of which it is a subset. Uttley postulates a further system which works out conditional probabilities of one unit firing given that another unit has fired: connections between

units are altered in such a way as to represent conditional probabilities. Such a system could account for generalisation occurring between patterns one of which was not a subset of the other. Once again, the detailed model would be of little use in prediction: we could predict only by knowing the details of the previous inputs to the system, and our predictions would not involve following transformations of an input through a mechanism to arrive at an output, but merely considering previous inputs in terms of their class relationships of inclusion and exclusion and the working out of conditional probability relationships between different sets.

It should be noticed that all three theories make the assumption that the actual classificatory mechanism used by animals arises wholly as a result of learning. This is not accidental: in any general system of this kind where initial connections are random or are arranged in all possible ways, it would be very hard to account for any innate classification. For example on Hebb's system it cannot be determined at birth what cell assembly will be used for what classification: thus it would be impossible genetically to specify any connections which would lead to a given response being given to a given stimulus. Even if some such system as Hebb, Taylor and Uttley suggest is at work in parts of the brain - and from the degree of perceptual relearning that can occur in human beings as evidenced by studies with distorting lenses Kohler (1951, *ref. 24*), it seems likely that a system of this sort may operate - it would be extremely interesting to discover what its limitations are: these limitations indicate non-randomness of connections and therefore suggest analysing mechanisms of the specific kind at work which will explain and predict behaviour and which do not vary from individual to individual in a given species. It is important therefore to examine the evidence for stimuli innately producing specific reactions since any examples of this must severely limit the applicability of the general systems I have been describing, and open the way to the postulation of more specific systems.

EXPERIMENTAL TESTS FOR GENERAL ANALYSING SYSTEMS

Innate Stimulus-Response Connections

Evidence has been accumulating that at least in many sub-mammalian organisms there is a considerable degree of specificity in the classificatory mechanisms at work. The studies cited by Tinbergen (1951 *ref. 62*) on innate fright reactions of gallinaceous birds to a —+— figure moving in the direction of the short arm, and on other innate releasing stimuli often of a complex configurational kind are well known. Unfortunately, there is still some doubt about their validity because often these reactions were not studied under strict laboratory conditions and in only a few cases were precautions taken to exclude the possibility of the reaction coming

into being as a result of early learning. However, more recent studies are not open to these objections and have confirmed that some classification can occur without learning. Thus Fantz (1957, *ref. 15*) has shown unequivocally that dark reared chicks exhibit a preference in their pecking behaviour for round objects as opposed to square or triangular ones, under conditions where the preference could not have been influenced by rewards and punishments. Wells (1958, *ref. 66*) has shown that newly hatched Sepia have a very specific preference for attacking Mysis, a small crustacean with a complex and specific form: although unrewarded for their attacks, the latency of attacks decreased with the number of attacks made, and it was difficult to persuade them to attack any other shape. Rheingold and Hess (1957, *ref. 43*) have shown that chicks' preference for water is determined by its visual properties, and that chicks' rely upon the same visual properties before and after experience with water.

By fitting chicks' with prisms, Hess (1956 *ref. 21*) has demonstrated that the direction and distance at which a chick will peck to a stimulus are both innately determined. Some years ago Sperry (1942-4, *refs. 51, 52 & 53*) demonstrated that fly catching behaviour in the newt was determined by a highly specific neural organisation: if the eye ball was rotated through 180^{0} and the optic nerve severed and allowed to regenerate, reactions to objects moving in the visual field were made to a position 180^{0} from the position of the stimulus in the visual field. If the severed optic nerve can re-establish connections with such precision despite their biological uselessness to the animal, this suggests that the original connections may have had the same degree of specificity. Thus there can no longer be any question but in many submammalian species there is a degree of specificity present in the arrangement of the connections in classificatory mechanisms which means that theories of the general type cannot satisfactorily account for stimulus classification in these animals.

Perceptual Deprivation

The situation with regard to mammals is still not resolved. Several investigators-- Riesen (1949-51, *refs. 44, 45 & 46*), Riesen, Kurke and Mellinger (1953, *ref. 47*), Riesen and Mellinger (1956, *ref. 48*), Chow and Nissen (1955, *ref. 6*) have demonstrated that some mammals (cats and chimpanzees) discriminate less well between visually presented shapes if they have been brought up without prior experience of patterned light. Although this is sometimes taken to mean that these animals have to learn to classify stimuli or more specifically to build up specific connections out of initially random ones, this conclusion is far from forced upon us by the evidence. Thus there are at least four alternative explanations for why animals brought up without pattern vision should learn a given visual discrimination less readily

than a normally reared animal. These are: (1) Possible degeneration in the system due to lack of use: it is impossible to know whether this has been eliminated even where animals have been brought up in diffuse light rather than total darkness. (2) The disruptive effects on learning of emotional responses given to completely novel stimuli: Miller (1948, *ref.* 33) found that rats brought up in the dark and trained on a maze habit in the dark actually performed worse when run in the light. (3) The possibility that there are specific connections present initially, but that they become more specific only with use, and irrespective of how they are used. This possibility is underlined by the experiments of Wells (1958, *ref.* 66) and Hess (1956, *ref.* 21): Hess found that the pecking response increased in accuracy with use, even in the case where because of distortion introduced by a prism in front of the eye the increase in accuracy merely led to the chick pecking consistently in the wrong spot. It is also suggested by an experiment of Chow and Nissen (1955, *ref.* 6): they found incomplete interocular transfer where chimpanzees had been brought up with one eye receiving pattern vision, the other diffuse light: this suggests that the anatomical overlap of fibres from opposite eyes is at first incomplete, but that it becomes complete with use. The conditions necessary for its completion, however, do not include knowledge of results since it is completed if patterned light is given alternately to either eye. Riesen and Mellinger (1956, *ref.* 48) obtained a similar result with cats. (4) The possibility that increased learning time for a visual discrimination is brought about because animals have learned to switch in analysing mechanisms for other sensory modalities through previous experience: animals might switch off the input from vision when it first appeared because it could not initially be useful in solving a problem and would only interfere with a solution achieved through some other sensory modality. Recent results on the peripheral blocking of sensory input from a modality to which the animal is not attending (e.g. Sharpless and Jasper 1956, *ref.* 50) underline this possibility. These results will be referred to in more detail below.

The upshot of this is that at the moment there is no reliable evidence on the extent to which the classificatory system is of the extremely general sort suggested by Hebb and others in mammals, and on how far the actual classificatory system used in later life develops only as the result of learning. If mammals have few innate responses to stimuli and if the type of experiment which has been performed on perceptual learning to date is inconclusive, it must be asked whether it would be possible to obtain evidence which would help to decide how far the classificatory mechanisms at work in mammals were of the very general sort proposed by Hebb, how far they were more specific.

It seems to me that it is possible to perform experiments designed to test this, and we are planning such experiments as part of a programme on stimulus-analysing mechanisms at Oxford financed by the Nuffield Foundation. A number of different possible approaches to this problem are outlined below:

(1) If the classes of shapes which will be categorised together are determined as a result of learning, we might expect that animals brought up without visual experience will either exhibit very different kinds of transfer from adult animals or will exhibit no transfer to new shapes at all. For example if the equivalence between rectangles in different orientations is determined through learning, we would expect that an animal brought up without pattern vision and then trained to discriminate between a square and a horizontal rectangle would show no transfer to a vertical rectangle. Similarly we would expect no transfer from a triangle in one orientation to a triangle in another orientation. Such experiments must, of course, be very carefully controlled: in particular if an approach-avoidance habit is being learned it is necessary in transfer to prevent the animal performing correctly by avoiding or approaching the figure which remains constant (i.e. the square in the first example). The method of successive discrimination avoids this difficulty since only one shape is shown on one trial and the animal learns either to approach or avoid it. Again it is necessary to be certain that what is transferred is a method of classifying the shapes and not the habit of attending to the shapes: this possibility can again be eliminated by proper controls (e.g. comparing transfer to a series of different shapes only some of which bear any resemblance to the original: the degree of transfer to shapes bearing no resemblance to the original then gives a base line for measuring genuine transfer due to a method of classifying the shapes). If the order of ease of transfer to different shapes were markedly different from that found in an adult animal, this would provide very good evidence for the importance of learning in determining the ways in which shapes are classified and hence for the more general type of mechanism: it would be impossible to explain such changes in order of transfer by means of any of the four possible alternative explanations listed above, because although they might reduce all transfer there is no reason to suppose that they would reduce transfer to different shapes differentially. Although many experiments involving bringing animals up without pattern vision have been performed, there has been no attempt made to investigate how such animals transfer to new shapes.

(2) If a very general mechanism is at work, it should be possible to alter drastically the discriminability of shapes and the ways in which transfer will occur by giving animals unusual perceptual environments.

Gibson and Walk (1956, *ref. 17*) performed an experiment of this kind in which they showed that rats which has been kept in cages in which black circles and triangles were exposed were subsequently better able to discriminate these shapes than a control group brought up in cages without these forms being exposed. Unfortunately, the experimental animals may have learned to orient themselves in their cages by means of the exposed forms so that in the discrimination learning situation they may have transferred a habit of attending rather than a method of classifying. This could be controlled against to some extent by keeping the shapes exposed in constant movement round the cage, and also by finding out whether the superiority of the experimental group on shape discrimination was confined just to the pair of shapes to which they had been exposed by testing both groups on other shape discriminations.

It might be worth trying a different method of producing perceptual learning. If equivalence relationships are determined by the time sequence of shapes on the retina then it should be possible to build into animals very unusual equivalences. Thus animals could be brought up in a cage with a shape in view which was constantly distorting to another shape: for example if animals were allowed to view a triangle distorting into a circle during the early part of their life, it would be predicted that the triangle and circle would thereafter be equivalent figures for them. Discrimination between these figures would presumably be more difficult for such animals on a general mechanism and much more transfer from one to another should be exhibited than in normally reared animals.

(3) A third possible way of discovering how far specific mechanisms are involved in shape discrimination is by testing for interocular transfer under conditions where the animal has never experienced the same shape simultaneously on the two eyes. In a mammal because there are both crossed and uncrossed fibres in the optic pathways corresponding parts of both eyes project to the same hemisphere: because of this, results of experiments using this technique are inconclusive since the interocular transfer found may be due to the same excitation being set up in the primary projection area irrespective of which eye is stimulated. For what it is worth, Chow and Nissen (1955, *ref. 6*) and Riesen and Mellinger (1956, *ref. 48*) have found almost complete interocular transfer in animals which had had alternating pattern vision on both eyes, but had never had pattern vision with both eyes simultaneously. It would be possible to overcome the difficulty in interpreting these experiments by cutting the optic chiasma before any pattern vision was given: this would mean that the left eye projected only to the left hemisphere and the right eye only to the right hemisphere. If under these conditions interocular transfer were found with previous experience limited to pattern vision alternating on the two eyes, then it would be established that further connections from area 17 were highly specific and that classification of shapes was not effected by growth processes in initially random connections.

Myers (1955, *ref. 36*) in fact has found that interocular transfer occurs after severing the optic chiasma, but his animals had had normal pattern vision and hence the chance to establish equivalence between the same pattern projected to the two hemispheres since during the animals' previous binocular experience the same patterns had been being transmitted simultaneously to the two hemispheres.

Before leaving general systems it is perhaps worth making two further points. Firstly, it is sometimes supposed that because classifying systems seem to reflect the probability of environmental events, therefore they must have come into existence as a result of learning in the individual. For example a variety of animals - monkeys (Harlow 1945, *ref. 19*), rats (Lashley 1938, *ref. 27*), octopuses (Sutherland 1959, *ref. 59*) - have been found to classify right-left mirror images together more readily than up-down mirror images. This might arise because in an animal's normal environment if the animal goes behind a given shape it will receive a right-left mirror image of the same shape, whereas it would have to stand on its head to receive an up-down mirror image of a shape: clearly the former happens more frequently so we might expect animals to classify right-left mirror images together more frequently than up-down mirror images. This cannot be used as an argument for a general type of system in which equivalences develop as a result of learning on the part of the individual organism because we would expect a specific system to exhibit the same features since it will have come into being as a result of an evolutionary process. A specific system which treated right-left mirror images as the same would clearly be useful to an animal since they often are given by the same object whereas a system which treated up-down mirror images as the same would be less useful to an animal because in general they will emanate from different objects. Thus it might be expected that specific mechanisms built in as a result of evolution would be adaptive in much the same way as the actual classifications that come to be made in a general mechanism as a result of learning.

Secondly it is sometimes supposed that it might be possible to work out the number of discriminations an animal can make and to correlate this with the total number of possible connections in the brain or in some part of the brain. In this way it might be possible to work out what were the neural elements involved in a general system, and to make a correlation between some form of neural connectivity and the number of things an animal can discriminate. Unfortunately it is completely impossible to set any upper limit to the number of discriminations an animal can make by means of behavioural experiments. The literature of the subject is strewn with assertions that a particular animal could not perform this or that discrimination when all that was justified by the data was that the animal could not perform a particular discrimination within the limits of the particular experimental situation used. Thus in 1930 Munn *(ref. 35)* wrote

"The inability to learn the discriminations was due not to the characteristics of the apparatuses per se, but to a deficiency in the rat's ability to discriminate visual detail": Lashley was shortly to demonstrate that the discrimination between the shapes Munn used with his rats (that between a cross and a square) was in fact one of the easiest pattern discriminations for the rat in the jumping stand situation (Lashley 1938, *ref.* 27). Yet there is no guarantee that the jumping stand is itself the optimal learning situation for shape discrimination in the rat, and it would be as misguided to claim that the limits of the rat's capacity for shape discrimination found with the jumping stand represented the upper limit of the rat's capacity as it was to claim that the rat has no capacity for detail vision on the basis of experiments with a modified Yerkes apparatus as used by Munn. Apart from variations in the experimental situation used to train animals in visual shape discrimination, a second factor may contribute largely to what limits are found for their discriminatory capabilities. This factor is the extent to which animals are pretrained on shapes exhibiting gross differences along the same dimensions as the shapes they will ultimately be required to discriminate. Thus Lawrence (1952, *ref.* 30) found a group of rats trained to discriminate black and white cards and then transferred gradually to more and more similar shades of grey reached a much better criterion of learning on the two closest shades of grey used than a group given the same total number of trials on the two closest shades of grey from the outset of training. Saldhana and Bitterman (1951, *ref.* 49) found that whether or not rats learned a given series of discriminations depended on the order in which they were presented. Thus the approach of arriving at the number of switches in the brain by analysing the total number of discriminations an animal can make is one which it is not possible to follow, though it is an approach which might suggest itself to the engineer interested in applying information theory to organisms.

EVIDENCE FOR SPECIFIC ANALYSING MECHANISMS

Economy of Specific Analysing Systems

Some of the evidence which has already been presented suggests that specific analysing systems must be at work at least in the case of certain submammalian organisms which exhibit innate responses to some classes of stimuli. Although the crucial experiments have not been performed on mammals there is evidence from other directions that some specific classifying mechanisms must be at work even here. Both behavioural and physiological evidence indicate that part of what is involved in discriminating shape is the switching in of an appropriate analysing mechanism. In what follows, I shall argue that within any sensory modality, there are probably different

specific ways of processing incoming information: these different methods correspond to different ways of classifying incoming information. Probably part of what an animal faced with a discrimination task learns is to switch in the correct analysing mechanism: a second part is to attach a response according to which output the mechanism switched in is giving. This general idea has some plausibility on the grounds of economy. Presumably many of the operations to be conducted on incoming data will be the same from one sense to another and also for different ways of classifying information from any one sense, though here the sequence of operations may differ from one method of classifying to another. If this is the case it would presumably be most economical to use the same actual analysing mechanisms in a variety of different classifications rather than to have a separate analysing mechanism for each classification (the latter is implicit in the general type of theory put forward by Uttley). Thus in the use of computers it is most economical to have one computer capable of carrying out operations on different kinds of data and of varying the sequence of the operations it carries out according to the way in which it is programmed, rather than to have different computers for every variety of information which is to be fed in and for every variation in the sequence of operations to be carried out. To summarise this crudely, in the model envisaged the brain is being viewed as containing at least three different boxes: (1) A number of different analysing mechanisms. (2) A control centre which determines which of these mechanisms shall be switched in on any given occasion and in what sequence they shall be switched in. (3) A further box which is responsible for selecting the response to be attached to the output from the analysing mechanisms. This is obviously a gross oversimplification and in practice the boxes may turn out to be not so very discrete, but it is worth seeing how far available evidence supports this conception.

Evidence for Specific Analysing Mechanisms

One consequence of this crude model is that the nervous system would not be able to process information in two different ways at once, since there would then be competition for the common analysing mechanisms. There is plenty of evidence to suggest that this is in fact the case. Thus Mowbray (1953 *refs. 37 & 38*) found that when the eye and ear are presented with complex stimuli at the same time, the information presented to one or the other is made effective in the response but not the information presented to both: Mowbray points out that this finding is against the type of theory put forward by Hebb. Broadbent (1954, *ref. 2*) shows that digits simultaneously presented to the two ears are not accepted by the analysing mechanisms in their temporal order of presentation but all digits presented to one ear are accepted and then all presented to the other ear; he has extended this finding to digits presented simultaneously to ear and eye (Broadbent 1956, *ref. 3*). To explain the finding that information presented

simultaneously on two channels can be accepted although it is accepted successively not simultaneously he postulates a short term memory store in which information can be stored until the central analysing mechanisms are ready to receive it. The idea that the same central analysing mechanisms may be used in processing information from different modalities gives a rationale for the finding that it is not possible to accept information from two channels at once. Davis (1956/7 *refs. 8 & 9*) has shown that where two stimuli requiring different but peripherally compatible responses are presented with intervals of less than about 200 milli-seconds, the response to the second one is delayed: in an ingenious series of experiments involving different modalities, he has shown that the amount of delay is approximately the same as the amount of overlap between the time the first stimulus occupies central pathways and the time the second stimulus would have occupied central pathways if it could have been accepted immediately: this suggests that the second stimulus cannot get access to central analysing mechanisms until the first one is cleared.

A second line of evidence suggesting the same general conception of the working of the central nervous system is that provided by recent studies on the recticular formation and on the peripheral blocking of input on sensory pathways. Since there have been a number of recent reviews of this evidence (e.g. Lindsley 1957, *ref. 31*), it is unnecessary to go into it in detail here. Hernandes-Peon, Scherrer, and Jouvet (1956, *ref. 23*) found that a click given to a cat's ear evokes a markedly reduced potential at the cochlear nucleus if the cat is simultaneously shown a mouse or given a whiff of fish. This suggests that stimuli unimportant for the animal can be blocked at a peripheral level if more important stimuli are being received: the blocking is itself under central control possibly mediated by the reticular formation (Hernandos-Peon and Scherrer, 1955, *ref. 22*). Once again the rationale behind this can only be that central analysing mechanisms can only be set in one way at a time, and if different stimuli were given access to them simultaneously their efficiency in dealing with any one stimulus would be impaired. In addition to giving support for the existence of specific analysing mechanisms used for a variety of stimuli, such experiments suggest in themselves a considerable degree of specificity in the innate organisation of the central nervous system and thus constitute evidence against the very general systems proposed to carry out stimulus analysis. It should further be noted that the existence of analysing mechanisms which although specific could be used for a variety of purposes is in line with the findings on mass action, i.e. the failure of lesions outside the primary sensory and motor areas of the cortex to produce loss of some specific functions only. In addition the idea that the same analysing mechanisms might be used in processing information from different sensory modalities may account for some examples of synaesthesia.

Learning to Switch in Analysing Mechanisms

The degree to which the same analysing mechanisms actually function when incoming information is being processed in different ways is of course extremely speculative, and is almost impossible to test at the moment: it would only become possible to test the idea when we had begun to work out in detail for the different senses what the specific analysing mechanisms at work were. The main point of importance for our present purpose is that it is likely that there are specific analysing mechanisms at work irrespective of the degree to which these are used in common when stimuli are being categorised in different ways. There is excellent evidence to support this more general point drawn from another realm of experimentation. This evidence indicates that in learning to make a discriminatory response animals both learn how to analyse the stimuli and also learn to attach a given response to differential outputs from the analysing mechanism once found. This is one explanation for results which support non-continuity theory as opposed to continuity theory of discrimination learning. Non-continuity theorists as represented by Lashley (1942, *ref. 28*) and Krechevsky (1932, *ref. 25*) maintained that animals only learn to attach a response to cues to which they are attending, continuity theorists (e.g. Spence 1945, *ref. 54*) that they learn to attach a response to any differential cues impinging upon the organism.

Early attempts to test between these two possibilities were not very successful since the experiments yielded conflicting results (Krechevsky 1932 *ref. 26*), McCullogh & Pratt 1934, *ref. 32*, Spence 1945, *ref. 54*, Ehrenfreund 1948, *ref. 13*, etc.): these experiments sought to discover whether animals learned anything about the relevant cues during the first few trials of training in a discrimination during which they tend to react more in terms of spatial position than in terms of the differential shapes. Since the results of these experiments are ambiguous they will not be further discussed here, though it should be noticed that the finding that some learning does occur in early trials is only against an extreme non-continuity theory position, because even in a sequence of trials where animals are attending on most trials to spatial position, they might on some trials be attending to shape (or on the model here proposed, be switching in the appropriate analysing mechanism) and therefore some effect of early training on later learning would be demonstrable.

Lashley (1942, *ref. 28*) tried to solve the problem by giving animals a set to solve discrimination problems in terms of one sort of categorisation (size) and then demonstrating that they learned little or nothing about other aspects of the stimuli to be discriminated provided they could continue to solve the problem in terms of the original way of categorising. Unfortunately, Lashley used an insensitive test of the amount of learning of other aspectsoof the stimuli - he used transfer tests rather than relearning, and the results of his own experiment can be disputed.

Lawrence (1950, *ref. 29*) in an extremely well controlled experiment tried a different approach to the problem: he gave animals a set to respond in terms of one cue, subsequently another cue was made relevant to the discrimination. When tested with the cues in isolation animals performed better when the original cue was relevant than when the additional cue was relevant. More important however, was the finding that when the original positive stimulus was made negative and the original negative was made positive animals learned to reverse their responses quicker than when they had to learn to reverse their responses to the additional cue. Unless some other learning had occurred than learning to attach a given reponse to all cues present it is impossible to explain this finding. It can, however, readily be explained if the group reversed on the preferred cue had learned to switch in the appropriate analysing mechanism and merely had to learn to attach responses differently to the outputs from that mechanism, whereas the group reversed on the less preferred cue would be more likely to switch in the wrong analysing mechanism and so take longer to switch in the less preferred one and also learn to reverse responses to that one.

A number of experiments have recently confirmed this type of finding. Thus Reid (1953, *ref. 42*), Pubols (1956, *ref. 40*), and Capaldi and Stevenson (1957, *ref. 5*) have all shown that the more training is given on a given discrimination the more readily rats learn to reverse their responses on the same discrimination. This finding is unintelligible if we assume animals are merely learning to attach a response to a given output from stimulus analysing mechanisms since we are faced with the paradox that the better the response is attached the easier it is to reverse it. It becomes intelligible, however, if we suppose animals have learned to switch in a given analysing mechanism: animals which have learned this most thoroughly will presumably continue to switch this mechanism in when their responses begin to give the wrong results and so will have a chance to learn to reverse their responses to the outputs of that mechanism. Animals which have not learned so thoroughly may switch in other analysing mechanisms and hence take longer to relearn since they will meanwhile not learn anything about the relationship of correct response to the outputs from the original analysing mechanism.

Bruner, Matter and Papanek (1955, *ref. 4*) demonstrated that the obverse of this is true: rats were given different amounts of training on one cue. They were then given 20 trials training with a second cue added to the first. They then had to learn to respond in terms of the second cue only (i.e. the first was removed). The more training they had had on the initial cue, the less they learned about the new cue during the training with both cues present. This again suggests that the more thoroughly they had learned to switch in the analysing mechanism which would detect the original cue, the less they switched in different analysing mechanisms when

both cues were present and therefore the less they learned during this period about the second cue.

It has been known for some time that if an animal is repeatedly conditioned and extinguished on the same habit, the length of time necessary for successive conditioning and extinction becomes less and less. More recently it has been shown by Harlow (1944, *ref. 18*) that if monkeys are trained on a discrimination and then the correct response is reversed successively they will eventually come to reverse their response in one trial: Pubols (1957 *ref. 41*) has demonstrated that rats also are capable of learning to do this. Again this suggests that animals do not learn simply to attach a response to a stimulus, though in order to explain these findings it is necessary to suppose an extra complication in the crude model here put forward. Not only must there be a control mechanism for switching in analysing mechanisms which mediates the learning of which analysing mechanism to switch in, but there must be a further control mechanism which can switch the relationship between the outputs from the stimulus analysing mechanism and the response.

THE PROBLEM OF SPECIFIC ANALYSING MECHANISMS

I have now reviewed some of the evidence for supposing that there are specific analysing mechanisms at work and that on different occasions different analysing mechanisms may be switched in. This raises a serious obstacle to working out in detail what the analysing mechanisms used for any stimulus modality are. The technique which I have used with octopuses (Sutherland 1957/8, *refs. 56, 57 & 58*) is to try to discover by experiments which shapes they can most readily discriminate and also what are the properties of the shapes they are analysing when they do discriminate between them by running transfer tests with new shapes in which the properties of the original shape are systematically altered. For example (Sutherland 1958, *ref. 57*) if an octopus is trained on a circle and a square of equal area, discrimination is unimpaired by altering the size of the original figures and this establishes that octopuses were not discriminating originally in terms of absolute length of outline or absolute breadth or height; on the other hand they will not transfer from the square to a diamond (i.e. a square rotated through 45°) and this establishes that they were not originally discriminating the properties of having straight lines and corners as against the properties of not having straight lines and not having corners. On the basis of this sort of work it is possible to suggest hypothetical analysing mechanisms which would account for one's results, and to draw predictions from such analysing mechanisms about the discriminability of further pairs of shapes. Unfortunately, though, in the case where animals succeed in discriminating shapes which on the original analysing mechanism

should not be discriminable one does not know whether this is because one's original guess at the analysing mechanism at work was wrong or whether it is because a second analysing mechanism is at work which is switched in where the original one does not give differential outputs for different pairs of shapes (for an example of this v. Sutherland 1959, *ref. 60*). This means that where one predicts that an animal cannot distinguish shapes the finding that it cannot distinguish them is some confirmation for the existence of the original analysing mechanism, while the finding that it can discriminate them is not complete disconfirmation. On the other hand where one predicts that a pair of shapes should be readily discriminable the finding that they are is only partial confirmation of the theory, whereas the finding that they are not would be a complete refutation of the theory. However, possibly because of the existence of different possible analysing mechanisms it is difficult to discover pairs of shapes which are not discriminable. Although the existence of independent analysing mechanisms makes investigation of the specific analysing mechanisms at work difficult, it does not make it impossible: one approach to this problem which has not been tried would be to preset the analysing mechanisms in a given way by training on shapes whose discrimination involves this or that type of analysing mechanism, and then to test for whether when animals are trained on further shapes the ways in which they transfer to new shapes are altered by the way in which the analysing mechanisms have been preset.

A further problem which confronts investigators who are trying to make hypotheses about the specific analysing mechanisms at work is to decide in what ways the nervous system is most likely to code information. Thus coding in terms of intensity of firing (i.e. rate of firing) and position of firing certainly occurs in the peripheral nervous system: it is difficult to know how far the nervous system may operate at more central levels on a coding in terms of time. Since peripheral response mechanisms would appear to work only in terms of the position of impulses and their intensity of firing decoding into these methods of carrying information would be necessary before the effector system was reached. A concrete example may help both to make this clear, and also to illustrate how far our ignorance of specific mechanisms goes.

Suppose we tell a human being that we are going to present two lights to his eye, one of which will be brighter than the other and he is to respond by pressing one of two keys to his right and left: if the bright light is to the left of the dim light he is to press the left hand key, if it is to the right of the dim light he is to press the right hand key. If we now control the subject's fixation and flash the lights on different parts of the retina, one would expect him to press the correct key for any retinal position occupied by the otwo lights provided their separation is greater than the minimum separable for the part of the retina on which they are projected. If we take the minimum separable to be 3 minutes or better out to a peripheral angle of 60^{o} (Polyak, 1941, *ref. 39*) this means that each

light can occupy over 1,000,000 discriminably different positions on the eye and the two lights could occupy over 10^{12} different positions. We give the left hand response for half of these possible input states, and the right hand response for the other half. Now, on Uttley's theory of how the equivalence for each group of $10^{12}/2$ input states has arisen, there would be one unit to represent each of these $10^{12}/2$ states and all these units would be connected to one further unit which is common to all of the $10^{12}/2$. We would in fact require more units than there are neurons in the central nervous system to cope with this one very simple discrimination. This in itself suggests that some specific mechanism of a more economical kind is operating, and it is the purpose of the example to throw light on the difficulties which arise in specifying the more specific mechanisms.

Despite the simplicity of the problem, there are as far as I know no discussions in print of what sort of analysing mechanism could mediate this generalisation or even this type of generalisation. The problem is to get rid of the information about the actual spatial positions occupied by the two lights and preserve only the information about their relationship to one another along the horizontal axis. One way of doing this would be to convert spatial positions into a time series: for example points on the cortex representing retinal points from left to right might fire successively into the same channel: the two lights will now be represented by a small excitation and a large excitation and a detector mechanism could sort out whether the large light was on the right or on the left by whether a large excitation was succeeded or preceded by a small one. Both Deutsch (1955, *ref. 10*) and Dodwell (1957, *ref. 11*) have proposed mechanisms for shape recognition which get rid of spatial position occupied on the retina by coding position in terms of time. The trouble with this sort of system is that we have to assume some sort of pacing mechanism which must operate in a very regular way, and also that the central nervous system is capable of detecting very small differences in time. There is evidence that the nervous system can decode very small time differences produced by external stimuli (e.g. time differences of 30 microseconds can be detected in auditory localisation, Wallach, Newman, Rosenzweig (1949, *ref. 65*), but here the time differences are produced by an external stimulus and this does not show that the nervous system can itself recode information in terms of time with such accuracy.

Another possible method of recoding the information would be to code position on the retina in terms of intensity: successive positions say to the right of the fovea would be represented by greater and greater intensities. We now need a second mechanism for associating the brightness of the light with the intensity of firing associated with it in the system representing its spatial position in intensity terms, and to respond merely to the spatial position of the light we have to detect whether the intensity associated with the bright one in the system representing horizontal retinal

position is greater or less than the intensity associated with the dim one. This sort of recoding at least for visual discrimination is possibly more plausible than recoding in terms of time, since it does not involve any sort of regular pacing mechanism. Moreover, it is already known that at one stage the nervous system does recode spatial information on the retina in terms of intensities. Thus one primitive response we make to objects stimulating the retina is that of fixation movements: one pair of muscles (lateral and medial recti) are mainly responsible for the horizontal components of eye movements, while two further pairs (superior and inferior recti and superior and inferior oblique) are mainly responsible for the vertical components of eye-movements. The immediate efferent control of these muscles must be in terms of intensity of firing in efferent nerves. Moreover there is some evidence that there are separate efferent tracts for the vertical and horizontal components of eye muscles: thus it seems likely that the efferent tract for lateral eye movements runs through the pons, whereas that for vertical eye movements runs through the superior colliculus (Duke-Elder 1949, *ref. 12*). If information about the position on the retina of a point to be fixated has to be divided into the two coordinates vertical and horizontal in order to be fed over efferent pathways onto the eye muscles, it might represent some economy if the same system were used for analysing the position of stimulated points on the retina relative to one another. Such a system could be developed to account for shape recognition irrespective of retinal position, and would make many predictions in common with the theory put forward by the present writer (Sutherland 1957, *ref. 55*) for recognition of shape and orientation in the octopus. The approach of discovering how information from a given modality must be coded in order to govern some primitive response in order to gain suggestions for analysing mechanisms which govern more complex responses is one which might be used more than it has been in the past.

There is one further point which may be worth making about specific systems of shape recognition. It seems likely that any such system must operate by comparing relative quantities somewhere in the nervous system rather than by taking into account any absolute quantity at any given stage. This is indicated not only by the fact that in all species tested it has been found that having learned to discriminate between shapes of a given size they will transfer the discrimination readily to shapes of different sizes, but by more general considerations about the operation of the nervous system. Any discrimination task which necessarily involves the storage of information about absolute quantities is in fact very poorly performed in terms of the information which can be extracted on any one presentation of the stimulus. Thus Carner (1953, *ref. 16*) showed that where human beings were asked to make judgments of absolute loudness, judgments were most accurate as measured by the amount of information transmitted per judgment where only five categories were used: increasing

the number of categories led to a decrease in accuracy which was not compensated for in terms of extra information transmitted by means of the additional number of categories used. Similarly Ericksen and Hake (1955, *ref. 14*) found that where subjects were asked to judge the size of squares, increasing the number of categories used above five did not lead to any increase in the amount of information transmitted per judgment: it led to a decrease in the accuracy of judgments which was partially compensated for by the increase in number of categories used. In both studies about 2.1 bits of information per stimulus were transmitted. It is obvious that in human beings and many animals very much more information per stimulus can be transmitted where visual patterns are being classified, particularly if there are no severe limitations on the length of time the stimulus is exposed or the length of time within which a response must be made. This can only mean that the nervous system performs more efficiently in terms of information transmitted where it is analysing relationships between quantities simultaneously present in it, than where it is analysing one absolute quantity. The explanation of this feature of the nervous system may have to do with changes in states of adaptation: it may be impossible to analyse accurately absolute quantities due to changes in states of adaptation of parts of the nervous system, but the changes in states of adaptation might be such that the relation between different quantities of excitation is preserved provided they are transmitted over the same parts of the nervous system. A mechanism for distinguishing shapes which depends upon an analysis of relative quantities is therefore more plausible than one which depends upon an analysis of absolute quantities.

CONCLUSION

It has not been the purpose of this paper to examine in detail any specific theories of how the nervous system classifies shapes. I have tried to show that the very general type of analysing system which engineers have tended to propose for pattern recognition may not correspond to the way in which the nervous system works, and to give reasons for supposing that there are in fact more specific and more economical analysing systems at work. I have also tried to suggest ways of testing between the two alternatives, and ways of working out what the more specific analysing mechanisms are. The engineer could obviously be of enormous help to the physiologist and the psychologist in setting up hypotheses about specific analysing mechanisms, but to do so he would have to start by taking into account the known facts about which shapes are classified together and which are classified apart, and to develop theories about analysing mechanisms in collaboration with experimentalists who could test the specific predictions made from this type of theory.

ACKNOWLEDGMENT

This paper was written while the author was engaged on a research project on stimulus analysing mechanisms financed by the Nuffield Foundation: the author wishes to express his gratitude to them for their financial support.

REFERENCES

1. ASHBY, R. C. Design for a Brain. *Chapman & nall: London.* (1952).
2. BROADBENT, D. E. The role of auditory localisation in attention and memory span. *J. exp. Psychol.*; 1954, **47**, 191.
3. BROADBENT, D. E. Successive responses to simultaneous stimuli. *Quart. J. exp. Psychol.*; 1956, **8**, 145.
4. BRUNER, J. S., MATTER, J. & PAPANEK, M. L. Breadth of learning as a function of drive level and mechanization. *Psychol. Rev.* 1955, **62**, 1.
5. CAPALDI, E. J. & STEVENSON, H. W. Response reversal following different amounts of training. *J. comp. physiol. Psychol.*, 1957, **50**, 195.
6. CHOW, K. L. & NISSEN, H. W. Interocular transfer of learning in visually naive and experienced infant chimpanzees. *J. comp. physiol. Psychol.*; 1955, **48**, 229.
7. COLLIER, R. M. An experimental study of form perception in indirect vision. *J. comp. Psychol.*, 1931, **11**, 281.
8. DAVIS, R. The limits of the "psychological refractory period". *Quart. J. exp. Psychol.*; 1956, **8**, 24.
9. DAVIS, R. The human operator as a single channel information system. *Quart. J. exp. Psychol.*; 1957, **9**, 119.
10. DEUTSCH, J. A. A theory of shape recognition. *Brit. J. Psychol.*; 1955, **46**, 30.
11. DODWELL, P. C. Shape recognition in rats. *Brit. J. Psychol.*; 1957, **43**, 221.
12. DUKE-ELDER, W. S. Textbook of Ophthalmology. Vol. IV. The neurology of vision. Motor and optical anomalies. *London: Henry Kimpton.* (1949).
13. EHRENFREUND, D. An experimental test of the continuity theory of discrimination learning with pattern vision. *J. comp. physiol.; Psychol.*; 1948, **41**, 408.
14. ERICKSEN, C. W. & HAKE, H. W. Absolute judgments as a function of the stimulus range and the number of stimulus and response categories. *J. exp. Psychol.*; 1955, **49**, 323.
15. FANTZ, R. L. Form preferences in newly hatched chicks. *J. comp. physiol. Psychol.*; 1957, **50**, 422.
16. GARNER, W. R. An informational analysis of absolute judgments of loudness. *J. exp. Psychol.*; 1953, **46**, 373.

17. GIBSON, E. J. & WALK, R. D. The effect of prolonged exposure to visually presented patterns on learning to discriminate them. *J. comp. physiol. Psychol.*; 1956, **49**, 239.
18. HARLOW, H. F. Studies in discrimination learning by monkeys: I. *J. gen. Psychol.*; 1944, **30**, 3.
19. HARLOW, H. F. Studies in discrimination learning by monkeys: III. Factors influencing solution of discrimination problems by rhesus monkeys. *J. gen. Psychol.*, 1945, **32**, 213.
20. HEBB, D. O. The Organisation of Behaviour. *New York: Wiley.* (1949).
21. HESS, E. H. Space perception in the chick. *Scientific American*, 1956, **195**, 71.
22. HERNANDOS-PEON, R. & SCHERRER, H. *Federation Proc.*; 1955, **14**, 71.
23. HERNANDOS-PEON, R. & SCHERRER, H. & JOUVET, M. Modification of electric activity in the cochlear nucleus during "attention" in unanaesthetised cats. *Science*, 1956, **123**, 331.
24. KOHLER, I. Uber Aufbau und Wandlugen der Wahrnehmungswelt. *Oesterr. Akad. Wiss. Philos - Histor. Kl. Sitz - Ber.*, 1951, **227**, 1.
25. KRECHEVSKY, I. "Hypotheses" versus "choice" in the pre-solution period in sensory discrimination learning. *Calif. Univ. Publ. Phychol.*; 1932, **6**, 27.
26. KRECHEVSKY, I. The genesis of "hypotheses" in rats. *Calif. Univ. Publ. Psychol.*; 1932, **6**, 45.
27. LASHLEY, K. S. The mechanism of vision: XV. Preliminary studies of the rat's capacity for detail vision. *J. gen. Psychol.*; 1938, **18**, 123.
28. LASHLEY, K. S. An examination of the "continuity theory" as applied to discriminative learning. *J. gen. Psychol.*, 1942, **26**, 241.
29. LAWRENCE, D. H. Acquired distinctiveness of cues. II. Selective association in a constant stimulus situation. *J. exp. Psychol.*; 1950, **40**, 175.
30. LAWRENCE, D. H. The transfer of a discrimination along a continuum. *J. comp. physiol. Psychol.*, 1952, **45**, 511.
31. LINDSLEY, O. R. Psychophysiology and motivation. In Jones (Ed): *Nebraska Symposium on Motivation. V.* 1957, pp. 44-105.
32. McCULLOGH, T. L. & PRATT, J. G. A study of the pre-solution period in weight discrimination by white rats. *J. comp. Psychol.*; 1934, **18**, 271.
33. MILLER, M. Observation of initial visual experience in rats. *J. Phychol.*, 1948, **26**, 223.
34. MILNER, P. M. The cell assembly: mark II. *Psychol. Rev.*, 1957, **64**, 242.
35. MUNN, N. L. Visual pattern discrimination in the white rat. *J. comp. Psychol.*, 1930, **10**, 145.

36. MYERS, R. E. Interocular transfer of pattern discrimination in cats following section of crossed optic fibres. *J. comp. physiol. Psychol.*, 1955, **48**, 470.
37. MOWBRAY, S. H. Simultaneous vision and audition: the comprehension of prose passages with varying levels of difficulty. *J. exp. Phychol.*, 1953, **46**, 365.
38. MOWBRAY, S. H. The perception of short phrases presented simultaneously for visual and auditory reception. *Quart. J. exp. Psychol.*, 1954, **6**, 86.
39. POLYAK, S. L. The Retina. *Chicago: Univ. of Chicago Press.* (1941).
40. PUBOLS, B. H. The facilitation of visual and spatial discrimination reversal by overlearning. *J. comp. physiol. Psychol.*, 1956, **49**, 243.
41. PUBOLS, B. H. Successive discrimination reversal learning in the white rat: a comparison of two procedures. *J. comp. physiol. Psychol.*, 1957, **50**, 319.
42. REID, L. S. The development of noncontinuity behaviour through continuity learning. *J. exp. Psychol.*, 1953, **46**, 107.
43. RHIENGOLD, H. I. & HESS, E. H. The chick's preference for some visual properties of water. *J. comp. physiol. psychol.*, 1957, **50**, 417.
44. RIESEN, A. H. The development of visual perception in man and chimpanzee. *Science*, 1949, **106**, 107.
45. RIESEN, A. H. Arrested vision. *Sci. Amer.*, 1950, **183**, 16.
46. RIESEN, A. H. Post and partum development of behaviour. *Chicago Med. School. Quart.*, 1951, **13**, 17.
47. RIESEN, A. H., KURKE, M. I., & MELLINGER, J. C. Interocular transfer of habits learned monocularly in visually naive and visually experienced cats. *J. comp. physiol. Psychol.*, 1953, **46**, 166.
48. RIESEN, A. H. & MELLINGER, J. C. Interocular transfer of habits in cats after alternating monocular visual experience. *J. comp. physiol. Psychol.*, 1956, **48**, 516.
49. SALDANA, E. L. & BITTERMAN, M. E. Relational learning in the rat. *Amer. J. Psychol.*, 1951, **64**, 37.
50. SHARPLESS, S. & JASPER, H. Habituation of the arousal reaction. *Brain*, 1956, **79**, 655.
51. SPERRY, R. W. Reestablishment of visuomotor coordination by optic nerve regeneration. *Anat. Rec.*, 1942, **84**, 470.
52. SPERRY, R. W. Visiomotor coordination in the newt (Triturus vividesceus) after regeneration of the optic nerve. *J. comp. Neurol.*, 1943, **79**, 33.
53. SPERRY, R. W. Optic nerve regeneration with return of vision in anurans. *J. Neurophysiol.*, 1944, **7**, 57.
54. SPENCE, K. W. An experimental test of the continuity and non-continuity theories of discrimination learning. *J. exp. Psychol.*, 1945, **35**, 253.
55. SUTHERLAND, N. S. Visual discrimination of orientation and shape by the octopus. *Nature*, 1957, **179**, 11.
56. SUTHERLAND, N. S. Visual discrimination of orientation by octopus. *Brit. J. Psychol.*, 1957, **48**, 55.

57. SUTHERLAND, N. S. Visual discrimination of shape by octopus. Circles and squares, and circles and triangles. *Quart. J. exp. Psychol.*, 1958, (in press).
58. SUTHERLAND, N. S. Visual discrimination of the orientation of rectangles by *Octopus vulgaris* Lamarck. *J. comp. physiol. Psychol.*, 1958, (in press).
59. SUTHERLAND, N. S. Visual discrimination of orientation by octopus: mirror images. *Brit. J. Psychol.*, 1959, (in press).
60. SUTHERLAND, N. S. A test of a theory of shape discrimination in octopus. *J. comp. physiol. Psychol.*, 1959, (in press).
61. TAYLOR, W. K. Electrical simulation of some nervous system functional activities. In Cherry, C. (Ed.): Information Theory. Third London Symposium, 314-327. *Butterworths Publications: London.* (1956).
62. TINBERGEN, N. The Study of Instinct. *Oxford: Clarendon Press.* (1951)
63. UTTLEY, A. M. Conditioned probability machines and conditioned reflexes. In Shannon, C. E. & McCarthy, J. Automata Studies, 253-275. *Princeton: Princeton Univ. Press.* (1956).
64. UTTLEY, A. M. Temporal and spatial patterns in a conditioned probability machine. In Shannon, C. E. & McCarthy, J. Automata Studies, 277-285. *Princeton: Princeton Univ. Press.* (1956).
65. WALLACH, H., NEWMAN, E. B. & ROSENZWEIG, M. R. The precedence effect in sound localization. *Amer. J. Psychol.*, 1949, **62**, 315.
66. WELL, M. J. Factors affecting reactions to MYSIS by newly hatched SEPIA. *Behaviour*, 1958, **8**, 96.

DISCUSSION ON THE PAPER BY DR. N. S. SUTHERLAND

DR. CROSSMAN: Dr. Sutherland appears to be concerned that the brain might not make use of specific mechanisms for stimulus analysis, and that generalized ones with learning properties might prove difficult or impossible to study. But up to date no single specific mechanism has been elucidated neurologically, with the possible exceptions of one in the frog's retina (by Barlow) and the octopus tactual system (by Wells). Therefore it does not seem to be at all obvious that specific mechanisms are easier to elucidate than generalised ones such as have been proposed, for example, by Hebb and Uttley. The latter might even be easier, if it proves possible to use a statistical approach on large assemblies of identical units, as is done in statistical mechanics with notable success.

DR. JOHN BROWN: In his very interesting paper, Dr. Sutherland argued that, if a general analyzing mechanism underlies classification in the nervous system, then the actual circuits for particular classification will vary from individual to individual and that, if this is so, all we can do is to try to establish correlations between earlier stimulus sequences and subsequent behaviour. This seems too pessimistic. The same kinds of circuits will probably be established in different brains even if different cells are involved in each case. And from hypotheses about circuits we may well be able to deduce functional properties such as the speed with which classifications of a given complexity will be performed. It is important to be clear on the difference between a general analyzing mechanism and a specific one. Even the general mechanism must have its limitations since no mechanism can have infinite capacity and a worthwhile job is discovering what these are. What distinguishes a specific analyzing mechanism is that it only handles information of certain sorts or at least handles certain sorts more quickly and efficiently. It is of interest that the nervous system seems to be poor at handling order information, whether spatial or temporal, as I have suggested in my paper (paper 7).

PROF. J. Z. YOUNG, CHAIRMAN: It is quite true that the specific characteristics of wiring will not be what we are looking for, but there are other relevant things which unfortunately, at present, we anatomists have not told you. For example, there are very few data indeed about the variance of characteristics of neurones. Even such simple things as

whether the gross cell size varies during lifetime, or in tissue which has learned or not learned would be interesting to know. There are all sorts of things to be found out about numbers of boutons and other synaptic points, numbers of branches and connections and the variances of all these. Discussion such as the present stimulate us to look for things which we have not thought about before.

DR. N. S. SUTHERLAND: Dr. Crossman's main point is that specific systems have not been discovered, but you can hardly use this as an argument against their existence, unless you are also going to use the argument that general systems which work have not been discovered either, and that is an argument against general systems. In fact, specific systems have been proposed and have led to a lot of experiments and, heuristically at least, they have been extremely useful. I proposed a specific system about shape recognition in the octopus, which has led to 20 or 30 experiments, and so far I have not succeeded in disproving it, although it does make fairly precise predictions. I do not say that it is right, but at least it has uncovered quite a lot of evidence about the animal. On the point about treating systems statistically, I do not want to suggest that because specific systems are at work they do not function in a probabilistic way. This is not an alternative in any way at all. On Dr. Brown's points, I am not sure that I have grasped them very well. He said does it matter if specific cells are involved in different members of the same species. I think it certainly does, and if they are involved that is something which we ought to know about. His second point was that even on general systems you may get the same kind of circuits in different brains. This may be true. I think it would be nicer to find specific systems, but I do not want to use that as an argument for the fact that they are there - only as a reason for looking to see. My arguments in favour of specific systems are based on a whole lot of evidence, which I think it is very difficult to explain given general systems. In particular, I think it is almost impossible to explain by a general system how any animal can have a built-in response, which must be attached to some output from a classificatory mechanism. If you postulate a general system and all shape recognition comes from learning, I just do not see how you can specify the connection between a way of categorising patterns and a way of responding. It does not seem possible to specify this connection genetically, unless the classificatory system is also specified genetically.

On the question of the limits of performance, which Dr. Brown raised, I tried to show in my paper that I did not think that was a very useful way of talking. I think that any experimentalist will tell you that it just is not possible to specify the limits of performance, at the moment. It depends so much on the experimental situation: if you get a better

experimental situation you get better performances. I quoted in my paper something that Munn said in 1930, that he found rats were unable to discriminate between squares and crosses in a particular apparatus, and he came to the conclusion that the inability of the rat to discriminate between squares and crosses was a defect of its visual mechanism, and that it was not a defect of the apparatus he used. Two years later Lashley found that this was one of the easiest pattern discriminations for rats to perform using a different apparatus, so in view of all this it is almost impossible to set limits to what animals can do

DR. A. J. ANGYAN: I would like to ask Dr. Sutherland whether he finds it possible to explain specific analytic mechanisms on the basis that there are some innate mechanisms in animals. For instance, observations on newly hatched birds show the action of three specific stimuli: mechanical stimulation of the nest; pitch of voice which is specific for their parents; and air vibrations caused by the wings of the parents. I would suggest that if any specific mechanisms are involved in discrimination, they can develop on the basis of these innate stimulation patterns. The conditional probability system of learned discrimination might, therefore, be built on this basis.

Another question concerns the fact observed in most young animals, particularly rodents, that highly specific pitch sensitivities evoke unconditioned motor responses. If these are the starting points for stimulus analysing mechanisms in the sense that the so-called specific mechanisms may only develop on the basis of innate automatised behaviour, we have to deal with automatic analogical sequences conditioned by a system which behaves like a 'conditional probability computer'. We might assume that both the above mechanisms must be involved in any acquired behaviour. I would be very glad to hear his opinion about this hypothesis.

Facts of psychology and physiology seem to indicate that one must deal with a combination of specific and non-specific mechanisms and model them analogically. If we accept the example shown by the preceding speaker, a correct analysis of the stimulation may only be obtained if both mechanisms act together. For instance, colour discrimination as tested by conditioning is a very difficult, if not insoluble, task for animals. If these cells are to be supposed to fire in a random way the stimulus resulting at the output would only be a probabilistic one. In consequence, discrimination of stimulus intensities can be more refined than discrimination of quality. To find common elements as they are brought together in objects reflecting coloured light or simply colours, analogical feedback mechanisms must be supposed to act upon them. Facts show by experiments of Sperry, and especially by more recent work of Dr. Sznetogathai and Szekely in Hungary, on the example of the opto-kinetic reflexes indicate that if we suppose the cells to be coupled in a sequence in a negatively fed back manner, then

we have to deal with given fixed relations as a basis for stimulus analysis. Something of this kind must occur in the analytic-synthetic discrimination of the stimuli, about which we have common concepts.

MR. B. L. M. CHAPMAN:[ø] I would like to say that I am in agreement with Dr. Sutherland that we should try to spend our time studying specific systems at this stage of our present knowledge of the nervous system. I am glad that he raised this particular point about discrimination between lines of different lengths for it is one which, I think, may be very important to a proper understanding of visual processes. However, I do not think that this is a satisfactory explanation of discrimination of line lengths in human beings, although I am not prepared to argue about the octopus.

Consider these two lengths which Dr. Sutherland has marked off representing images on the human retina. He has said that there is a mapping from a line of cells on the retina on to a block of cells and that there is a sequential mapping of the two images from the retina on to that same block of cells. If that were so, then one would expect that, this being essentially a digital system, human beings would have a very accurate discrimination not only of difference in length but of the actual ratio of this difference to the original length of one line. We ought then to be able to look at two lines and say; "Yes, this line is 12.7% longer than the other". I think a more likely explanation of this problem is that line length is judged by means of eye movements.

If two lines of different length are placed parallel and especially if one end of each is on an imaginary line perpendicular to them, it is possible, by means of eye movement in one plane only, to move the eye at a fairly constant speed and judge time intervals. By moving the eye along each line in turn one is able to compare the two time intervals taken to scan them. The farther apart that these two lines are, the longer the interval between successive scannings, and during that interval our memory of the time taken to scan the preceeding line will decay as we move the image across the visual field. Further, if two lines are placed in random orientations in the visual field then our two scanning processes

---

[ø] Mr. Chapman's comments arise out of a point made by Dr. Sutherland in his opening remarks but not made in his paper. Dr. Sutherland drew attention to the following phenomenon: if two parallel lines of unequal length are presented to the human eye, the difference in length is much more readily discriminated if the lines either start or finish together (i.e. if the line joining together one of the two pairs of ends of the original lines is at right angles to the original lines) than if the lines neither start nor finish together. Dr. Sutherland pointed out that this fact is so obvious that it has hitherto escaped experimental investigation, and that by considering what sort of machine for recognising line lengths would have this property, light might be thrown on how the stimulus analysing mechanisms for visual patterns operate in man.

are not made with the same set of muscular movements which will make their comparison more difficult.

DR. W. K. TAYLOR: I will go some way with Dr. Sutherland in his postulation of specific mechanisms because obviously there is sufficient genetic information to determine the organization of the retina and the system responsible for vertical tracking. I would go a little further and suggest that there is a specific mechanism for giving size generalization. In the case of humans, however, when it comes to learning the names of objects, I think it extremely improbable that there is a specific mechanism that gives a French baby an initial preference for learning the French names for objects rather than their English names. In this case we are obviously looking for mechanisms that are initially non-specific.

DR. D. M. MACKAY: There is at least one 'discussion in print' of the problem mentioned by Dr. Sutherland on p.595, in the Macy Foundation conference proceedings of 1951 *(ref. 1)*. I should be interested to know what he thinks of the suggestion I made there. Briefly, the idea was that filtering the *input* to extract invariants entailed needless trouble, because what is wanted is to organise an invariant *response* to the input. The alternative is to make the abstractive mechanism itself part of the response system, so designed that it can organize an internal matching response invariant with respect to position. To show what this means, I have used *(ref. 1)* the example of a self-guiding trolley, set to drive around a triangle: wherever the triangle lies in the field, the trolley finds that it has to present itself with the same succession of commands, in order to run round it. My suggestion was that an analogous but internal matching mechanism could account for position-invariance in visual perception. This general principle of making stimulus classification a matching task for part of the response mechanism seems to me to be rather more promising in terms of the demand on neural complexity, than making it entirely a matter of filtering the input. It does not, of course, solve all the problems. We do not understand the details of the specific neural mechanism which enables a man, for example, to *draw* the same figure in different positions. I think, however, that physiologists are rather happier with that kind of problem than they have been with the problem of designing a specific neural filter to give an invariant signal from a figure anywhere in the visual field.

REFERENCE

1. MACKAY, D. M. In search of Basic Symbols. Proc. 8th Conference on Cybernetics. *Jos. Macey Jr. Foundation, New York.* (1951).

DR. A. M. UTTLEY: We have had a number of mentions of specific mechanisms at birth from Dr. Angyan. One thing I can add to that is the small bird crouching. I believe I am right in saying that a card with four black dots on it will cause crouching, but if one is removed it will not. It seems to me quite inescapable that there must be a dendritic system somewhere connected to four receptors, or small groups of receptors; and I would like the anatomists to start looking for one *now*. Surely the engineers have convinced them and given them enough faith to start looking somewhere for some of these specific mechanisms?

DR. N. S. SUTHERLAND: I think I am certainly in agreement with Dr. Uttley over this, and I think I am in agreement with Dr. Angyan, so I do not think I need say anything about their points. Dr. Taylor's is rather important, because it involves a misunderstanding of what I was saying. Of course, one has to learn to make responses to stimuli as classified. I was not suggesting for a moment that Frenchmen came into the world speaking French. In this paper I have not really been concerned at all with how the responses are attached to the outputs from stimulus classifying mechanisms. I am just concerned with trying to find out what are the stimulus analysing mechanisms. The question of how - given that there are certain stimulus analysing mechanisms which have got certain outputs - an animal attaches a response to the outputs of these is a different question. I would like to take up Chapman's points on the lines. I think his first point exposes a confusion about this sort of thing. It is not true that, just because this a digital system is involved, animals must be able to respond extremely accurately. Whether they respond accurately or not depends on what further mechanism you attach. You cannot make predictions very readily from just the first stage of an analysing mechanism. The information preserved here might very easily be lost later on, depending on what sort of mechanism you put on to analyse these differences in the lines. The second point about the lines is that he suggested eye movements were concerned. They may be concerned normally, but they certainly are not always, because if you try exposing lines with a tachistoscopic flash, you will find that you can still discriminate the lengths of the lines. I am not sure how much worse you can do this than if you allow eye movements. Dr. MacKay suggested that the stimulus analysing mechanisms might be part of the response system, and this is something which I agree with and which I have in fact suggested, myself, in my paper - that the stimulus analysing mechanisms, the ones that exist, may very well be based on how you code information in order to get some primitive response, like the response of making eye movements. It would obviously be an economy, if you have got to code the information in a certain way in order to govern this primitive response, to use this way of coding it as part of an analysing mechanism.

He suggested this example of a trolley, say, running round the triangle, and this is rather like the Hebb system of shape recognition. In Hebb's system the eye movements, it seems to me, actually had to be made before he could get shape recognition, and of course you can very readily show again that triangles can be recognised without any eye movement at all, but this does not mean to say that the way the information is processed, in order to make eye movements, may not give you a clue as to what the stimulus analysing mechanism is.

DR. D. M. MACKAY: That was not the idea at all. It was an internal process.

DR. N. S. SUTHERLAND: I realise that yours was. It is Hebb's which is external. I agree with Dr. Mackay that the way in which information is coded for a response may throw light on how it is coded in the classifying mechanism. However, he has discussed the problem only in very general terms: I am suggesting that we ought to look at how information is actually coded in built-in response mechanisms, for example those controlling eye-movements, in order to give us a clue to the perceptual analysing mechanisms. Dr. Mackay did not discuss the specific example I raised of a left or right response given to the brighter of two lights on the retina in his article in the Macey Foundation conference on Cybernetics. He thinks this sort of problem may be solved by an internal matching response which is invariant with respect to position, but this is a far cry from specifying in detail a mechanism which will actually achieve this. It is this sort of detailed specific mechanism for which I think pyschologists ought to be looking.

MR. W. LAWRENCE: I wonder if I might put in a point there? It seems to me that this is very relevant to the question of apprehension of speech, where the response mechanism of being able to speak the same words, oneself, has a very great bearing on one's ability to recognise the stimulus.

SESSION 4A

PAPER 3

AGATHE TYCHE
OF NERVOUS NETS - THE LUCKY RECKONERS

by

DR. W. S. McCULLOCH

Dr. Warren S. McCulloch, born 1898 Orange, New Jersey, U.S.A., was educated at Yale University (B.A. 1921) and Columbia University (M.A. 1923, M.D. 1927). His psychiatric training was at Rockland State Hospital (N.Y.), 1932-4. Until 1941 he held several fellowships at Yale University, Laboratory of Neurophysiology, on activity of the central nervous system, becoming Assistant Professor 1940-1. From 1941 to 1952 he was Professor of Psychiatry and Physiology and Neurophysiologist at the University of Illinois. Since 1952 he has been staff member of the Research Laboratory of Electronics at Massachusetts Institute of Technology.

He is the author of numerous articles on functional organization of the brain, and on facilitation, extinction and functional organisation of the cerebral cortex.

AGATHE TYCHE
OF NERVOUS NETS - THE LUCKY RECKONERS

by

DR. W. S. McCULLOCH

SUMMARY

VENN diagrams, with a jot in every space for all cases in which given logical functions are true, picture their truth tables. These symbols serve as arguments in similar expressions that use similar symbols for functions of functions. When jots appear fortuitously with given probabilities or frequencies, the Venn diagram can be written with 1's for fixed jots, 0's for fixed absence, and p's for fortuitous jots. Any function is realizable by many synaptic diagrams of formal neurons of specified threshold, and the fortuitous jots of their symbols can be made to signify a perturbation of threshold in an appropriate synaptic diagram.

Nets of these neurons with common inputs embody hierarchies of functions, each of which can be reduced to input-output functions pictured in their truth tables. The rules of reduction are simple, even for fortuitous jots, and thus formalize probabilistic logic. Minimal nets of neurons are sufficiently redundant to stabilize the logical input-output function despite common shifts of threshold sufficient to alter the function computed by every neuron, and to secure reliable performance of nets of unreliable neurons. Both types of nets are flexible as to the functions they can compute when controlled by imposed changes of threshold.

The neurons, the variations of their thresholds, their excitations and inhibitions are realistic; and there remains sufficient redundancy for statistical control of growth to produce the synapsis of these stable, reliable and flexible nets.

NEUROPHYSIOLOGISTS are indebted to John von Neumann for his studies of components and connections in accounting for the steadiness and the flexibility of behaviour. In speaking to the American Psychiatric Association *(ref. 11)* he stressed the utility and the inadequacy of known

mechanisms for stabilizing nervous activity, namely, (a) the threshold of nonlinear components, (b) the negative feedback of reflexive mechanisms, (c) the internal switching to counteract changes - "ultrastability" - *(ref. 1)*, and (d) the redundancy of code and of channel. He suggested that the flexibility might depend upon local shifts of thresholds or incoming signals to components that are more appropriate to computers than any yet invented. His Theory of Games *(ref. 13)* has initiated studies that may disclose several kinds of stability and has indicated where to look for logical stability under common shift of threshold. His "Toward a Probabilistic Logic" *(ref. 12)* states the problem of securing reliable performance from unreliable components, but his solution requires better relays than he could expect in brains. These, his interests, put the questions we propose to answer. His satisfaction with our mechanisms for realizing existential and universal quantification in nets of relays *(refs. 6,15)* limits our task to the finite calculus of propositions. Its performance has been facilitated by avoiding the opacity of the familiar symbols of logic and the misleading suggestions of multiplication and addition modulo two of the facile boolean notation for an albegra that is really substitutive *(refs. 8,9,10)*. Our symbols have proved useful in teaching symbolic logic in psychological and neurological contexts *(ref. 3)*. Familiarity with them undoubtedly contributed to the invention of the circuits whose redundancy permits solution of our problems.

The finite calculus of propositions can be written at great length by repetitions of a stroke signifying the incompatibility of its two arguments. The traditional five symbols, '$\sim$' for 'not'; '$\cdot$' for 'both'; '$\vee$' for 'or'; '$\supset$' for 'implies'; and '$\equiv$' for 'if and only if', shorten the text but require conventions and rearrangements in order to avoid ambiguities. Economy requires one symbol for each of the sixteen logical functions of two propositions. The only necessary convention is then one of position or punctuation.

Since the logical probability and the truth value of a propositional function are determined by its truth table, each symbol should picture its table. When the place in the table is given, any jot serves for "true" and a blank for "false". When the four places in the binary table are indicated by '$\times$' it is convenient to let the place to the left show that the first proposition alone is the case; to the right, the second; above, both; and below, neither. Every function is then pictured by jots for all of those cases in which the function is true. Thus we write $A \overset{\times}{\cdot} B$ for contradiction; $A \overset{\times}{\cdot} B$ for $A \cdot \sim B$; $A \overset{\times}{\cdot} B$ for $A \cdot B$; $A \overset{\times}{\cdot} B$ for $B \cdot \sim A$; $A \overset{\times}{\cdot} B$ for $\sim A \cdot \sim B$; $A \overset{\times}{\cdot} B$ for $A \cdot (B \vee \sim B)$; $A \overset{\times}{\cdot} B$ for $(A \vee \sim A) \cdot B$; $A \overset{\times}{\cdot} B$ for $\sim A \cdot (B \vee \sim B)$; $A \overset{\times}{\cdot} B$ for $(A \vee \sim A) \cdot \sim B$; $A \overset{\times}{\cdot} B$ for $(A \cdot \sim B) \vee (\sim A \cdot B)$; $A \overset{\times}{\cdot} B$ for $A \equiv B$; $A \overset{\times}{\cdot} B$ for $B \supset A$; $A \overset{\times}{\cdot} B$ for $A \vee B$; $A \overset{\times}{\cdot} B$ for $A \supset B$; $A \overset{\times}{\cdot} B$ for $\sim(A \cdot B)$; and $A \overset{\times}{\cdot} B$ for tautology. The $\times$ or chi, may be regarded as an elliptical form

of Venn's diagram for the classes of events of which the propositions A and B are severally and jointly true and false; for in *fig. 1*, the chi remains when the dotted lines are omitted. Similar symbols can therefore be made, from Venn symbols, for functions of more than two arguments. Each additional line must divide every pre-existing area into two parts. Hence, for the number of arguments δ there are 2^{δ} spaces for jots and $2^{2\delta}$ symbols for functions. (*See fig. 1.*)

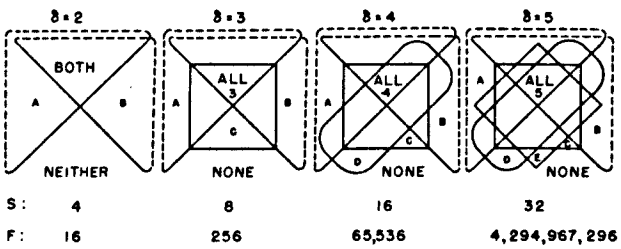


Fig.1. Venn figures with spaces for all intersections of δ classes. S is the number of spaces; F, the number of functions.

Formulas composed of our chiastan symbols are transparent when the first proposition is represented by a letter to the left of the × and the second to the right. When these spaces are occupied by expressions for logical variables, the formula is that of a propositional function; when they are occupied by expressions for propositions, of a proposition; consequently the formula can occupy the position of an argument in another formula.

Two distinct propositions, A and B, are independent when the truth value of either does not logically affect the truth value of the other. A formula with only one × whose spaces are occupied by expressions for two independent propositions can never have an × with no jots or four jots. The truth value of any other proposition is contingent upon the truth values of its arguments. Let us call such a proposition "a significant proposition of the first rank."

A formula of the second rank is made by inserting into the spaces for the arguments of its × two formulas of the first rank; for example, $(A \cdot\times B) \dot{\times} (A \times\cdot B)$. When the two propositions of the first rank are composed of the same pair of propositions in the same order, the resulting formula of the second rank can always be equated to a formula of the first rank by putting jots into the × for the corresponding formula of the first rank according to the following rules of reduction:

Write the equation in the form $(\ldots x_1 \ldots)\ x_2\ (\ldots x_3 \ldots) = (\ldots x_4 \ldots)$; wherein the x_j are chiastan symbols:

(1) If x_2 has a jot on its left, put a jot into x_4 in every space where there is a jot in x_1 and no corresponding jot in x_3. Thus,

$$(A \dot{\times} B) \times (A \times B) = (A \times B)$$

(2) If x_2 has a jot on its right, put a jot into x_4 in every space where there is a jot in x_3 and no corresponding jot in x_1. Thus,

$$(A \times B) \times (A \dot{\times} B) = (A \times B)$$

(3) If x_2 has a jot above, put a jot into x_4 in every space where there is a jot in both x_1 and x_3. Thus, $(A \dot{\times} B) \dot{\times} (A \dot{\times} B) = (A \dot{\times} B)$

(4) If x_2 has a jot below, put a jot into x_4 in every space that is empty in both x_1 and x_3. Thus, $(A \dot{\times} B) \times (A \dot{\times} B) = (A \dot{\times} B)$

If there is more than one jot in x_2 apply the foregoing rules seriatim until all jots on x_2 have been used. Put no other jots into x_4.

By repetition of the construction we can produce formulas for functions of the third and higher ranks and reduce them step by step to the first rank, thus discovering their truth values.

Since no other formulas are used in this article, the letters A and B are omitted, and positions, left and right, replace parentheses.

In formulas of the first rank the chance addition or omission of a jot produces an erroneous formula and will cause an error only in that case for which the jot is added or omitted, which is one out of the four logically equiprobable cases. With the symbols proposed for functions of three arguments, the error will occur in only one of the eight cases, and, in general, for functions of δ arguments, in one of 2^δ cases. If p_δ is the probability of the erroneous jot and P the probability of error produced, $P = 2^{-\delta} p$. In empirical examples the relative frequency of the case in question as a matter of fact replaces the logical probability.

In formulas for the second rank there are three $\times$'s. If we relax the requirement of independence of the arguments, A and B, there are then 16^3 possible formulas each of which reduces to a formula of the first rank. Thus the redundancy, R, of these formulas of the second rank is $16^3/16 = 16^2$. For functions of δ arguments, $R = (2^\delta)^\delta$.

To exploit this redundancy so as to increase the reliability of inferences from unreliable symbols, let us realize the formulas in nets of what von Neumann called neurons (3). Each formal neuron is a relay which on receipt of all-or-none signals either emits an all-or-none signal or else does not emit one which it would otherwise have emitted. Signals approaching a neuron from two sources either do not interact, or, as we have shown *(refs. 5,7)*, those from one source prevent some or all of those from the other source from reaching the recipient neuron. The diagrams of the nets of *fig. 2* are merely suggested by the anatomy of the nervous system. They are to be interpreted as follows.

A line terminating upon a neuron shows that it excites it with a value +1 for each termination. A line forming a loop at the top of the neuron

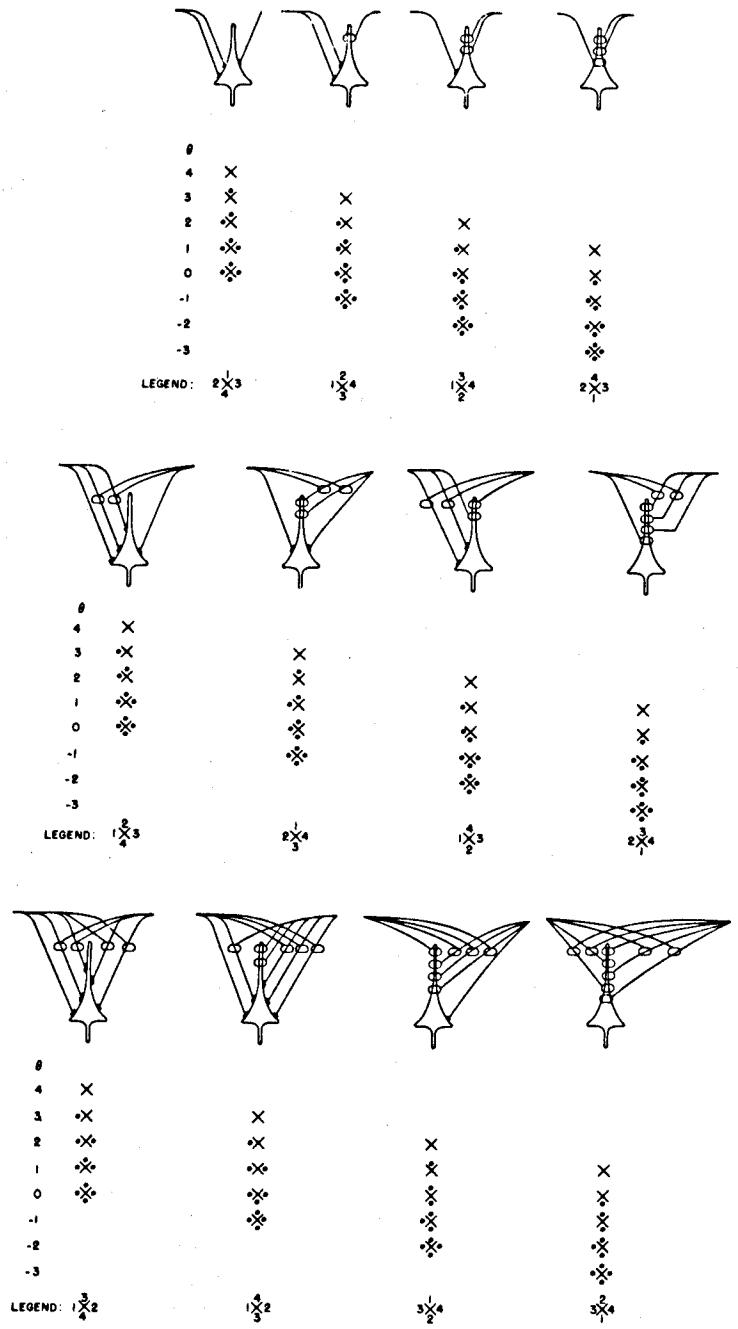


Fig.2. Synaptic diagrams for × appearing before ×.

shows that it inhibits it with a value of excitation of -1 for each loop. A line forming a loop around a line approaching a neuron shows that it prevents excitation or inhibition from reaching the neuron through that line.

Each neuron has on any occasion a threshold, θ measured in steps of excitation, and it emits a signal when the excitation it receives is equal to or greater than θ. The output of the neuron is thus some function of its input, and which function it is depends upon both its local connections and the threshold of the neuron. These functions can be symbolized by ×'s and jots beginning with none and adding one at a time as θ decreases until all four have appeared in the sequence noted in the legend for its diagram in *fig. 2*. These are the simplest diagrams fulfilling the requirement. All simpler diagrams are degenerate, since they either fail to add one jot or else add more than one jot for some step in θ. Because all 24 sequences (of which only 12 left-handed are drawn) are thus realized, we can interpret the accidental gain or loss of a jot or jots in an intended × as a change on the threshold of an appropriate neuron.

Any formula of the second rank is realized by a net of three neurons each of whose thresholds is specified; for example, see *fig. 3*. The formula can be reduced to one of the first rank whose × pictures the relation of the output of the net to the input of the net.

When all thresholds shift up or down together, so that each neuron is represented by one more, or one less, jot in its × but the reduced formula is unaltered, the net is called "logically stable."

The redundancy of formulas of the second rank provides us with many examples of pairs of formulas and even triples of formulas that reduce to the same formula of the first rank and that can be made from one another by common addition or omission of one jot in each ×, and the diagrams of *fig. 2* enable us to realize them all in several ways: For example, there are 32 triples of formulas and 64 logically stable nets for every reduced formula with a single jot. Even nets of degenerate diagrams enjoy some logical stability; for example $\dot{\times}\ \dot{\times}\ \dot{\times} = \dot{\times}$ goes to $\dot{\times}\ \dot{\times}\ \dot{\times} = \dot{\times}$.

If such nets are embodied in our brains they answer von Neumann's repeated question of how it is possible to think and to speak correctly after taking enough absinthe or alcohol to alter the threshold of every neuron. The limits are clearly convulsion and coma, for no formula is significant or its net stable under a shift of θ that compels the output neuron to compute tautology or contradiction. The net of *fig. 3* is logically stable over the whole range between these limits. Let the causes and probabilities of such shdfts be what they may, those that occur simultaneously throughout these nets create no errors.

Logically stable nets differ greatly from one another in the number of errors they produce when thresholds shift independently in their neurons and the most reliable make some errors; for example, the net of *fig. 4*.

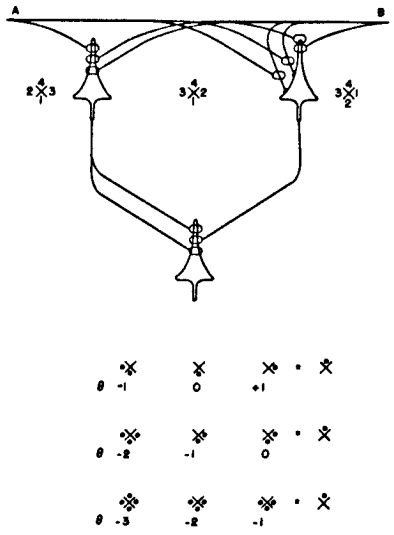
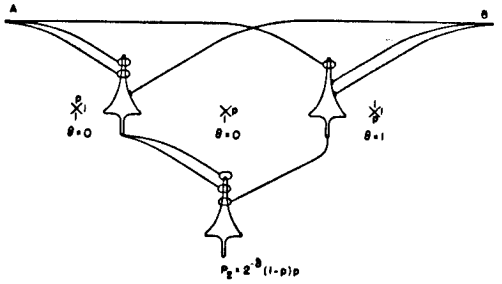
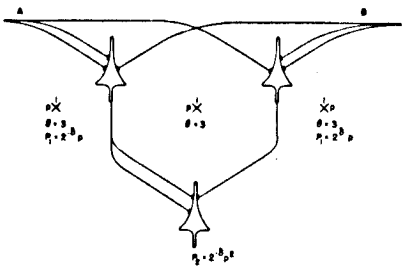


Fig.3. A logically stable net for $\dot{\times}$.



CONDITION	REDUCED SYMBOL	ERRORS	
		CASE	PROBABILITY
$\times\ \times\ \times$	$\times$		
$\times\ \times\ \times$	$\times$		
$\times\ \times\ \times$	$\times$		
$\times\ \times\ \times$	$\times$	$\times$	$(1-p)p(1-p)$
$\times\ \times\ \times$	$\times$	$\times$	$(1-p)p(p)$
$\times\ \times\ \times$	$\times$		
$\times\ \times\ \times$	$\times$		
$\times\ \times\ \times$	$\times$		
$\times\ \times\ \times$	$\times$		

Fig.4. A best stable net for $\times$.



CONDITION	REDUCED SYMBOL	ERRORS	
		CASE	PROBABILITY
$\dot{\times}\ \dot{\times}\ \dot{\times}$	$\dot{\times}$		
$\dot{\times}\ \dot{\times}\ \dot{\times}$	$\dot{\times}$		
$\dot{\times}\ \dot{\times}\ \dot{\times}$	$\dot{\times}$		
$\dot{\times}\ \dot{\times}\ \dot{\times}$	$\dot{\times}$		
$\dot{\times}\ \dot{\times}\ \dot{\times}$	$\dot{\times}$		
$\dot{\times}\ \dot{\times}\ \dot{\times}$	$\dot{\times}$		
$\dot{\times}\ \dot{\times}\ \dot{\times}$	$\times$	$\times$	$(p)p(1-p)$
$\dot{\times}\ \dot{\times}\ \dot{\times}$	$\times$	$\times$	$(p)p(p)$

Fig.5. A best unstable net for $\dot{\times}$.

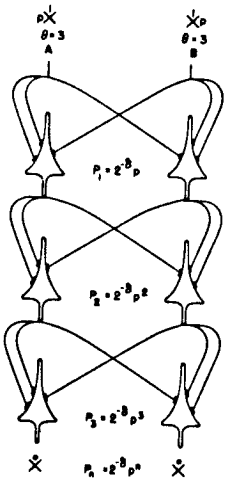


Fig.6. Unstable improvement net for $\dot{\times}$.

To include independent shifts, let our chiastan symbols be modified by replacing a jot with 1 when the jot is never omitted and with p when that jot occurs with a probability p, and examine the errors extensively as in *fig. 4*. Here we see that the frequency of the erroneous formulas is $p(1-p)$, and the actual error is a deficit of a jot at the left in the reduced formula in each faulty state of the net, i.e., in one case each. Hence we may write for the best of stable nets $P_2 = 2^{-2}\, p(1-p)$. The factors p and $(1-p)$ are to be expected in the errors produced by any net which is logically stable, for the errors are zero when $p = 0$ or $p = 1$. No stable net is more reliable.

No designer of a computing machine would specify a neuron that was wrong more than half of the time; for he would regard it as some other neuron wrong less than half of the time; but in these more useful of logically stable circuits, it makes no difference which way he regards it, for they are symmetrical in p and $1 - p$. At $p = 1/2$, the frequency of error is maximal and is $P_2 = 2^{-2}\ 1/2(1 - 1/2) = 1/16$, which is twice as reliable as its component neurons for which $P_1 = 2^{-2}\ 1/2 = 1/8$.

Among logically unstable circuits the most reliable can be constructed to secure fewer errors than the stable whenever $p < 1/2$. The best are like that of *fig. 5*. The errors here are concentrated in the two least frequent states and in only one of the four cases. Hence $P_2 = 2^{-2}\, p^2$.

Further improvement requires the construction of nets to repeat the improvement of the first net and, for economy, the number of neurons should be a minimum. For functions of δ arguments each neuron has inputs from δ neurons. Hence the width of any rank is δ, except the last, or output, neuron. If n be the number of ranks, then the number of neurons, N, is $\delta(n-1) + 1$.

Figure 6 shows how to construct one of the best possible nets for the unstable ways of securing improvement with two output neurons as inputs for the next rank. The formulas are selected to exclude common errors in the output neurons on any occasion. In these, the best of unstable nets, the errors of the output neurons are $P_n = 2^{-\delta} p^n$.

[Whether we are interested in shifts of threshold or in noisy signals, it is proper to ask what improvement is to be expected when two or three extra jots appear in the symbols. With our nondegenerate diagrams for neurons a second extra jot appears only if the first has appeared, and a third only if the second. If the probability p of an extra jot is kept constant, the probability of two extra jots is p^2 and of three is p^3. Examination of the net in *fig. 6* shows that $P_2 < P_1$, if $p + p^2 + p^3 < 0.15$ or $p < 0.13$. To match Gaussian noise the log of successive p's should decrease as $(\Delta\theta)^2$, or 1, p, p^4, p^9 giving $P_2 < P_1$ for $p < 0.25$. The remaining errors are always so scattered as to preclude further improvement in any subsequent rank.]

When common shifts of θ are to be expected, or all we know is $0 < p < 1$, a greater improvement is obtained by alternating stable and unstable nets as in *fig.7*, selected to exclude common errors in its output neurons. For n even

$$P_n = 2^{-\delta}\, p^{n/2}\, (1-p)^{n/2}$$

and the expected error is

$$2^{-\delta} \int_0^1 p^{n/2}\, (1-p)^{n/2}\, dp = 2^{-\delta}\, \frac{\left(\frac{n}{2}\right)^2}{(n+1)!}$$

which is less than with any other compositions of $\delta = 2$ nondegenerate diagrams.

When $\delta = 3$, the redundancy, $R = \left(2^{2^\delta}\right)^\delta$, provides so many more best stable and best unstable nets that the numbers become unwieldy. There are $\left(2^{2^\delta}\right)^{\delta+1}$ nets for functions of the second rank each made of 4 neurons to be selected from 8! diagrams with 9 thresholds apiece. Formerly *(ref. 4)* I said it was clear that the best stable and unstable nets for $\delta < 2$ are better than those for $\delta = 2$ only in the factor $2^{-\delta}$ for error in a single case. That is only true if the nets are composed of nondegenerate diagrams alone. With neurons $\delta = 3$, a single degenerate diagram for the output neuron permits the construction of a logically stable net with $P_2 = 0$, even with independent shifts of θ sufficient to alter the logical function computed by every neuron, as seen in *fig. 8*. The same degenerate diagram for the $\delta = 3$ output neuron receiving inputs from three nondegenerate $\delta = 2$ neurons, selected to make but one error in each case, is likewise stable and has an error-free output despite independent shifts of θ, as is seen in *fig. 9*.

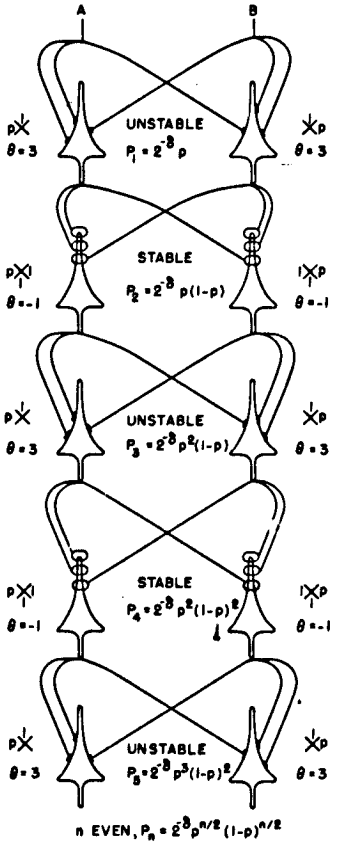


Fig.7. Alternating improvement net for $\times_{\cdot}$.

None of these nets increases reliability in successive ranks under von Neumann's condition that neurons fire or fail with probability p regardless of input; but they are more interesting neurons. They are also more realistic. In the best controlled experiments the record of a single unit, be it cell body or axon, always indicates firing and failing to fire at

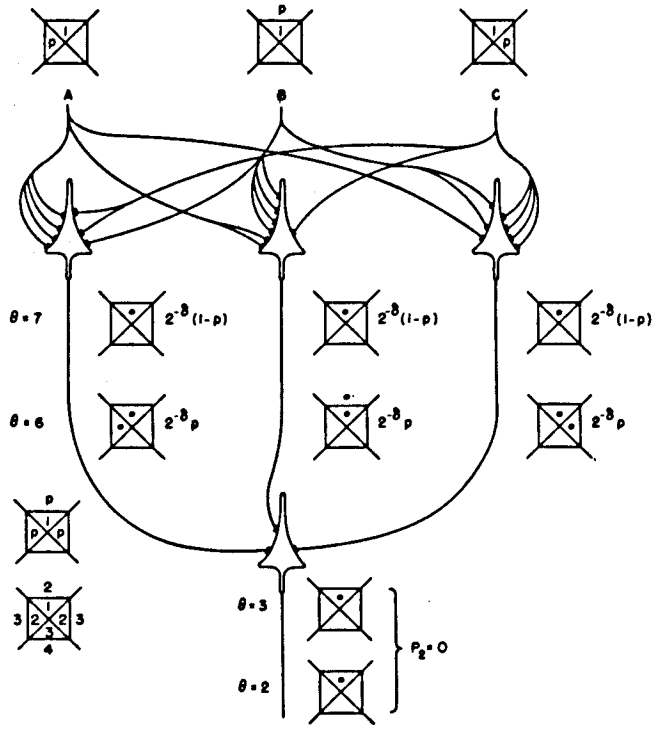


Fig.8. Input $\delta = 3$, output degenerate $\delta = 3$ neuron for $\overset{*}{x}$.

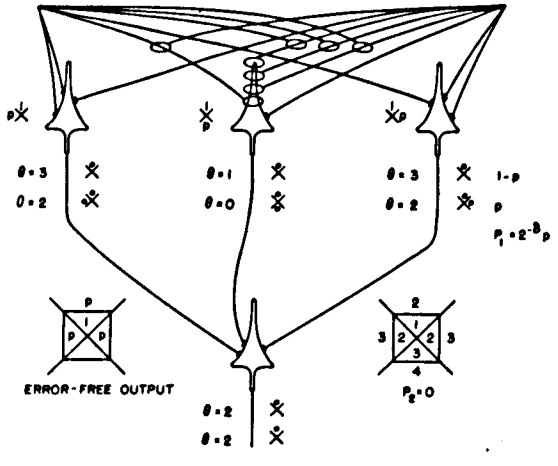
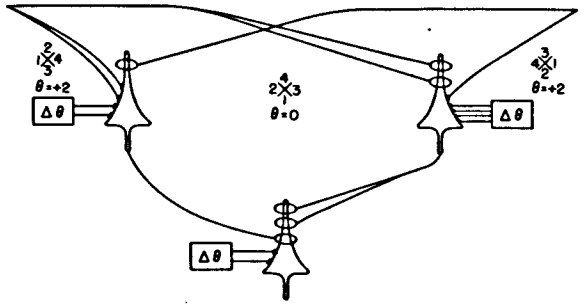
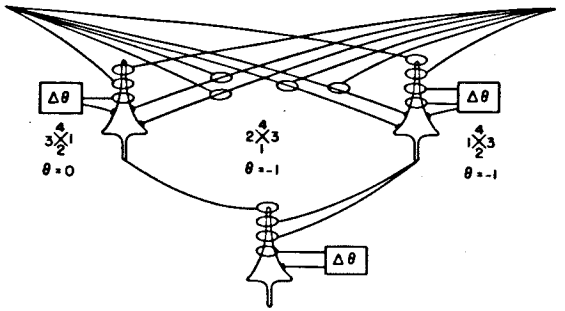


Fig.9. Input $\delta = 2$, output degenerate $\delta = 3$ neuron for $\overset{*}{x}$.



CONDITION Δθ	Δθ	Δθ	14 REDUCED SYMBOLS
0	0	-1	
0	0	-2	
-1	0	-1	
-1	0	-2	
-2	0	0	
0	-1	-1	
0	-1	-3	
-1	-1	-2	
0	-2	-1	
0	-2	-4	
-1	-2	-3	
-1	-2	-4	
-2	-2	-1	
-2	-2	-3	

Fig.10. Flexible net, always unstable.



CONDITION Δθ	Δθ	Δθ	9 REDUCED SYMBOLS
+1	+1	+1	
0	0	0	
-1	-1	-1	
+1	+1	+2	
0	+1	-1	
+1	0	+2	
+1	0	+1	
0	-1	0	
0	-1	+2	
0	-1	+1	
0	-1	0	

Fig.11. Flexible net, logically stable for $\dot{\times}$.

near threshold values of constant excitation, whether it is excited transynaptically or by current locally applied. At present, we do not know how much of the observed flutter of threshold is due to activity of other afferent impulses and how much is intrinsic fluctuation at the trigger point. We have not yet a reasonable guess as to its magnitude in neurons in situ, and for excised nerve we have only two estimates of the range: one, about 2 per cent; and the other, 7 per cent *(refs. 2, 14)*. Our own measurements on a single node of Ranvier would indicate a range of 10 per cent. To discover regularities of nervous activity we have naturally avoided stimulation near threshold. Now we must measure the intrinsic jitter, for this determines both the necessity of improving circuits and the maximum number of afferent terminals that can be discriminated. Eventually we will have to take into account the temporal course of excitation and previous response, for every impulse leaves a wake of changing threshold along its course.

Despite the increase in reliability of a net for a function of the second rank some of these nets can be made to compute as many as 14 of the 16 reduced formulas by altering the thresholds of the neurons by means of signals from other parts of the net, as in *fig. 10*, and even some logically stable nets for triples of formulas can be made to realise 9 of the 16, as in *fig. 11*. Even this does not exhaust the redundancy of these nets, for both of them can be made to compute many of these functions in several ways.

The diagrams of *fig. 2* were drawn to ensure a change in function for every step in θ. Actual neurons have more redundant connexions. We are examining how to use this redundancy to design reliable nets the details of whose connexions are subject to statistical constraints alone. This is important because our genes cannot carry enough information to specify synapsis precisely.

For the moment, it is enough that these appropriate, formal neurons have demonstrated logical stability under common shift of threshold and have secured reliable performance by nets of unreliable components without loss of flexibility.

This work was supported in part by Bell Telephone Laboratories, Incorporated; in part by The Teagle Foundation, Incorporated, and in part by National Science Foundation.

REFERENCES

1. ASHBY, W. ROSS, Design for a Brain, *John Wiley and Sons, Inc., New York.* (1952).
2. BLAIR, E. A. and ERLANGER, J. A, A comparison of the characteristics of axons through their individual electrical responses. *Am.J.Phys.*, 1933, **106**, 524.
3. McCULLOCH, W. S. Machines that know and want. Brain and Behaviour, A Symposium, edited by W. C. Halstead; Comparative Psychology Monographs 20, No.1, *University of California Press, Berkeley and Los Angeles.* (1950).
4. McCULLOCH, W. S. *Quarterly Progress Report of the Research Laboratory of Electronics, M.I.T. (Neurophysiology)*, October 1957.
5. McCULLOCH, W. S., HOWLAND, B., LETTVIN, S. Y., PITTS, W. and WALL, P. D. Reflex inhibition by dorsal root interaction. *J. Neurophysiol.*, 1955, **18**, 1.
6. McCULLOCH, W. S. and PITTS, W. A logical calculus of the ideas immanent in nervous activity. *Bull..Maths. and Biophys.*, 1943, **5**, 115.
7. McCULLOCH, W. S., WALL, P. D., LETTVIN, J. Y. and PITTS, W. H. Factors limiting the maximum impulse transmitting ability of an afferent system of nerve fibers. Information Theory, Third London Symposium, edited by C. Cherry. *Academic Press, Inc., New York.* (1956).
8. MENGER, K. Algebra of Analysis, Notre Dame Mathematical Lectures, No.3. *University of Notre Dame, Notre Dame, Ind.* (1944).
9. MENGER, K. Analysis without variables. *J. Symbol. Logic*, 1946, **1**, 30.
10. MENGER, K. General algebra of analysis. *Reports of a Mathematical Colloquium, Series 2, issue 7.* (1946).
11. NEUMANN, J. von. Fixed and random logical patterns and the problem of reliability. *American Psychiatric Association, Atlantic City, May 12.* (1955).
12. NEUMANN, J. von. Probabilistic Logics and the Synthesis of Reliable Organisms from Unreliable Components. Automata Studies, edited by C. E. Shannon and J. McCarthy. *Princeton University Press, Princeton, N.J.* (1956).
13. NEUMAN, J. von. and MORGENSTERN, O. Theory of Games and Economic Behaviour. *Princeton University Press, Princeton, N.J.* (1944).
14. PECHER, C. La fluctuation d'excitabilité de la fibre nerveuse. *Arch. Internat. de Physiol.*, 1939, **49**, 129.
15. PITTS, W. and McCULLOCH, W. S. How we know universals. The perception of auditory and visual forms. *Bull. Maths. and Biophys.*, 1947, **9**, 127.

DISCUSSION ON THE PAPER BY DR. W. S. McCULLOCH

MR. E. A. NEWMAN: From the point of view of a computer designer dealing with logical circuits, the ideas expressed in Dr. McCulloch's paper seemed to me brilliant in their inherent simplicity and great power. If one uses a number of two-input devices together, as Dr. McCulloch does, and wants the overall result of one two-input device, then evidently the redundancy is very great, and the overall system can be chosen in a way that gives great stability.

Any complicated data-processing system must store and handle a hierarchy of ideas. Dr. McCulloch's scheme enables such a hierarchy to be handled in a way that gives little or no redundancy for the mass of non-vital detail, and hence maximum overall storage efficiency, and at the same time great redundancy for those basic patterns which are absolutely vital.

DR. GREY WALTER: I should like to comment on the notion that a high degree of trustworthiness is essential and attainable for a nervous system, even when its neurons are liable to failure. In Dr. McCulloch's schemata we see that this can be so but there are interesting limits and exceptions. The story is, I believe, that some of these ideas occurred to Dr. McCulloch in a conversation with von Neumann about the latter's ability to drive his car when he had been drinking. To me it seems equally surprising that small quantities of alcohol and other drugs do actually affect behaviour rather dramatically - in concentrations that is, that do not affect the individual nerve cells appreciably. This vulnerability to general assault by infiltration has to be accounted for as well as the apparent indifference of the nervous system to the fate of individual neurons.

One can conceive of the input to a nervous system as consisting, among other things, of a chain or cascade of neurons. Now if the passage of an impulse from one end to the other is a question of probability, then in the ideal normal case when transmission is perfect, the probability of an impulse which enters this chain from the periphery getting to some part of the nervous system is unity. In physiological terms this means that the impulse from each neuron is of supraliminal intensity at the synapse with each succeeding neuron. Now if some general influence - such as alcohol - is brought to bear so that the probability of transmission for each synapse is slightly reduced, then the probability of transmission along the whole chain will be diminished by a greater factor, depending on the number of links or synapses; the cascade acts as a probability amplifier,

or more precisely attenuator since the value of the probability will usually be fractional. For example, if there are ten neurons in the cascade and the general influence, von Neumann's martinis or a sleepless night or whatever, reduces the impulse passage probability from unity to 0.5 for each synapse, then the probability of an impulse getting right through the ten links will be reduced to $(1/2)^{10}$, and a chance of a thousand to one against means not merely collision but coma. A series chain of elements is therefore very vulnerable and furthermore when it makes up part of a network, the system is likely to be anisotropic - its lengthways probability conductance can easily become quite different from its sideways conductance, which in the example given will still be 0.5 for impulses crossing the chain through a single synapse.

In the type of neuron circuitry schematised by Dr. McCulloch, this might be a valuable aid to understanding the specific effects of general influences such as drugs, hormones and the like. It has been shown that most of the psychotropic compounds exert their action mainly by way of the complex neuron networks in the base of the brain. These reticular formations control the inputs and outputs of the brain and are known to be particularly vulnerable also to mechanical disturbance in spite of their situation. If we consider these systems in the terms suggested by Dr. McCulloch we can see that his schemata allow for individual component failure but expose the system to interference by general influences. This can account for the variety and intricacy of the homeostatic mechanisms and structural safeguards that support the brain as a whole, while the fate of the individual neuron is left to chance. The analogies with computer design and with social systems are obvious.

DR. W. K. TAYLOR: Dr.McCulloch recognises many of the limitations of his model neurones but I should like to know whether he has considered one point about them that particularly worries me and which I can demonstrate with a simple example of two inputs and a threshold of 2. If both inputs consist of 1 millisecond duration pulses at a steady rate of 200 per second, the unit will give 200 pulses per second out if the inputs are in step but no output at all if the pulses are more than one millisecond out of step, since they will then never summate to reach the threshold of 2. We know that there is no accurate timing device in the nervous system, as there is in a digital computer, and that variable delays can occur in nerve fibres and at synapses. If the delay is more than the pulse width at one input of the McCulloch model neuron unit the output falls from 200 pulses/sec. to zero in the example given and I suggest that the unit would be a particularly unlucky reckoner.

DR. W. S. McCULLOCH: I should like to answer those two separately first. In the first place, the dissimilarity of properties from neuron to neuron, where we happen to know them, is very, very, great. Patrick Wall has been working, during the last year, on the afferents to the spinal cord, and the first relays that ascend in the dorsolateral portion of the spinal cord presumably destined for the cerebellum. They can be excited, some by stretch receptors and some by touching the skin. If you map the area that any one axon brings into the spinal cord, you get a reasonable area. It is small. If you map the area of the skin covered by any one of these cells whose axons go up towards the cerebellum - this is the skin I am talking about - you find it is about 25 times as great. Now comes the curious thing. You can take that animal and soak him in barbiturates and you do not get a change in the area. You can strychninise him till he is ready to jump off the table, and you do not get a change in the area. Exactly the reverse is true, of course, of the motor neuron. Here you have two neurons, each of which has direct connections from the periphery but they are entirely different in their behaviour: one has no subliminal fringe that we can discover, the other certainly has. Now I am well aware that the behaviour of these cells is also quite different in response to an input. A motor neuron under these circumstances, when you give it a boomp coming in - the typical behaviour of these cells is to put out a whole series of impulses when they receive one; and what you change when you drug that animal is you increase the length of the barrage with strychnine and you decrease it with barbiturates, but as long as it will respond at all it will respond to the whole area - quite a different organisation from that of motor neurons.

Now my neurons are very far removed from a good Hodgkin-Huxley neuron. The reason is perfectly obvious. I only want to embody in these neurons a certain logical aspect of a theory. That is all that I propose to do. I do not think they are particularly realistic. I think they are more realistic than the Eccles-Jordan proposals of mere flip-flops for neurons, in that they do introduce a notion of threshold; and I think it is out of this notion of threshold that the power of this kind of logic comes.

Fundamentally, we in biology are famous for our inability to handle mathematics. That is certainly true of us, by and large; but the other thing is this, that almost all mathematics was made for the physicist and not for the biologist. We do not generally have the mathematics we need, and we go on to invent it for ourselves, playing with spools of string and bits of rubber. What I wanted with my map was to get the rubber in the right place. I wanted not $f(x)$ where x was probable but where the f was probable. That is all that those normal neurons are supposed to do.

The next thing I want to say pertains to the second question. The second question is of this kind: what a neuron will do under these circumstances depends on the threshold of the neuron. If the threshold of

the neuron is 2 and there are two afferents to it, it is a coincidence detector, and if we say one afferent has a rate A and the other has a rate B, the output will be some constant times the product of A and B (these are frequencies) - it becomes a product-taker. All coincidence detectors will do this. If the value of the threshold was 1, then you will get the simple sum, so until you have specified thresholds of these kind of components, you do not know what they are going to do.

DR. A. M. UTTLEY: And being a product taker, it will be a very lucky reckoner if you want cross-correlation?

DR. W. S. McCULLOCH: That is right.

DR. F. ROSENBLATT: I would like to add just a couple of comments to Dr. McCulloch's defence of his neurone. First of all, we have made several observations on the effect of changing thresholds in statistically connected networks in connexion with our work on the perceptron; and one of the interesting things is that it makes very little difference in such networks what the threshold is, within rather wide limits, provided the threshold is fairly high. Given a threshold sufficiently greater than zero, then we can practically double this threshold and a network which is made up of essentially random connections still learns in about the same way, with a greater or a lesser efficiency. It is also possible to change the properties of the neurones in rather drastic ways without seriously altering their performance. For example, if we substitute the model which I proposed in my own paper the continuous transducer neurone which responds on a frequency basis, in place of one which responds at fixed time increments - which I think really takes care of Dr. Taylor's problem rather nicely in most cases - it turns out that there is a negligible difference between such neurones and Dr. McCulloch's neurone. As a matter of fact, the difference is so negligible that it has proved to be no longer worthwhile to maintain this distinction in our own work because we now find we can analyse these circuits much more satisfactorily using something much more like a conventional McCulloch-Pitts neurone and come out with almost identical numerical results. The difference between a continuous transducer neurone and a neurone which responds at fixed time increments makes very little difference. However, some non-zero threshold is essential; as soon as the threshold falls to zero the behaviour of the system goes to pot and we get essentially no learning out of it at all.

DR. M. L. MINSKY: I do not quite follow one of Dr. Walter's remarks. It seems to me that if, under the influence of different drugs, you can find grossly different local neural behaviors and yet the same functional behavior, this is evidence that you do have some sort of mechanism of the

kind Dr. McCulloch suggests, in which the properties of neurones can vary without changing the properties of larger sub-nets.

In connection with Dr. Rosenblatt's remark, I do not see a direct connection between that kind of reliability and this. If you have a machine in which the initial properties of the neurons do not critically affect the learning behavior, then indeed early changes in threshold should not have much effect. But it does *not* follow from that alone that after the machine has become organized you can go ahead and change thresholds grossly and expect it to maintain the same behavior. I would like to ask him if there is evidence for a high degree of this kind of stability in a "Perceptron" that has learned.

DR. F. ROSENBLATT: Most of our evidence does concern initial changes in threshold; this is quite true. However, we also have some evidence that we can change the threshold (within reasonable limits) *after* the system has learnt, and the performance will not be too seriously affected. There will be some deficit: if we lower the threshold we are allowing additional cells to respond, which would not otherwise have responded and this does correspond to some introduction of noise into the system. However, the statistical bias which serves as our 'memory' phenomenon tends to be retained in spite of the introduction of additional noise due to the changes in threshold. This does not have quite the type of stability Dr. McCulloch proposes, it is true. However, we do have here a system which is particularly insensitive to threshold changes, even after learning.

DR. M. B. BARLOW: There is one small point - Dr. McCulloch is a physiologist, and I think it would be rather dreadful if everybody thought that all physiologists would agree with a model neuron like that. I wonder whether he would stand up and say that he does not believe real neurons are as simple as the neurons he is talking about.

DR. W. S. McCULLOCH: I thought I had made that very clear. I have no notion that these neurons of mine are anything more than an embodiment of arithmetic. I know that a real neuron has an 8th order non-linear differential equation, which I cannot on inspection handle at all. That is not the kind of device with which to explain a logical problem when what you are really trying to do is set up a probabilistic logic.

PROF. J. Z. YOUNG, CHAIRMAN: Would it be better not to use the word "neuron" then?

DR. W. S. McCULLOCH: I am sorry. I call them formal neurons because this is the word von Neumann had used and I followed him through on it, thats all. The great point is they are not to be confused with ordinary relays that

merely close or open a circuit. Everyone says, "Oh, is this not the same as Shannon's Hammock net?" No, it is not. Shannon and I have been over this problem. It is an entirely different problem. His is one of maintaining or obtaining continuity through a system, and its solution is entirely different from mine. They do not even map on each other.

PROF. Y. BAR-HILLEL: I should like to call Dr. McCulloch's attention to other existing pictorial representations of the truth-functional connectives, especially to the diagonal symbolism of Charles S. Peirce (of some 80 years standing), the wheel symbolism of the late Polish logician Stefan Lesniewski, and the recent trapezoid symbolism of W. T. Parry. Some of these symbolisms might perhaps be slightly more convenient for Dr. McCulloch's purposes than his own. Compare, e.g., the following three notations for material implication, if A then B, the customary Russelian symbolization of which is A ⊃ B:

McCulloch	Peirce	Lesniewski	Parry
A B	A  B	A B	A ⊃ B

You will notice that they are all based on the same principle, i.e. the exploitation of the Boolean expansion (the developed disjunctive normal form).

For all this, see the two papers by W. T. Parry, *(ref. 1)* and G. B. Standley *(ref. 2)*. The second paper describes an ideographical method of computation with Parry's trapezoidal symbols which is again closely reminiscent of Dr. McCulloch's procedure.

MR. E. A. NEWMAN: I would refer to the comments of Dr. Taylor and Dr. Barlow.

As an engineer it seemed to me evident that when Dr. McCulloch refers to, say, an "and" element with two inputs and one output, he is referring to a generallised element capable of performing the "and" operation. He could, for example, quite well be referring to a device in which the two inputs took the form of pulse trains having a frequency proportional to the logarithm of the drive, and the output a pulse train having the sum frequency - or in fact to any of the hundreds of devices with the same logical implications. I am very surprised therefore that physiologists should think that, when Dr. McCulloch adopts a simple logical notation for the functional behaviour of his neurons, he should be implying that they obtain their functional behaviour in precisely the way directly implied by the diagrams.

REFERENCES

1. PARRY, W. T. A new symbolism for the propositional calculus. *Journal of Symbolic Logic*, 1954, **19**, 161.
2. STANDLEY, G. B. Ideographic computation in the propositional calculus. *Journal of Symbolic Logic*, 1954, **19**, 169.

DR. M. B. BARLOW: I was prompted to make my remarks by the fact that engineers and others have, in the past assumed rather too simple properties for neurons. As long as they all stop doing so, I am very happy.

DR. W. S. McCULLOCH: I should like to make one remark, and I should like to address a question to your chairman. We expected, from what we thought to be the dimensions of a node of Ranvier along an axon that we should necessarily run into a jitter of thresholds of less than one per cent. On making the measurements under ideal conditions, the actual jitter that we encounter is of the order of ten per cent. Are we over-estimating the area of the trigger point?

PROF. J. Z. YOUNG, CHAIRMAN: I should think very likely. It is a technical point, but the actual available membrane surface there is extremely difficult to compute. By electron-microscopy it looks very different from by light microscopy because there are folds of the Schwann cells all over, and it may be many times less than one would suppose from the actual gap seen in the light microscope.

DR. W. S. McCULLOCH: You see, it is this jitter of threshold which compels me to look for a probabilistic way of handling it, because if that threshold is jittering - the function that the cell is going to compute, if that's its striking point is certainly going to be shifting; and unless I'm prepared to take that into consideration I think my imitations of real neurons in words or in chalk are very unrealistic.

PROF. J. Z. YOUNG, CHAIRMAN: It is a very striking thing if there is that threshold change. Is there other evidence of that?

DR. McCULLOCH: The other evidence goes way back to Blair and Erlanger in those early papers. They have about 20 or 30 nodes of Ranvier in series. They came up with values of the order of 3 per cent.

SESSION 4A

PAPER 4

MEDICAL DIAGNOSIS AND CYBERNETICS

by

DR. FRANÇOIS PAYCHA

BIOGRAPHICAL NOTE

Dr. François Paycha, born at Narbonne, studied medicine at the University of Montpellier. His first researches were concerned with the embryology of the eye, later using the distribution of radioactive phosphorus P32 to study the structure of the tissues and for the detection of tumours.

He was then appointed to the National Centre of Scientific Research. While in charge of a hospital clinic, he noted the considerable differences in the diagnoses of conscientious and knowledgeable practitioners and those advanced by the hospital. In view of the special need for exact diagnosis in medicine he made a study of the causes of these differences.

After theoretical research, he made the first "Medical Memory' in 1953 with the help of Bull and later of I.B.M. He studied the structure of a three-symbol logic which is applicable to medical problems and in general.

After a year in the service of Prof. G. E. Jayle, he abandoned pure research and entered industry.

MEDICAL DIAGNOSIS AND CYBERNETICS

by

DR. FRANÇOIS PAYCHA

SUMMARY

I am going to analyse briefly, and describe, the logical structure of Medicine.

On the basis of this study, I shall show how the results thus obtained may be wholly applied to other subjects.

Then I shall state in detail how and why certain branches of activity recognize other forms of logic.

Lastly, I shall show how one may conceive a general system of logic, which is normative, but only in terms of the nature and development of each science, regarded as a special case of a general rule.

A convenient and concise way of introducing to Medicine those versed in various other subjects, is to outline its history. This will be brief and incomplete, giving no names or different stages (which would be too ... involved), my principal aim being to show the individual character of Medicine.

This form of presentation is necessary in order to make the nature of Medicine clear to those who have had treatment because the patient does not see things in the same light as the doctor.

HISTORY OF MEDICINE

The first man or woman to pour fresh water on a painful wound was performing the first piece of therapy by that act; and the first man or woman to become aware of the approaching demise of a fellow-creature thereby made the first prognosis.

The desire to relieve pain is probably as old as the world itself, and concern with suffering undoubtedly dates just as far back.

Regarded in this way, on the basis of human suffering, Medicine is probably the oldest of the sciences.

In the beginning, its aims were unformulated, its activities undirected, but it has since become a discipline governed by the desire to relieve pain.

Medicine has gradually dissociated itself - though incompletely as yet - from magic, voodooism and superstition; in short, from a series of practices - highly irrational, to say the least - whose existence was justified by the ineffectiveness of the drugs which the "doctors" (they must be given this name) used.

Some of the first practitioners turned their activities towards the making of drugs, but their work soon became out of date; the secrecy with which they surrounded their ridiculously ineffective recipes and the naive character of those which have been handed down to us - likewise very ineffective - now only have an anecdotal value. Nowadays our attention is directed ironically enough, much less to the drugs themselves than to the phials and bottles which contained them and which are the delight of archaeologists.

Others, more moderate in their aims wished to know about the diseases of man before endeavouring to cure him. Today, the nosological framework of their descriptions seems vast, tenuous and shapeless; but the slender thread of their clinical observation, made two thousand years ago, still remains valid today and is recorded in its entirety in the huge network of innumerable subjects and interrelationships forming our present knowledge.

We can already perceive the division which is going to take place. Even in the time of Hippocrates, it was difficult for one brain to know everything, for one man to do everything, diagnosis as well as therapy.

Gradually this tendency took hold, and nowadays we have two distinct branches, both equally indispensable: medicine and pharmacy.

In passing, stress must be laid on this process whereby a single discipline subsequently divides into two or sometimes several different parts, under the pressure of increasing complexity of the relevant data.

In this study, we shall consider Medicine and Pharmacy as a whole.

Medicine, then, is a discipline defined by its particular aim of curing the sick. It is this aim which governs the activities of the doctor in terms of opportunities for action, thus all methods and all techniques are justified.

It must be noted, however, that such definitions, made and presented "a posteriori", do not correspond to any logical arrangement within the discipline.

It is the aim which gives it its unity; the purpose of Medicine is to cure.

HOW IS MEDICINE PRACTISED?

This ancient branch of knowledge is represented by the medical practitioner; we shall therefore establish the logical structure of Medicine by studying his activities.

The point where Medicine and pain come together is in the consulting room, where on the one hand we have the patient and on the other, the representative of Medicine, the practitioner.

What takes place during the consultation?

There are two indisputable facts: at the beginning of the interview the doctor knows nothing about his patient except that he is ill, and at the end of it the patient goes out provided with a prescription with which he obtains the medicaments to cure him, or intended to do so within the limits of our knowledge. If we confine ourselves to the traditional system, the consultation consists of various parts, as follows: the questioning, the general examination, palpation, inspection, examination with instruments. When these have been carried out, the doctor makes out a case-sheet which, above all else, must be complete. He makes his diagnosis, arranges for further examinations, perhaps, and prescribes treatment.

It must be stressed that in this description, the part devoted to establishing symptoms is fully developed, but the part leading to diagnosis, i.e. to the affirmation that a patient is suffering from such and such a complaint, is skimped. Much emphasis is laid upon the value of a proper examination, a complete record of symptoms, palpation carried out gently and correctly, but no indication is given of the way in which all this material is put together.

Thus is the point to which I would like to draw attention. To make the study easier, we shall transcribe the medical data into cybernetic language.

This we shall call all the particulars we have about the patient and the ailments "information".

We shall call all the actions by which the doctor obtains information about his patient the "acquisition of information", which thus comprises the general examination, palpation, questioning the patient, special examinations: in brief, all the semiological and laboratory techniques. In this connection, we should note that knowledge of a blow on the right side is just as much a bit of information as a laboratory report stating: inductance 45 Henries, or a detailed report from the heart specialist.

We shall call "information processing" all the mental processes whereby use of the information acquired leads to the affirmation that "this patient is suffering from such and such a complaint".

It must be noted that for the time being I have merely given cybernetic names to functions already known for a long time. One may therefore ask whether introduction of these new terms is justified, whether any new contribution is made by this simple change in vocabulary.

. More generally, we must prove the existence of a problem in diagnosis and of therapy - a problem which is far from obvious, especially to the medical practitioner.

The latter in fact regards diagnosis as a self-evident affirmation; in most cases he will remark that for centuries diagnoses have been made and treatments prescribed, without anyone giving their attention to the mechanism, but that the system has, nevertheless, worked. This view of the problem, which consists in ignoring it, is associated - paradoxically enough - with the high professional scruples of the doctor.

Let us suppose, in fact, that a practitioner has made a diagnosis *L* and prescribed treatment *K* for a patient; he has allowed for all contingencies and all the pathological and therapeutic possibilities before arriving at *K* and *L*. Thus in all good faith, he does not think that there can be any possible conclusions other than these.

However - and in this lies the justification for this study - it has never been shown that the nature of Medicine is such that the whole of it can be known by a single doctor.

Indeed, we have already seen, in the brief history given above, how Medicine, or the art of tending the sick, is already divided into two branches: Medicine proper and Pharmacy, and it has been so divided for a long time. Now during the last thirty years we have seen Medicine proper split up in its turn into ophthalmology, neurology, pediatrics, geriatrics, obstetrics, gynaecology, oto-rhino-laryngology, and so on.

And now a new tendency is becoming apparent, a kind of super-specialisation, the result of which is that within each special field, new, independent branches are tending to form, one making a study of and treating binocular vision, another, phonation and so on. Fragmentation of this kind is an excellent thing, in the sense that it leads to a good knowledge of the subject concerned; it is justified - and this is the important thing - by the multiplicity of data applied.

There is an unfortunate corollary to it, however, and that is, the different specialists are obliged to ignore the rest of Medicine.

Now an ailment is never confined to a single organ, and such specialisation inevitably leads to a Medicine of organs, a therapy for organs, neglecting the essential indivisibility of the human being.

How could it be otherwise? The specialist has all his attention, all his faculty of memory, all his actions directed towards a tiny sphere of activity; he cannot multiply himself by a number corresponding exactly to the number of specialist fields.

We thus reach an impasse, with continued progress in the various branches of knowledge on the one hand, and our inability to use them all at once on the other, while they are all necessary for the proper practice of Medicine.

Although specialisation may be a means of study, it cannot be the best way to making cures.

After all, the problem would be a minor one if we could be sure that the specialist could now somehow or other link up the smallest details in his own field with the whole of pathology.

Now this is by no means the case; already, details are escaping the attention of the doctor within his own special field, and he is finding it more and more difficult to keep up with ideas in the wider field.

To express this more specifically, we can consider the fact that there are some 4,785 periodicals published in the world. If we allow for a monthly issue, with four articles of interest in each issue, anyone who wished to keep up to date would have to read - and remember - some 19,140 articles a month, or 638 a day. This assumes that by previous study he knows all the basic works and especially that there is no defect in his memory; in particular, that he never has to re-read a work in order to recall it.

The magnitude of these figures alone shows that a problem exists, but they are confirmed by experience as well. Errors in diagnosis occur, unfortunately; we all know of such cases.

The doctor is thus faced with the problem of diagnosis, and every day he sees the difficulties of his work increasing. The logician is faced with the problem, as well, when he is searching for rules and conditions.

Considered in this light, the problem is a very general one which, as we shall see, concerns Medicine solely because it represents a particular aspect of the problem and is really that of many different forms of knowledge.

INFORMATION IN MEDICINE. NOTATION SYSTEM

It seems that Medicine is "par excellence" a field undergoing continual change, and it seems difficult to reduce it to logical terms. Now it is in no way a question of reducing the very substance of Medicine, clinical observation and other factors, but very much the opposite procedure of adapting logical symbols to this complexity.

Here, logic is not conceived as a more or less arbitrary order imposed on facts, but as a way of transcribing these facts so that by considering them as a whole, it becomes possible to bring out laws and relationships between general opinions which have hitherto been hidden.

We can give an arbitrary number - any number - to each pathological symptom, provided there is a 1:1 correspondence. To simplify the question, let us suppose that we choose the following series of natural whole numbers:

(a) 1 2 3 4 5 6 .....

Each of these figures then corresponds to a symptom, and once the conventional symbols have been decided upon, writing the figure 5 corresponds to writing, for example, "headaches in the vertex".

On the basis of this conventional system, we can write an ailment M of which a headache in the vertex is one of the symptoms:

(b) M = 3, 5, 8, 17 .....

Thus in this series in addition to a headache in the vertex we have other symbols, 3, 8, 17 ....

But we can write this ailment M in another way. If we agree to make the symbol "1" (i.e. the binary symbol 1) express existence so that, for example, if there is a "1" underneath the number "5", the symbol "5" is by definition present, we can write series *(a)* and mark with a "1" underneath, the symbols which really are present, for example, in the ailment M:

(a) 1 2 3 4 5 6 7 8 ..... 17 18 19 .....

(b) M = 1 1 1 1

We can also agree to indicate with an 0 the absence of a symptom: hence we get:

(a) 1 2 3 4 5 6 7 8 ..... 17 18 19 ....

(b) M = 0 0 1 0 1 0 0 1 ..... 1 0 0 ....

It is then possible to eliminate the *(a)* series (which can always be found again easily), the position of each 1 and 0 indicating the pathological symptom which they represent.

We then get

(c) M = 0 0 1 0 1 0 0 1 .... 1 0 0

With these conventions, which are very simple, if not childish, it will be easy for us to describe, with a code, all the ailments, in the form of relationships, such as *(c)*:

N = 0 0 0 0 1 1 0 0 1

P = 0 1 1 0 1 0 1 1 1

However, we can describe the patients themselves, as well:

Mr. K.... = 0 0 1 1 1 0 0 0 ....

Mr. H.... = 0 0 0 1 1 1 0 0 ....

For example a person in good health will be represented thus:

X = 0 0 0 0 0 0

Are we entitled to write in this way?

Are we entitled to reduce to logical symbols entities as complex as pathological symptoms? Are we entitled to put on one and the same footing symptoms which are obviously not all of the same value?

It seems that we are not - at first sight; and this is one of the most important arguments put forward.

Now if we study the question more closely, we find that there are various criteria of the value of pathological symptoms. For example, it is perfectly legitimate for one to give an extremely limited prognostic value to the appearance of a pre-auricular ganglion during an ocular neoplasm, or to attribute great diagnostic value to photophobia in Weeks' kerato-conjunctivitis. At the same time, however, it must be borne in mind that the appearance of the same pre-auricular ganglion in a case of Parinaud's conjunctivitis is perfectly normal; on the other hand, if revealed in syphilitic interstitial kerotitis, the same photophobia would be of no interest for the purpose of diagnosis.

Thus we see that the notion of the value to be attached to a symbol depends upon both the criteria envisaged and especially the clinical context of the symbols; and this context can only be arranged when the diagnosis has been made.

Now we have a strange inconsistency here: we would attribute a value to the pathological symptoms, knowing that this value depends upon the diagnosis, and we would use this value to make the diagnosis.

Such a procedure must be rejected, because it uses the unknown factor in proof: the assumption is not the symptoms expressed in terms of a diagnosis, but, of course, the symptoms alone.

For these reasons, therefore, it is legitimate to use the above system of notation.

Introduction of a symbol expressing the absence of information

In addition to the symbols 1 and 0 about which we have just spoken, mention must be made of the question mark '?'. This third symbol indicates the pathological symptoms about which we have no information at all.

If, for example, the number 6 is used by convention for the wave-recording system of the electro- encephalogram, and if we have not yet made the EEG examination of the patient *K*, we may write the following, if we already know of the existence or absence of the other symptoms:

(d) $K = 1\ 1\ 0\ 1\ 0\ ?\ 0\ 1$

Apart from the role of this question-mark, we shall see that its presence characterizes a logical structure which is peculiar to certain disciplines. The doubtful, non-informative nature of the ? enables us to do without studying the characters with logical validity further on.

Moreover, we shall see that this does not give rise to any logical operation, and that it therefore cannot lead to any erroneous conclusion.

Conditions for the validity of positive diagnosis

Let us suppose that a doctor has diagnosed the complaint *M* for a patient *A*.

Let us study in detail the mental operations and actions which lead him to this affirmation.

When the patient enters the doctor's consulting room, the doctor knows nothing about him, which we may express as follows:

(e) $A = ? \; ? \; ? \; ? \; ? \; \ldots\ldots$

The doctor then examines his patient. Let us agree to use this term "examine" to designate the whole of the questioning, the examination proper, palpation, percussion, auscultation ... in short, the techniques currently employed by practitioners.

When these actions are completed, if the doctor makes out a complete record, we find in it details of the facts which we write down using the values 1 and 0, according to a code of equivalence agreed on in advance:

The record for the patient will then be:

(f) $A = 1 \; 1 \; ? \; 0 \; 1 \; 0 \; 1 \; 0 \; 1 \; ? \; 0 \; 1 \ldots.$

- which is the same as saying that patient A shows the symptoms which we designate with the numbers 1, 2, 5, 7, 9, 12.... and that the patient does not have those which are designated by the numbers 4, 6, 8, 11.... the latter being those for which the doctor's search yielded a negative result. For example, he has looked for tenderness in the right iliac fossa, number 4: there is no tenderness there, so it is indicated by 0.

The question marks represent the symptoms which the doctor has not specified, either because he did not consider it worthwhile to do so, or because it has not occurred to him.

To simplify our reasoning, let us now suppose that the doctor is thoroughly acquainted with four ailments only. This is admittedly a surprising assumption, but it is perfectly valid. In fact the number of ailments is much greater; it could be put at 10,000, for example. But however great the number is, it is finite, which means that we can be sure of succeeding in drawing up a complete, exhaustive list of all the ailments. It is by regarding 4 and 10,000 as comparable in the sense that they are both finite, that assimilation becomes possible: the reasoning applied to 4 may by extension be applied to 10,000 and even more. The knowledge which the practitioner has about the 4 ailments (which we shall call *M*, *N*, *P* and *Q*) has been acquired from the books on Medicine, improved by the examination of patients and kept up to date by reading specialist reviews.

This knowledge is recorded in the doctor's memory. He knows, for example, that the ailment *M* includes a headache in the vertex, and that in the same ailment *M* there is no tenderness of the right iliac fossa.

Confining ourselves to twelve symptoms, so as to make the reasoning easier, using the above conventions, in combination with a code, we may write:

(g) $M = 0\ 1\ 1\ 0\ 1\ 0\ 0\ 1\ 1\ 1\ 1\ 0$

- a relationship which represents symbolically the knowledge which the doctor has of the ailment *M*.

We shall see that, if it is suitable, the arbitrary limit (12) which we have imposed on the number of symptoms retains all the demonstrative value required for the argument.

In fact, it is easy to confirm that, although there may be a large number of symptoms, they are not infinite in number. However great the number is, we shall see that it in no way invalidates the argument below.

We may write three more relationships on the lines of *(g)*.

(h) $N = 1\ 1\ 1\ 0\ 0\ 1\ 1\ 0\ 1\ 1\ 0\ 1$

(i) $P = 1\ 1\ 0\ 0\ 1\ 0\ 1\ 0\ 1\ 1\ 0\ 1$

(j) $Q = 1\ 1\ 1\ 0\ 1\ 0\ 1\ 1\ 0\ 0\ 0\ 0$

Let us take as an example the patient who goes into the doctor's consulting room. Some doctors, who are famous names in the short history of Medicine, had acute powers of observation which enabled them to see details normally overlooked. However, if we ignore these men - who after all, are exceptional - and consider the more common cases, then from the very beginning, after the first words spoken, after the first replies to his questions, the doctor will generally have some information - vague perhaps, but of such a kind as to enable him to direct his investigation along certain lines.

We shall dwell upon this first change in the doctor's attitude. At first, he is passive and contents himself with recording the data as he finds it, or which he may even obtain after a single question even though he poses it without any preconceived idea. During a second period, when a working hypothesis has already been formed, the doctor directs his questions and examinations along definite lines.

The length, difficulties and methods of the initial consultation period vary greatly. They depend upon two factors: on the one hand, the mental make-up of the doctor and on the other, the form of the ailment.

This period, during which the doctor makes a very random search for indications which may restrict the field of subsequent investigation, we call the "semiological period". It is during this period that what is known as the "clinical sense" seems to appear. If the patient comes in with his head down, covering his eyes with his hand, and wiping them (they are covered in a muco-purulent secretion) with a handkerchief, declaring that there is a pain in his eyes, just as if he had sand in them, then the semiological period is reduced to the time taken by a simple thought

association reflex. The doctor immediately thinks of infectious conjunctivitis. If the patient has pains all over his head, in no particular part, and a general feeling of fatigue or asthenia - but still has a good appetite - further information will be needed before an exact diagnosis can be made.

Let us suppose then, that the doctor has just completed the semiological period. He thinks to himself: This ailment could be *N*, *P* or *Q*.

What has made him think of these ailments?

This is one of the essential points of diagnosis.

These ailments have occurred to him, because the patient has shown the symptoms which he knows to be part of the ailments *N*, *P* and *Q*.

That is, he has associated in his mind the constituents of patient *A*'s ailment with the constituents of *N*, *P* and *Q*. We may write this mental process as follows: let us suppose that the knowledge which the doctor has of his patient at time $t = 2$ is such that:

(k) $A\ t\ 2 = 1\ 1\ ?\ 0\ ?\ ?\ ?\ ?\ ?\ ?\ 0\ ?$

That is, the patient has symptoms 1 and 2, but not those numbered 4 and 11. The same table applies to ailments *P* and *Q* and *N* as well.

As soon as the doctor has made a mental association between one or several ailments and the case of his patient, the search for symptoms is no longer carried out at random. On the contrary, the doctor makes a precise search for such and such a symptom which he knows to be part of the ailment concerned. His reasoning runs as follows: such a symptom belongs to this ailment, and my patient already shows these symptoms; let us see whether he has this one as well.

We can draw up a table showing these steps in the mental process:

No. of symptoms:	1	2	3	4	5	6	7	8	9	10	11	12
*At*o =	?	?	?	?	?	?	?	?	?	?	?	?
*At*1 =	?	1	?	?	?	?	?	?	?	?	?	?
M =	0	1	1	?	1	0	0	1	1	1	1	0
N =	1	1	1	0	0	1	1	0	1	1	0	1
P =	1	1	0	0	1	0	1	0	1	1	0	1
Q =	1	1	1	0	1	0	1	1	0	0	0	0
*At*2 =	1	1	?	0	?	?	?	?	?	?	0	?

*At*o = the patient enters the consulting room; he is about to cross the threshold; the doctor is aware of his presence, but he has not seen him yet and knows nothing about him.

*At*1 = symptom No.2. is immediately apparent. Nevertheless, since it occurs very frequently (in the present case, it occurs with *M*, *N*, *P* and *Q*, i.e. in all the ailments), the doctor cannot draw any valid conclusion from it. *At*1 thus represents the state of his knowledge in the semiological period.

$At2$ = the presence of symptoms 1 and 2. The absence of Nos. 4 and 11 connects this case with ailments N, P and Q. Ailment M must not be considered, because it includes symptoms No. 11, which the patient does not show. Comprising $At2$, N, P and Q, the doctor then looks for symptom No. 5 in the patient, for example. If it is present, the possibility of the ailment being P or Q must still be entertained; if it is absent, the doctor will have to consider ailment N.

This thought process - which utilizes the facts held in the memory, comparing them with the actual case of the patient, in order to direct the semiological investigation - we call the "differential diagnosis period". It is a period in which the doctor is guided by his knowledge of pathology.

I would emphasise that these different stages of diagnosis were never described by the classical writers. It seems that they were not aware of these mental activities. (This assumption is very probably true, since habituation to thought processes removes them from the conscious mind).

To revert to our example; let us suppose that patient A shows symptom No. 5, which we write as follows:

(l) $At3$ = 1 1 ? 0 1 ? ? ? ? ? 0 ?

This means that the patient is suffering from ailment P or Q.

In order to be able to make a positive diagnosis, the doctor then looks for symptom No. 12 in the same way. If it is present, the positive diagnosis is made, still by comparison: patient A is suffering from ailment P, because symptom No. 12 is present in this affection.

This procedure, and the various periods involved, is followed in its entirety in all consultations, and it is the basis of all diagnosis. Only the length of the various periods varies.

Certain points thus brought out should now be specified in detail. I shall ignore those conclusions which are of interest only to doctors, and dwell upon those which are of logical significance.

1. If we designate by $\Sigma\ (K)$ the number of symptoms defined by 1 or 0 in an ailment K or in the case-sheet of a patient, we may write:

(m) $\Sigma\ (At3) < \Sigma\ (N)$
(n) $\Sigma\ (At3) < \Sigma\ (M)$
(o) $\Sigma\ (At3) < \Sigma\ (P)$
(p) $\Sigma\ (At3) < \Sigma\ (Q)$

This means that, in actual fact, the doctor does not use all he knows about the ailment for making the diagnosis. The difference represents the symptoms which the doctor has omitted or neglected to specify, since the diagnosis appears to him to be decisively established without them.

2. This is the essential point which I emphasize; namely, that the diagnosis is the outcome of a series of comparisons between what the doctor knows about the ailments and what he knows about his patient. There is no fabrication at any time.

This conclusion is very important, and we shall see below its consequences and applications.

In all the above, we have reasoned about four ailments and eliminated three of them. It is easy to see that we could just as well have reasoned in the same way about five ailments, or six; in short, about any desired number of them, on the one condition that the number of symptoms characterizing the ailment increases at the same time. In passing, let us note that we may conceive the number of these symptoms being increased to thirteen, then to fourteen, and so on, in the same way.

In practice, the number of symptoms is much greater than the number of ailments, which enables the ailments to be distinguished unequivocally.

Conditions for valid diagnosis

With this ternary system of symbols, it will be easy for us to study the conditions to which the doctor must subject himself, if he is not to make an error in his diagnosis.

This does not mean that these conditions are essential for exact diagnosis, but that they are logically necessary. The diagnosis may, of course, be correct without all the conditions being fulfilled, but it may also be wrong.

If the police have a detailed description of an offender - if, for example, they know that he has a scar twelve centimetres long, pigmented, in the right lumbar region, and if they have some very good photographs of him as well - the offender may be arrested in the street, thanks to the photographs. There will not be sufficient identification, however, until it has been confirmed that he has the scar. Nevertheless, the diagnosis of the policeman who recognized him in the street will have been correct, even though it was not based on all the requisite data. On the other hand, it invalidates the arrests which others may have made, arrests which were not justified in the absence of the scar (wrong diagnosis).

Taking the above notations, we may write:

$$(q)\quad 2^{\Sigma(Kt)} > V$$

where Kt represents what the doctor knows about the patient at the time of the diagnosis, and V the total number of ailments.

This formula represents the minimum of knowledge which is necessary in order to ensure that no error in diagnosis is made.

It is also the mathematical expression of the two requirements illustrated by the following examples.

Let us suppose that a doctor has diagnosed the ailment M in patient A. He has studied, in all good faith, all the possibilities, all the hypotheses for symptoms, with which his memory has been able to furnish him. If, during this study, he had considered an ailment Q which perhaps appeared to him to be more closely identifiable with the case of his patient A, he would have ruled out the diagnosis M in favour of Q.

However - and this is an important point - for any given patient, the search amounts to going through all the ailments together, since we are left in ignorance, if the very ones we rule out do not include just the one from which our patient is suffering.

How is it possible to satisfy these conditions, when for the cornea alone, the number of ailments may be put at 1,000? Then again, as we have seen, and as happens in every diagnosis, the diagnosis is made with a relatively small number of symptoms.

In our example, we have put the number at 5. Here, it is easy to study the possibilities, because only four ailments are involved, M, N, P and Q, but when the number of ailments runs into thousands, it is difficult for the practitioner to answer this question. Would there not be a risk of the diagnosis which I have made with all these particulars about the patient, being invalidated, if I looked for such an such another morbid symptom?

We thus have a second condition. In order to be sure about his diagnosis, the doctor must look for all the symptoms.

Now some of them involve certain danger; how can they be attributed to the patient systematically.

Therefore to be sure about a diagnosis, it would be necessary both to consider all the possible affections and to look for all the pathological symptoms in the patient.

Formula (q) provides a solution for these requirements which is neat and simple, mathematically speaking.

It also defines a threshold, Δ, which we shall find in the further treatment of these applications below.

Value of the Diagnosis

Having thus specified the mental processes leading to the diagnosis, we should now study the value of the affirmation that the patient is suffering from such and such an ailment.

There is one point which must be brought out straight away: it is not a question of a probability here.

A patient would never allow himself to be left permanently in doubt. Of course, there are cases - and they are frequent - where a consultation cannot lead to a definite conclusion, even with searching examinations; cases in which further examinations are essential; and here the patient will readily allow the diagnosis to be deferred until the results are available. But once all the symptoms have been specified, a conclusion must be drawn, and it must be definitive. That is, the reply given must not be confined merely to probabilities; it must be absolutely positive. To give a patient who is anxiously awaiting a diagnosis a reply that "there is one chance in three that you have the complaint M, and two chances in three that it is N and finally one in ten that it is another complaint" invites the following kind of reply: "make the necessary

investigation to put an end to this ambiguity, then". If, for example, *M* is characterized by a shift to the left in Arneth's formula, let us let blood."

I would stress this very special nature of medical diagnosis.

For a long time it has been thought that statistical studies could be of no little value in Medicine. This is true; statistics are an incomparable tool of knowledge in medicine; but not in the field of diagnosis.

Statistics are on a level different from the individual level of the patient.

When the statistical method has been applied in classifying groups of patients and studying correlations, it cannot give any answer to the question everyone asks: namely, "which group do I come under"?

Statistics are the tool of the public health specialist, who studies man as part of a whole, but they cannot give any help in the highly individualized branch of diagnosis.

However, it is certain that for some cases, though statistics cannot give a final answer, they can be a guide. We shall study below these special data which in actual fact correspond to a particular material state.

Diagnosis and Prognosis; the Conditions for Prognosis

Besides diagnosis, and subsequent to it, we have the development of prognosis, which consists of forecasting the course of a given ailment. Here, statistics attain their full value, being based upon the analysis of existing and comparable facts.

Thus it makes use of the diagnosis results and interprets them, sometimes with the aid of new facts which have nothing do do with making the diagnosis.

For example, we have here a patient who has had an accident. After the general examination, questioning and supplementary examinations - radiography, in particular - the doctor diagnoses open spiral fracture of the two bones of the leg.

At this moment, the diagnosis is made unequivocally.

In order to make the prognosis however; to be able to say to the patient, "we are going to reduce this fracture, put it in plaster, and in so many days' time you will be able to return home", further investigations have to be made; if the blood sugar level is high in a diabetic the prognosis will be less optimistic.

Thus, as far as the elements of evaluation are concerned, the rules for prognosis are the same as those for diagnosis, but instead of leading simply to an affirmation, all the data are together combined through a factor derived from the study of earlier cases, and lead to the statement of a probability.

PRACTICAL APPLICATIONS OF THIS THEORETICAL STUDY IN LOGIC

Once the conditions for diagnosis and the means of arriving at it have been studied, the applications are easy to conceive.

The most striking and simplest illustration of this is the use of punched cards.

For each symptom we have a corresponding place on the card, determined by the row and column (on an IBM card there are 80 columns and 12 rows). There is a card for each ailment, with perforations for each of the symptoms present. In machine language, all the cards together form a "library".

We ourselves can read the contents of these cards, thanks to the code, as well; but a simple sorting machine handles them much more rapidly.

The cards are made out, in holes, according to the data contained in the medical treatises, old and new, none being left out; and they are punched according to the latest data given in the most recent works. That is, the cards constitute a complete, up-to-date library.

We have a patient before us and we find that he has certain symptoms, Nos. 78 and 115, say, while those numbered 513 and 587 are absent.

We then send the cards through the sorter, having all the cards without perforations 78 and 115, and those with perforations 513 and 587, rejected in four successive passes.

We are then left with a number (N') of cards. Let us pick out one of them at random; it has, say, perforation 432. We see whether the patient has this symptom, and find that he does in fact have it; so we then use the sorter to eliminate from the cards left after the first four passes all the cards without this perforation. Thus by successive passes - each one of which reduces the number of remaining cards, we are eventually left with only one card, representing the diagnosis.

If we analyze this procedure, we see that we have satisfied the theortical conditions which we have shown to be desirable in the logical study, i.e. we have allowed for all the possible cases; in fact, all our cards together represent all the existing medical knowledge on the subject concerned. We are sure that looking for a further symptom cannot invalidate our diagnosis, and in fact, we have already eliminated them all but one. It is easy to make a check, by seeing whether the patient has or has not the symptoms entered as present or absent on the card.

Lastly, this mechanical system enables us to calculate by practical means the value of Δ the threshold - which we defined above. Thus this threshold corresponds in practice to the number of sorting passes needed to obtain one card only.

I have given this example of the operation of a punched card sorter, because it enables the processing of the information to be followed easily. However, this example is surpassed in the field of applications by machines using magnetic tape and quick-access memory drums. Combination of the two

memory systems allows more flexible and rapid operation, which I cannot describe in detail here.

Before I conclude this passage on the practical systems, I would like to stress two points - which I consider essential - in the use of machines: on the one hand, the subordinate role of the machine, which cannot be thought of as making the diagnosis by itself; the machine assists the doctor, who remains the prime mover in Medicine, - without whom the machine could not function; on the other hand, the quality of the service given by the machine: its memory is tireless and infallible.

Study of the General structure of the Machine

From the particulars given above, one can easily derive notions enabling one to state precisely the structure of medical knowledge.

We shall only give a brief glimpse of this structure, since a complete study would require technical treatment falling outside the essential point of this paper.

If we consider equivalences of the *(g)*, *(h)*, *(i)* and *(j)* type:

(g) $M = 0\ 1\ 1\ 0\ 1\ 0\ 0\ 1\ 1\ 1\ 1\ 0,$

we find that they are presented in the books and publications in the form sequence of M followed by the symptoms which it covers.

In fact, the name of M, of N etc. is the chapter heading under which the details which we denote by 1 or 0 are listed. However, we see that conversely, in making the diagnosis, the movement is in the opposite direction, from the symptoms to the name which is the diagnosis.

To denote this double use of the *(g)*-type equivalence, we can use a double arrow $\rightleftharpoons$ thus:

(r) $M \rightleftharpoons 0\ 1\ 1\ 0\ 1\ 0\ 0\ 1\ 1\ 1\ 1\ 0$

This is important, in my opinion, because it emphasizes symbolically the changes brought about in our mode of learning by progress in technology.

In fact, if we consider the development of information processing techniques, we see that for thousands of years the problem of storing information has been solved fairly successfully - Nebuchadnezzar had libraries even in his time; and printing has multiplied this means of storage, but so far there have been very few means of converting information and, above all, they have been ineffective. We have only had the information, in a form like that of the *(r)* formula, but following only one direction in use: from M to the constituent elements of M.

The only information-processing machine which we had at our disposal was our memory and intellect: a very flexible machine, with infinite resources (many of which escape us), it is true, but one with a serious defect in that its storage capacity is inadequate and it is not absolutely reliable.

Modern techniques have given us punched card machines, magnetic tape memory machines, magnetic drum memory machines .... in short, a set of units capable of processing information, of operating in both directions, as in *(r)*.

This notion of two-way utilization clearly appears in the study of the application, but the principle of it is used in the mental process. I think that it is important to emphasize that the operations which we can follow easily in their material form on punched cards exist, except for a few details imprinted on the mind when a diagnosis is made.

The only reason why they do not become apparent is that we are too much accustomed to carrying them out and they therefore remain in the subconscious.

We only have to see what happens in a difficult case which, owing to the difficulty, demands all the doctor's attention. He reasons the matter out as follows: This is not the ailment *P*, because there would have to be different pigmentations; nor is it *R*, which is characterized by a rise in the maximum humeral pressure. In short, in any unusual type of case, the doctor has consciously to think over the stages of diagnosis, and then these successive comparison are made with the eliminations, finally leading to identification.

Moreover, it is very probable that the practitioner has short cuts (still unconscious), which quickly lead him to the diagnosis, and that organizational structure of the data in his mind is less rigid than the equivalences of the *(r)* type.

However that may be, and with the reservations mentioned, we see that he remembers information in the *(r)* form.

Medicine does not consist only of diagnosis and prognosis, however; it includes - and this is the most important point - therapeutics, as well.

Now the latter is linked to the name of the ailment in a more complex way.

We shall not study the logical details of therapeutics, because they are questions involving medical explanations which would make this paper too long if included here.

I shall merely state that, in the decision "you must have such and such a treatment", one can see the same logical basis as in diagnosis - with the reservation that here, as in prognosis, probabilities must be allowed for.

If we try to tabulate these various data for each ailment we get a group of relationships in the following form:

(s) ailment $M \rightleftharpoons 0\ 1\ 1\ 1\ 0$ .... symptoms *(Z)*

ailment $M \rightleftharpoons$ treatment

Prognosis for ailment $M \rightleftharpoons 0\ 1\ 1\ 0$ ..... symptoms (not necessarily the same as in *Z*)

This set of relationships forms a large part of the sections on the ailment *M*.

During the consultation, the doctor's mental processes follow a path leading from the symptoms to the diagnosis whence they spread out towards the treatment and prognosis.

If we try to generalize this structure and derive a logical system from it, we find that the basic element is the diagnosis. It is the diagnosis which forms the focus of all the doctor's efforts; from it, he draws his conclusions for treatment and prognosis. The term "diagnosis" is the "turntable" for the dynamic logical structure of medicine.

It is interesting to note how general this structure is. I believe it can be found - more or less in its entirety - in all the disciplines.

In fact, Medicine is a science combining knowledge and action: action in the prescription of treatment, and surgical action. The knowledge classified under a term which defines it very often only includes the element of acquisition of information.

If we consider botany, for example, we can easily see that this science consists essentially of the logical part which corresponds to diagnosis. The botanist observes a plant, picks out its particular features, compares them with the classification tables in his memory, and when he has identified the plant under observation with a plant already known, he is able to name it.

Other forms of activity include both diagnosis and action, however. For example, if we consider the logical work done by a lawyer preparing his speech as counsel, we find that it includes (a) a diagnostic element: acquiring information on his client's case, followed by comparison of these particulars with the precedents set down in legal texts (the diagnosis may be said to be made when the lawyer can say "Here is the precedent which applies to my client's case"; (b) an action element - strangely distorted here, but which in actual fact would have to be confined to reading the text of the precedent which applied to the client's case.

The same reasoning is followed, moreover, by the judge in charge of the case.

The very same structures are to be found in administration. All the administrative regulations are founded on one and the same logical model: the first part defines the groups, and the second, the measures applying to these groups.

It is absolutely the same structure as that in legal texts. In some cases, a group is so defined that it is impossible to be mistaken - the problem does not exist - for example "no-one may plead ignorance of the law". In other cases - the most numerous - the definition - of the group is complex: and besides this, the definitions themselves are numerous. Classification thus involves the same problems as those involved in making a diagnosis.

For example, "when the injury heals without permanent disablement, or if there is permanent disablement at the time of healing, a medical certificate showing ..... shall be made out in duplicate" (Act of October 30th 1946).

In this passage, which states the law for accidents at work, the diagnostic element may be distinguished. Does this case, this victim of an accident have an injury which has healed without disablement, or not? If so, he belongs to the group defined by the act, and from this, the action element - comparable with the therapeutic element in Medicine - is derived: a certificate is made out in duplicate.

In most cases, however, the problem in Law is not such a simple one. Indeed, whereas in Medicine there is only one diagnosis, in Law there may be several answers.

In Medicine, a patient may for example be suffering from sciatica and a gastric ulcer. We shall make two easy diagnoses - separately, because in this case the indications and symptoms of the two ailments are distinct from one another. But if we examine a patient suffering from cardiac insufficiency with a chronic emphysema complication, it will be difficult for us to distinguish between the symptoms of the two ailments.

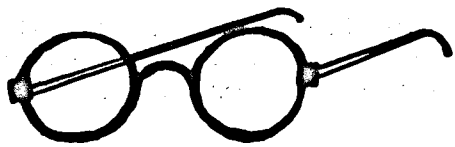
In Law, such intricacies are common, and it is these that the reasons adduced in the judgements specify. Each reason involves a separate "diagnosis". Later on, we shall study very briefly the consequences of this.

In passing, let us stress here the existence of a threshold - comparable with the one we defined for Medicine; a threshold which, derived from formula *(q)*, indicates the presence of necessary and sufficient conditions, as I have defined them for making a diagnosis.

Although this structure of knowledge may seem over-simple, it must be borne in mind that this elementary simplicity is the rule in our mental processes.

I would like to demonstrate how general the diagnostic process is, by means of a short example.

This mode of thought is in fact so general, so common, that we are not consciously aware of it. When we look at the picture below, the name of the object occurs to us immediately; it is a pair of spectacles.



This recognition - which is really only a diagnosis - must be studied in detail, however. To do this, let us complicate the data of the problem; the difficulties which we shall find will bring out the mental steps taken. Let us say that an object - unknown to us - is placed underneath a cloth by another person.

If we try to guess what the object is, without raising the cloth, we may be surprised when we reason as follows: It is a small object, hard to the touch, in the form of a rectangular parallelepiped, with its longest side about three centimetres long. But it is not a rubber because it is not flexible, it does not bend; on the contrary, it is crumbly, like sugar; yes, it is a lump of sugar.

Here, we can easily recognize the steps which we have already described for making a diagnosis.

Thus we see that this logical form of thought is very general; we find it not only in most of the disciplines but in the processes of everyday life, as well. The process is very rapid in its usual form, so rapid and so common, in fact, that we are not consciously aware of it.

For concluding our study, let us take the following general logical formula.

$$(r)\ M \rightleftharpoons 0\ 1\ 1\ 0\ 1\ 0\ 0\ 1\ 1\ 1\ 1\ 0$$

with two-way utilization, and the deductive type of conclusions attaching to the notion of *M*.

It is useful to study, on the basis of these forms of knowledge, the logical validity of various structural relationships found in disciplines of this kind - disciplines and unorganized forms of knowledge as well. Critical study of the validity of relationships of type

$$(r)\ M \rightleftharpoons 0\ 1\ 1\ 0\ 1\ 0\ 0\ 1\ 1\ 1\ 1\ 0$$

There are a very large number of relationships of this kind in all the disciplines; they are all valid by convention. In most cases, in fact, *M* corresponds to a fact or to a notion of the "entity" type.

In geology, for example, it may be said that oolithic limestone has such and such distinguishing features. Conversely, such and such characteristics found in a rock enable one to state that the rock is oolithic limestone.

This is in the case where *M* corresponds to a fact, an object. Another example, this time of an entity: the notion of "stress" corresponds to a set of defensive reactions in the body against some cause tending to disturb its equilibrium.

We see that these relationships are conventional relationships; they are postulates, convenient postulates, but postulates which cannot be challenged. In fact, their nature is not such that they rule out the co-existence of other conceptions.

We find these types of conventions, this type of structure in the descriptive sciences, such as botany and natural history.

They represent the initial, early form of a discipline, the latter in fact requiring efforts of comparison, observation and description, and today sciences which are confined to these functions are rare. We are turning more and more to action.

This is written for Medicine as follows:

(t) ailment $M \Longrightarrow$ treatment T

This relationship is valid, generally speaking, for all the sciences in which an action is an extension of knowledge:

(t') $M \Longrightarrow A$

Critical study of the validity of relationships of the type

$$M \Longrightarrow A$$

This is the essential point of the present study; the action consequent on a piece of knowledge is in fact all the more effective, the more specific the knowledge is.

Expressed in such general terms, such a law seems quite meaningless. One must consider specific examples, in order to be able to judge its value better.

The therapeutic action may vary very considerably. In France there are some 9,000 drugs, without counting surgical intervention. This means that if taken at random, a prescription will have one chance in 9,000 of being effective. The difference between this 1/9,000 chance and the cure effected through the doctor's diagnosis emphasizes the importance of precise knowledge.

Relationships of the *(r)* type are only modified slightly or at least gradually - because they represent basic concepts, "generally recognised ideas", frameworks for nosological classification. These ideas, which may only be contested by other ideas, have a reliable reference value. They are the basis of mutual understanding.

The *(t)*-type relationships, on the other hand, are always being questioned owing to technical progress. In Medicine, for example, they are overthrown by the appearance of a new anti-biotic.

It is the *(r)*-type relationships which determine the degree of complexity of a science. When the number of the values 0 and 1 corresponding to each relationship increases, the discipline becomes more complicated: after a certain point, it divides, and this is what has happened in Medicine.

It is the *(r)*-type relationships which indirectly underlie the notion of a threshold - from a diagnosis point of view, as well, because one must not lose sight of the fact that the threshold varies with V, the total number of ailments.

Certain objections must be considered. The principle of a fixed number of diagnostic hypotheses surprises one because we cannot readily conceive the

limits to our intellectual capital - which are hazy, even more so than those of books on pathology; and it is conceivable that putting pathology into the material form of *(r)*-relationships, then, for example, punched cards, will at first seem an arbitrary limitation.

Now the existence of a finite number of symptoms - however large the number may be - is the medical expression of determinism, a determinism, which is absolute, both for living bodies and for inorganic matter (cf. BERNARD).

The basic postulate of science is that "in Nature there are no contingencies, no capricious occurrences, no miracles, no free-will" (GOBLOT).

This postulate is the profession of faith of all science, and if it sometimes seems to us to fall down in the face of unexpected facts, it invites us to admit that our knowledge is still imperfect.

Determinism is not obvious in Medicine, because it expresses itself in a complex way, involving above all a very large number of factors; consequently we do not readily attribute to Medicine the same logical forms of knowledge as we do to the other sciences. In Medicine, $\Sigma\ (M)$, the quantity of knowledge is exceptionally large, as is V, the number of ailments.

In sciences other than Medicine, the value chosen for $\Sigma\ (M)$ is lower; but in Medicine this quantity is forced on us by the very complexity of the object of our study: Man. In the patient we have a complex whole, the different elements of which need to be specified separately, but which must be regarded as indivisible, as far as interpretation of it is concerned.

The interest of this study, of the ternary notation (0, 1, ?), lies both in the fact that the symbols describe the mental steps taken in the diagnosis, in the fact that they specify certain notions relating to the necessary and sufficient conditions for diagnosis, in the fact that they show the structural complexity of Medicine, but above all, in the fact that they make it possible to use machines.

The machines give us their power and above all, their reliability, their reliability in memorizing data which - as we find every day - is sadly lacking in the human mind.

Consequently, we can be certain that we cannot foresee the developments which this mechanization will bring to the problems very frequently encountered in diagnosis.

In conclusion, we may make the following assertions (and I think that these are the essential points of the present study):

(1) In an action as complex 'a priori' as diagnosis, embodying and utilizing all the facts of medical knowledge, and apparently an art in itself, it is nevertheless possible to describe a logical process.

(2) The principle of this logical process is amazingly simple, because it comprises a series of comparisons between what the doctor knows about ailments and what he knows about his patient. By successive eliminations, following comparisons which reveal differences, the diagnosis is made

when the case of the patient is found to be identical with that of the ailment M; we then say that the patient is suffering from ailment M.
(3) The simplicity of the process enables it to be mechanized without difficulty.
(4) However this simplicity must not be allowed to hide the difficulties already stressed by CARREL: "The very volume of the facts which we know about Man is itself an obstacle to their use."
(5) Mechanization can solve the difficulties arising from the inadequacy of our memory, and is the sole means of utilizing the whole body of knowledge.
(6) Diagnosis is a very general process, and strangely enough, it very frequently takes place in our sub-conscious, which explains why it is often uncontrolled and therefore a source of possible error.

DISCUSSION ON THE PAPER BY DR. F. PAYCHA

DR. F. A. NASH: I enjoyed reading Dr. Paycha's paper. I agree with many of the points he makes in his clear exposition of his views on the nature of the diagnostic process.

Whatever is asserted, especially in print, tends to be taken as fact unless it is questioned. I notice it is claimed (in the forenote to Dr. Paycha's paper) that he "made the first medical memory in 1953 with the help of Bull...". I should be foolish, perhaps, to try to establish with Dr. Paycha as to which of us has the greater right to unpopularity with our colleagues for inventing the first medical memory. However, it must be recorded that, in 1953, I constructed an apparatus to assist the logical faculties in differential diagnosis called "The Grouped Symbol Associator" (G.S.A.), and the patent applications were lodged officially with the Patent Office in London on 14th October. 1953 *(ref. 1)*. It is fully described in The Lancet *(ref. 2)* and the Mark III Model is commercially available.

Perhaps we can avoid any argument by distinguishing apparatus that remembers *by serial* operations from apparatus like the G.S.A. that not only "remembers" but associates what it remembers, and that instantaneously.

The G.S.A. makes visible not only the end results of differential diagnostic classificatory thinking, it displays the skeleton of the *whole* process as a *simultaneous* panorama of spectral patterns that coincide with varying degrees of completeness. It makes a map or pattern of the problem composed for each diagnostic occasion, and acts as a physical jig to guide the thought processes. *Figure 1* is a general view of the G.S.A.

I agree with Dr. Paycha that the diagnosis will always remain the decision of the doctor: but the machines can help with their infallible memories. I do not agree that the statistical approach is useless in differential diagnosis. It can help to reduce errors resulting from over-frequent diagnosis of rarities if one knows the frequency of occurrence of different diseases. Again if, as White and Geschickter state, 98% of death and disability in U.S.A. is caused by only 200 of the total of

REFERENCES

1. British Patent No. 28388/33.
2. NASH, F. A., The Grouped Symbol Associator. *The Lancet*, 1954, April 24, 874.

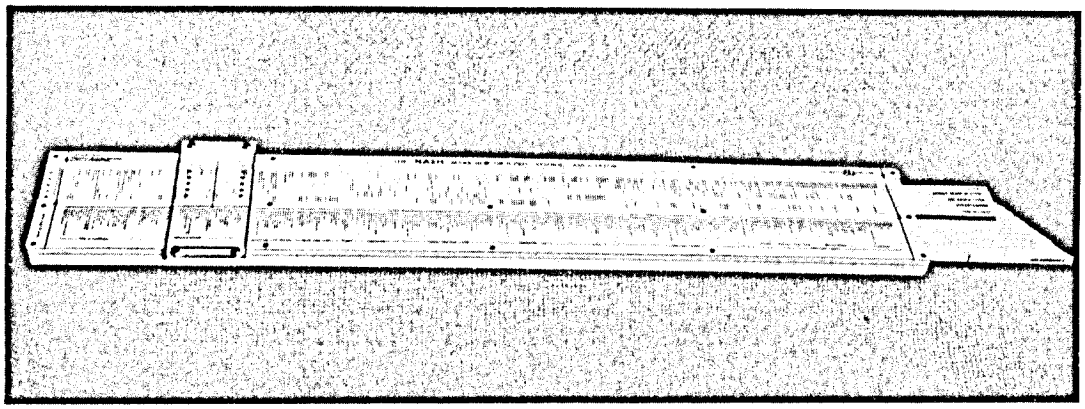


Figure 1.

2,000 diseases, obviously the doctor who knows which these are will, in the long run, make a better over all performance, other things being equal, than the one who does not.

In Cecil's Medicine, a standard American text book of medicine, there are described something in the order of 800 diseases.

DR. GREY WALTER: I should like first to congratulate Dr. Paycha on his presentation of a bold attack on a very controversial position. There are some general aspects of this work that I should like to bring into the discussion, as a physiologist, and not as a medical man.

It seems to me that we scientific workers in the medical field have rather neglected the way in which physicians go about their business; they are one of the few classes of people who are forced to appreciate complex patterns in human beings and build up from them notions of syndromes. This is rather a peculiar intellectual exercise and is very heavily weighted in the case of medicine with success or failure, since it deals directly with human lives. As scientists in the laboratory, we are reluctant to think in this way because the traditional statistical methods on which we rely so much do not help us to recognise complex individual patterns; they tend to efface individual differences rather than to emphasise them. But there are now available statistical methods which would help us to recognise syndromes in the more general sense, including those configurations of signs in normal people such as we generally call personal character or type.

These methods, whether clinical or statistical, are based, presumably, on the recognition of diagnostic signs, and I should like to ask a general

question of those skilled in this art whether the use of the adjective "diagnostic" is the same in medicine as it is, for example, in zoology, where I seem to remember that a diagnostic sign is found always and only in association with a certain genus or species. In medicine it is rather rare, I think, for a diagnostic sign to be so explicitly defined, but the application of the sort of methods Dr. Paycha has suggested might help us to recognise truly diagnostic signs. I think I am right in saying that Dr. Paycha's method was first applied to ophthalmology; that is, a group of organ diseases rather than organism diseases. Here the situation is considerably simpler than, for example, in neurology, where the problems are essentially of organism disease. Disturbance of the nervous system tends to affect many organs and hence the behaviour of the whole organism. This may raise rather special problems. I am not sure of this - it is a question I want to put to the meeting - as to whether methods which I suppose one can call cybernetic may be specifically adapted to the recognition of syndromes in organism behaviour. Consideration of an organism as opposed to an organ introduces special difficulties - not only difficulties of the same type but of a higher order than those obviously encountered in the application of this method to the diagnosis of *organ* disease. In neurology, one has an overlap and interaction of functions so that a similar type of functional disorder may result from a large number of central disturbances, because of the compensatory action of the nervous system. This seems to me where the word cybernetics may be justified, because one is bound to consider interaction between several systems of control. I would query whether the term cybernetic in Dr. Paycha's title is justified in this particular application. It might be justified in studying a system such as the nervous system, in which the interrelations of the components are as rich as possible, both between nominal inputs and nominal outputs and within the system itself. There the truly cybernetic methods might be essential in order to identify diagnostic signs and syndromes even including normal variations.

PROF. J. Z. YOUNG, CHAIRMAN: Are you suggesting from the general point of view that there are different systems of pattern recognition necessary, or different forms of classification? For example, you mentioned zoology or diagnosis. I was not clear what your general point was.

DR. GREY WALTER: There are two words that seem to be used rather freely - statistical and diagnostic. Dr. Paycha said that, for example, statistics interested the Medical Officer of Health but not the doctor. This statement can be true or false, according to how you define statistics; obviously statistics interest the doctor in the sense that his probabilistic judgment of a diagnosis is a statistical statement. There are tables of vital statistics drawn up by registrars that may not interest him vitally, but

they obviously influence his behaviour, for example, if he knows an epidemic is going on. It seems to me that "statistical" is used rather loosely in this sense and "diagnostic" is also used loosely. I know many of my own deeply respected medical colleagues will speak of a diagnosis or diagnostic sign in a sense that seems to me a good deal looser than you or I might approve in relation to the identification of a species or genus; should the usage be tightened up in application to medical diagnosis, or should we allow them freedom and not bother to reprove them if they use it loosely? It seems to be a useful word to have - it has a clear etymology and a strict usage in the basic sciences. It seems a pity that it should be used more loosely in medicine, because it confuses scientific workers who are trying to relate objective measurements of some kind to diagnostic signs and symptoms.

MR. G. B. NEWMAN: I consider that statistical considerations are an essential part of the diagnostic process, for much will depend on the relative detail of the diagnosis. Certain cancers are susceptible to their hormonal environment and before treating some patients it is necessary to know if the patient's cancer is of the "hormone dependent" type. Whilst it may be possible to say from observed signs that a patient obviously has Cancer of the Breast, it will not be possible to say with certainty that her cancer is "hormone dependent". Whether the patient is given treatment directed at the hormone dependent type of tumor must, therefore, be based on the a priori probabilities of "hormone dependence" in the type of patient concerned.

As a practical point, the patients case histories are far from being as complete as Dr. Paycha states and this, while presenting the great problem in the investigation of this subject, will also mean the more frequent use of probability considerations.

DR. A. REMOND: I have admired Dr. Paycha's work in France for several years and have often wondered how to apply it to my own practice. I understood then that it was a method for the future. I was not able, in the way I had been educated, to make use of it. In the neurological diagnosis at least, we are dealing with symptoms which are never entirely present or absent and which cannot be represented by either zero or one - they are always in between these extremes and we are seldom sure that they are present or not. They can be there once and when we look for them again they have disappeared or they are only half there. Take for instance, the Babinski symptom.

REFERENCE (by Mr. Newman)

BROADMAN, K., ERDMAN, A. J., LORGE, I., WOLFF, H. G., Cornell Medical Index, Health Questionnaire, *Journal American Medical Association* 1949, **140(6)**, 530 and 1951, **145** (3), 152.

Every morning, in a neurological department each patient is examined. At first when the interne looks for evidence of a pyramidal syndrome he may find it, it is "obviously" there; then, when later in the morning the "patron" comes to review the new cases, the interne says "This patient has a paralysis of such a kind, he has a babinski on the left". The "patron" looks for it ... it is'nt there any more. It is not the place here to discuss why it is so. In many instances we should rather deal with probability, instead of pure, clear cut observation. To come back to Dr. Paycha's work and before trying to make a medical diagnosis, a process often very difficult, we should dissect what we now call symptoms in more minute elements taken as bits which surely are there or not. But no clear definitions of these elements have been given as yet in medical books. Symptoms are too often complicated entities. The difficult question for me would be to redescribe every disease in terms of suitable code words being part of the language of a diagnostic machine.

DR. R. EFRON: As I understand the speaker, he is making a distinction between "prognosis", which he feels is a statistical element, and "diagnosis", which he feels is a more specific and individual act of decision. Speaking as a clinician, as a neurologist, I feel that this distinction is rather artificial. I think we make a probabilistic diagnosis because we are constantly checking back against the unfolding course of the disease. There is a distinction, therefore, between the speaker's concept of "diagnosis" that is set in time - at a particular instant, and the manner in which diagnoses are usually made. They are, in practice, more fluid things, taking place over a period of time and therefore there is always a feedback into the system of new information derived from observation of the patient's course.

There is one other point about which I am confused, and this may be because the teaching of medicine in Anglo-Saxon or English speaking countries is different from the way it is taught in France. We distinguish "signs" from "symptoms". A sign is something which is an observed phenomenon - observed by someone other than the patient. A symptom is something subjective - something which is complained of by the patient. In relation to Dr. Grey Walter's point of diagnostic signs, may I recall that one of the games which medical students often play is to think of so-called "pathognomonic signs". This is a special category of sign, only one of which permits the absolute diagnosis of a specific disease. An example: red teeth permit you to make a specific diagnosis of a certain metabolic disease. These are such striking signs that a medical student remembers them easily. Certainly, if there were more pathognomonic signs the probability of making a machine for medical diagnosis would be much higher.

DR. D. M. MACKAY: The argument that 10 signs are sufficient if you have 2^{10} possibilities from which to select is valid only if the signs are logically and statistically independent. Is it not essential to do some factorial analysis on the two hundred signs mentioned to find out what groups of signs are logically independent, before one can argue that ten or even many more could be sufficient?

DR. F. PAYCHA (in reply): I thank Dr. Nash for his remarks. But the part played by statistics in diagnosis must be bounded.

For a doctor, the use of statistics in diagnosis is, in fact, a solution of facility. If about 200 diseases can cause 90 per cent of mortality, I find it is quite immoral to neglect the remaining 10 per cent.

It should be very easy for a practitioner to limit his diagnosis to 200 diseases. But, when a patient enters his consultation room, he may be afflicted with any disease, even one of this 10 per cent. And the doctor does not know if this malady belongs or does not belong to this 200 diseases which cause 90 per cent of death.

If the doctor assumes that there is more probability for a certain disease to have occurred, then he is making an a priori hypothesis on his patient; and so, he eliminates them without any reasons.

Statistics are of interest to the Minister of Health, but they do not interest the practitioner. Statistics only have a part to play in the prognostic and therapeutic side. Statistics, of course, apply to a group of individuals, but the patient is a unique, a sole case; and the question is then to know to which group of the statistics this patient belongs.

(To MR. G. B. NEWMAN): The first consideration is very interesting in two points:

1. Because it contains in itself its answer: Mr. G. B. Newman says that statistical considerations are essential part of the diagnosis; and immediately, he takes an example, and he speaks of treatment. So, he demonstrates the part of statistics in therapeutics, but not in diagnosis.
2. Because such error is frequent; often one blends diagnosis and therapeutics.

This example is therefore instructive: if now the treatment of certain cancers is based on probabilities of hormonal dependence, it is because we do not know, before testing the action of the hormones, how to recognize hormone-dependent type of cancer. When we know a sign or a symptom to distinguish this type from the other, we shall be able to give more chance of life to the patients.

The second remark is very true: seldom, the patients' case histories are complete. But, in this eventuality, the interest of the Medical Memory is

that it gives more than one answer, and the comparison between these answers shows the missing signs and symptoms.

(To DR. EFRON): It is true that, for one patient, during a malady, the doctor may give several diagnoses. As Dr. Efron says, in practice, there are fluid things, taking place over a period of time, and therefore there is always a feedback into the system of new informations derived from observation of the patient's course. But, every diagnosis is done by a series of comparisons.

As regards the difference between signs and symptoms, French semiology is not very clear on this point, and that, of course, I am sorry about.

It is true that every medical student knows the sets of pathognomonic signs, special category of signs, only one of which permits the absolute diagnosis of a specific disease. But the practitioner may think that every disease may exist without these pathognomonic signs.

(To DR. MACKAY): There are two points of view in this contribution:

1. Logically 10 signs are sufficient, for there are about 1,000 diseases of the cornea, and $2^{10} > 1,000$.
2. Statistically signs appear not independent: certain groups of signs and symptoms appear together more frequently than alone or than other. In these conditions, it should be possible to study statistically, and by experiments, and also by factorial analysis every combination of the different signs. But, we do not forget there are about 200 signs: so there are 2^{200} combinations and this study should be tedious enough.

So, it is easier to try, with punched cards for instance how many sortings are necessary to obtain one card only: the number of sortings is about 10 in this case.

SESSION 4A

PAPER 5

MODELS AND THE LOCALISATION OF FUNCTION IN THE CENTRAL NERVOUS SYSTEM

---

by

R. L. GREGORY

BIOGRAPHICAL NOTE

Mr. R. L. Gregory, born 1923, was educated at King Alfred's School, Hampstead, London, then served in the R.A.F. (Signals) for five years during the war. He went to Downing College, Cambridge, read Philosophy for two years, and took Part II in Experimental Psychology in one year, under Professor Sir Frederic Bartlett.

He joined the Medical Research Council's Applied Psychology Unit, Cambridge, in 1950, his work including the relation between blink rate and both visual and non-visual attention (partly with E. C. Poulton). He was seconded to the Navy for six months, to work on psychological problems of submarine escape.

In 1953 he took up a Demonstratorship in the Department of Experimental Psychology, Cambridge, under Professor O. L. Zangwill. Apart from teaching, he has written papers on explanation models, colour vision, dark adaptation, hearing and deafness, and on special devices and instruments. His main experimental work has been concerned with trying to estimate neural noise levels in vision and hearing, and to relate these with human ageing. He produced, with Violet Cane, a theory of sensory discrimination in terms of signal/noise ratio in the nervous system, and was senior author of a thesis entitled "Increase in 'Neurological Noise' as a factor in Ageing" which received a major C.I.B.A. Foundation award for work on ageing (1956).

He was awarded Craik Prize for work in Physiological Psychology by St. John's College, 1958, and was appointed University Lecturer in Experimental Psychology, 1958.

MODELS AND THE LOCALISATION OF FUNCTION IN THE CENTRAL NERVOUS SYSTEM

by

R. L. GREGORY

SUMMARY

THE general question is raised: what sort of explanations are appropriate for biology? Teleology is rejected, while explanation in terms of conceptual models of the engineering kind is accepted. It is argued, however, that the engineering approach does not rid us of the necessity for making decisions on the purpose of observed structures or behaviour. Engineers generally know the design ends of their devices, but we have to guess the functional significance of biological systems. Certain difficulties are discussed. These considerations lead to the problem of distinguishing between 'accidental' and 'functionally important' features of biological systems. The criteria for distinguishing between these would seem to require conceptual models.

Localization of function is discussed in general terms. The use of the words 'localization' and 'function' is discussed, and a comparison is drawn between removing or stimulating parts of machines and brains. It is argued that the performance changes associated with removing parts of machines cannot be understood except in terms of functional models of the machine. The same would seem true for the central nervous system. In the case of independent parallel systems, such as telephone installations, much might be learned given only the crudest 'anatomical' model, and this is largely true of the peripheral nervous system. In the case of the cortex, this approach is likely to be misleading, especially if it is a tightly coupled system, for then removal of part of the system will either have little effect, except in certain conditions, or will introduce new functional features, for we now have a different system. These new features can only be understood, and only have significance, in terms of a functional model. It is concluded that ablation and stimulation of the brain may, and indeed does, bring out interesting facts, but that these must be interpreted in terms of models, for without a model we cannot say what is localised.

WHEN a biologist considers the fundamental question: "What *is* an explanation of behaviour?" he may have a sense of conflict, even of paradox. Biology seems to be a science in its own right, or set of sciences having common aims, and so it should have its own language and explanatory concepts; yet when any specifically biological concept is suggested and used as an explanatory concept it seems to be unsatisfactory and even mystical. There are many biological concepts of this kind: Purpose, Drive, *elan vital*, Entelechy, Gestalten.* Physicists and engineers seem, on the other hand, to have clearly defined concepts having great power within biology. Why should this be? Is it that biology is not sufficiently advanced as an explanatory science to have developed its conceptual systems sufficiently far? Or is it that biologists should not look for *special* concepts, uniquely applicable to living systems? This latter view implies that biologists should think of living systems as being examples of physical or engineering systems in which case engineering language should be applicable for description and explanation in biology.

If an engineer is presented with a piece of equipment about which he knows very little, he will at least know that it was designed by men for some purpose he could comprehend. He is unlikely to find any quite new, (or any highly inefficient), techniques involved in its design or construction. The 'black box' situation is artificial in engineering: the boxes an engineer sees are not as black as all that. In the case where the box *is* sealed, (perhaps as a game, or a test or in war) he still has a lot of information about it if he only knows that it was made for human beings by human beings. Biological systems are not designed by humans, and the purpose of many of their characteristics may be highly obscure. Thus in important respects the engineer is not on his home ground when he advises the biologist.

Biologists are shy of the notion of purpose; in particular, they reject teleological explanations as 'unscientific'. It is not clear, however, that classification of biological structure and function in terms of purpose should, or can, be avoided. At first sight engineering analogies may seem to do this, but this is not so. To take an example of Dr. Uttley's, it is a discovery, and a most important one, that the heart is a pump. It would

---

* Some biological concepts have been given 'engineering' definitions; for example drive and purpose. Whether they still cover the cases required by biologists is an important question. Perhaps Survival of the Fittest is a specifically biological idea, and in a sense an explanatory concept, though it does not attempt to explain the functioning of an organism but only how it came to be 'designed'. Although the idea started as a biological theory, it may be extended to other cases, in economics for example, and it may be expressed in non-biological, largely statistical, language. Thus it is not specifically biological.

be impossible to understand it without recognising this. Biological systems are adapted, through genetic experience, to serve ends which we might, or might not, discern. We have to discover what this end is before we can think up what sort of system it is, or assess its efficiency.

At this point we should try to clear up an old bogey. We can see what the purpose of a piston or a boiler is, once we understand steam engines, and we can see the functional purpose of the heart, or the eye. Why, then, do we not find purpose in inanimate nature? It is poetic, and most misleading, to say, for example, that the purpose of rain is to water flowers, and yet one might say that a purpose of rootlets is to suck up moisture. Is this because rootlets are alive, but rain is inanimate? Surely not. Rain does not fall especially upon flowers, but flowers do take advantage of what rain there is, by specially adapted structures. Only living things and machines are adapted, in this sense, to take advantage of their environment, and only they display purpose. To state a purpose is, evidently, to specify what a thing is adapted or designed to do and we only find adaptation, or design, in organisms and machines. If one supposed a car engine to be a hair drier the exhaust pipe would have an obvious use, but the rotating shaft would not. It would seem to be a very poorly designed hair drier; noisy, smelly, dirty and inefficient, with a most annoying rotating shaft. The noise, and smell and the heat will appear 'accidental' once the shaft output is recognised as such, and quite different estimates of its efficiency will be made.

It is often difficult to decide whether an observed feature is functionally important or merely accidental. Any physical system will have some colour, but generally this is unimportant to its function. It requires a knowledge of function to classify observed features into 'essential' and 'accidental' properties, yet without hypotheses of function this classification is impossible. The engineer's insights into the functional significance of observed features would thus seem of the greatest importance to the classification of biological observations and findings. Biologists have tended to make this distinction implicitly, as may be shown by looking at their choice of words. Let us consider an example of this. *'Fatigue'* indicates loss of efficiency after prolonged use; *'Adaptation'* suggests that a change which might be identical, and might occur under exactly the same conditions as fatigue, is useful. Thus, 'the retina has recovered from *'fatigue'* and, 'the retina has become *adapted*', may refer to the same facts, and yet these sentences have different meanings for they imply different decisions on the significance of the facts. A useful feature in a machine might be regarded as having been specially designed to that end. Useful biological features tend to be regarded as the 'essential' features of the system given by specially adapted mechanisms. The decision between 'functionally important' and 'accidental' is often difficult to make in biology. Consider the continual small tremors of the eye ball. Are these

useful, or do they represent a failure of precise control? Do they serve to aid vision, perhaps by preventing retinal fatigue at contours by producing re-stimulation of the retinal cells? If the tremor can be shown to be useful in this respect (and this is an empirical question open to experimental test) would it be supposed that there are special mechanisms evolved to produce the tremor, as there are mechanisms which have evolved to give accommodation changes, and colour vision? Would it be worth looking for such special mechanisms? If we found a bit of the brain which seemed to serve no function but to induce tremor in the eye, would we be more inclined to think that tremor has been specially developed? If we took this course, perhaps we should look at earlier eyes, for the tremor mechanism might *have* been useful though now it might not be. Suppose in fact the tremor *is* accidental, perhaps mere noise in the system; a biologist might still do well to study its effect on vision, but its status would affect his way of thinking about the visual system as a whole. It might affect the choice of explanatory model we would develop. The engineer may be able to help here by suggesting what special features of the phenomenon to look for to make the decision between 'essential' and 'accidental'. Thus hunting might be distinguished, by statistical or other criteria, from noise. This sort of suggestion would seem to be of the greatest benefit to biologists. The engineer can help not only in suggesting models, but perhaps even more important, by suggesting what to look for to develop criteria for deciding between 'accidental' and specially adapted or 'essential' features.

Within the field of behaviour, probably the most important decision of this kind is whether the slowness of learning is specially adapted or is a limitation of the brain mechanism. One-trial learning is rather rare; why is this? If we knew what takes place in the brain during learning, then we could probably say whether or not this is a weakness of the design of the system, or whether it is specially built in according to genetic experience. Without this knowledge of the mechanisms subserving learning we can only ask: would one-trial learning be, on the whole, an advantage? If we decide that it would be a good thing, then it will appear a weakness of 'design', or inadequacy of the materials of the brain, that it is so rare. This is a shaky argument, but it is frequently used in biology for lack of anything better. In fact, we can point to advantages in *not* having one-trial learning; for unreliable inductive generalisations are less likely to occur. It is notable that Pavlov required a silent laboratory, with nothing untoward occurring during his Conditioning experiments: distractions made the process longer.* We might suggest that slow learning is generally

---

* This is a signal/noise situation. The irrelevant events constitute noise which has to be ignored as far as possible if responses are to be appropriate. This is a very general problem for organisms making decisions on sensory or other data cf. Gregory (1955, 1956, *refs.3,4*), Barlow (1956, *refs.1,2*).

useful, we may then be tempted to think that the brain could perfectly well have provided fast learning but that it would have been inappropriate. If so, certain types of engineering models may be ruled out. The biologist's assessment of the utility of the things he observed may affect the choice of 'engineering' model. The engineer might well question the biologist, and ask whether there are *any* cases where, for example, very fast learning does occur. If he can find such cases the limits of the system will have been extended. This might open up further possibilities, or rule out others. 'Imprinting' is interesting in this connection, for here very fast learning does occur, and it is plainly useful.

ON HOW AN ENGINEER MIGHT LOOK AT THE NOTION OF LOCALISATION OF CORTICAL FUNCTION

1. *What neurologists mean by 'localisation of function'*

It is often suggested by neurologists that some function is localised in a more or less specific region of the central nervous system. Thus it may be suggested that speech, or some particular feature of speech behaviour, is localised in Broca's area, in the left frontal lobe. Or it may be said that the pre-frontal gyrus is a 'motor area'. It is not always clear what neurologists mean by saying that a function is localised in a specified region of the brain. If this is not clear, any test or experiment involving behaviour is made difficult, and discussion is likely to be confused. I shall now try to examine the language used to describe localisation of function. We may then see whether the techniques used by engineers for establishing and describing function might be helpful in relating behaviour with neural systems.

If we consider the phrase: 'localisation of (cortical) function', we can be puzzled by the use both of 'localisation' and of 'function'. The neurologist generally means, evidently, that *a given region of (the brain) is especially associated with some particular type of behaviour.* This seems to be clear from the contexts in which the phrase is used - mainly ablation and stimulation experiments in which behaviour changes are correlated with brain damage or stimulation. This idea that a given area, or region, of the brain is particularly associated with certain behavioural functions goes back at least to the phrenologists. They spoke of a 'bump of intelligence' where, in some sense, intelligence was supposed to reside. The bigger the 'bump' the greater the intelligence.

2. *'Localisation'*

At first sight, at least, it may seem that the neurologist speaks of regions of the brain in the same way that the Geographer speaks of regions

of the Earth. Thus: "The Sahara is a region in Africa" and, "the striate area is a region of the occipital lobes", are similar sorts of statement. A desert ends when the sand becomes soil: the striate area ends when the cell structure changes in a recognisable way. Suppose now that there is no visible dividing line to delimit 'functional' regions of the brain: How could we specify such regions? The same thing happens in geography, and the analogy may prove helpful. In *physical* geography natural features, such as seas or continents, are bounded by shores; deserts end where the climate and soil change. In *economic* geography countries, or trading areas, may not be so bounded: they are essentially *functional* areas. These are functional boundaries, and make sense only in terms of the life of the people in each country - their customs, economics and government. These things cannot be observed like shore lines, by simply looking. A further point - a country is usually fairly compact, for reasons of transport and communication, but this may not be so. Thus in a sense the British Empire might be said to be, or to have been, one country though scattered across the world. The bits were linked by connections which could be described as 'functional' but they could not be observed as, at any rate static, 'structure'.

The distinction made between physical and economic geography holds for the neurology of the brain. A region may be isolated and named either (a) by virtue of its special *structure*, or (b) by virtue of its special *function*.

3. *'Function'*

In normal speech, to specify an object's function is to say what causal part it plays in the attainment of some end. It would seem that an object's function may be specified in two quite distinct ways, and the difference between these is important. Consider a rheostat used as a dimmer to control the brightness of a lamp. The resistor may be referred to in terms of its property of changing the current in the circuit, or simply as a dimmer. The word can only be used in the first way within the context of electrical theory; it is essentially a *technical* sense of the word. The same component may also be referred to quite non-technically, as: 'the thing which dims the lamp'. Here no special knowledge is required, beyond the generalisation that when the knob is turned the lamp dims. The grounds for using the word 'function' are quite different in the two cases, depending upon whether it is used to designate an essentially *inductively* derived generalisation that A produces B (turning the knob has always dimmed the lamp, and so it is a dimmer) or whether it denotes a *deductive* inference within a theory: (added resistance reduces the current flowing in a circuit; the reduced current gives less power .... therefore the lamp must dim). If the current was not reduced, it could not *possibly be* a resistance. It should be noted that we may be mistaken in *calling* a given object a resistance, but if it *is* a resistance then it *must* have certain functional properties in a given system.

Now it is clear that this second, deductive, use of the word can only be used within a deductive system, or model. The burning question is: has the neurologist got such systems or models to enable him to use 'functional' in this technical way?

It seems clear that the word 'function' is generally used in neurology in the non-theory laden, inductive, sense. If it is said: "speech is localised in Broca's area", or: "the function of Broca's area is to (with the vocal chords and many other things) produce speech", then we are using an inductive argument. The evidence for the induction is based largely on speech defects which are observed with lesions in this region of the brain. Broca's area is said to be a speech centre, without a notion of just *what* it does or *how* it does it. Again, it may be said that the striate area is used for vision, but we have no idea what processes are going on in this region during visual perception.*

It often happens that with man-made machines we can recognise the various components for what they are, or at least some of them. We may see that there are, for example, motors, resistances, levers or bearings in the machine. Once the components are identified as functional units, then deductive inference is possible. Of course we may always be wrong at the identification stage of the process: a resistance may be mistaken for a condenser, and then the premiss for the argument will be wrong. But the argument is still truly deductive, though the conclusion may then be wrong. Now if nerve cells varied in form in more different ways, each type having corresponding functional properties, then the neurologist would be in the position of the engineer in having readily identifiable components from which he might infer the function of the circuit. Unfortunately, the neurologist is perhaps never in this position, for the brain is like porridge. This means that this way is generally not open, for we cannot, at least at present, identify and recognise function from appearance by looking at cells and their connections.

---

* The neurologist's concern with cortical function is the extreme case; it would be interesting to consider questions of localisation of function in simple cases. There are, however, comparatively simple examples to be found in the brain: thus certain regions can be described as 'chemo-receptors', serving to monitor the blood concentration of CO_2. The signals from these regions have a direct action on respiration. I do not question that some areas have specific functions, but to state what their function is we must have a model of the system within which they play their part. To name an area as a 'chemo-receptor' is to give it a (sensory) function.
It might be recognised as such either (a) by the recognition of distinctive properties (like recognising a rose, or a friend) or (b) by recognising its place in a system. (a) Requires prior knowledge of such components or parts, (b) requires a knowledge of the rest of the system in functional terms. In either case, to specify the function of part of a system we must know something of *what* it does and *how* it does it. This is true of the simplest, as well as the most complex, types of system.

If any change takes place upon the removal of part of the brain, the changes are either (a) *loss* of some feature of behaviour, or diminution or worsening of some skills, or (b) introduction of some *new* behavioural features. Now it is often argued that if some part of behaviour is lost, or diminished in efficiency, then in some sense this behaviour (or rather the causal mechanisms necessary for this behaviour), are localised in the affected region. But does this follow? We may assume that the association is no chance association but a causal one, and still seriously doubt whether before the advent of the lesion the region in question contained causally necessary mechanisms for the affected behaviour. To illustrate this, we can take an example from radio engineering. If a main smoothing condenser breaks down, shorting the H.T. to earth through a low resistance, the set may stop working or work in a peculiar manner. The oscillator may stop although the rest of the system may continue working normally. Would we then say that the condenser was functionally important for the radio's oscillator stage, but not for the rest? If so, we would be wrong. Its purpose in the system is to smooth the ripple for the whole system, but it happens that a part of the system is more sensitive to reduction in supply voltage than the rest. Suppose that when the condenser breaks down, the set emits piercing howls. Do we argue that the normal function of the condenser is to inhibit howling? Surely not. The condenser's abnormally low resistance has changed the system as a whole, and the new system may exhibit new properties, in this case howling.

There *are* physical systems where removal of a part removes a specific feature of the output. Thus consider a piano: if a string, a hammer, or a key is removed one note is lost, and the rest may play as before. A piano is largely an arrangement of independent parallel systems, each with its own input (a key) and output (a string). Here functional localisation is a simple matter - a piano tuner soon knows where to look for any trouble. This is not in general true of machines, where loss of a part may produce the most bizarre symptoms. Only by understanding the principles involved, and the causal functions of the parts, can the trouble be explained. Further, even where there are parallel semi-independent systems, a fault in a part serving the various parallel systems may have a selective effect on them, and this can be confusing. Thus, reduced air pressure on an organ might affect some pipes a great deal and others not at all.

Mr. P. E. K. Donaldson has suggested to me that this point applies most particularly to tightly coupled systems, and the brain would seem to be such a system.

4. *Arguments from Stimulation*

If a part is stimulated and something happens, such as a muscle twitching, or the patient reporting a cloud of balloons floating over his head, what can be inferred? The considerations here are similar to the above. We now have a different system which might have quite new properties.

The functional 'centres' of Hess, and of the ethologists are interesting. If stimulation of the given region produces a sequence of movements similar to, or identical with, an observed innate behaviour pattern then, it is suggested, this region is the locus or 'centre', of this behaviour pattern. But there are difficulties in this idea of a localised functional centre. The word 'centre' here suggests that the causal neural processes leading to the innate behaviour pattern are located in space, closely packed, and even that there are not other causal mechanisms in the same region. But, just as many races may live in the same country, perhaps talking the same language and perhaps not, so causal mechanisms subserving different end results could well reside entangled. They may not be grouped neatly but might perfectly well be strung about in tenuous filaments, very difficult to find, identify, or stimulate in a controlled manner. Perhaps most important, why should the stimulation by some arbitrary 'signal' leading to familiar behaviour patterns be regarded as more interesting than some bizarre behaviour? Surely all it could mean is that some organised set of causal sequences has been set off, leading to familiar behaviour; but this happens with lights shone in the eye, or sounds applied to the ear, or any other stimulus, yet no one wants to say that the causal mechanism, the 'centres', are in the eye or the ear, except in very special cases. Indeed, it would be surprising if stimulation of the cortex did *not* sometimes produce behaviour sequences, but it would not follow that the region stimulated contained the causal mechanisms responsible for the activity. Certainly it must under these conditions have something to do with the activity, but it may be far removed, both spatially and causally. One might suspect that if a complete normal behaviour pattern is elicited with an artificial stimulus, then the stimulated area is rather less likely to be directly responsible than if the behaviour patterns are bizarre, for the stimulus might be expected to produce disruption of normal function if introduced in the middle of a causal mechanism, even if the stimulus is the correct trigger for the mechanism.

'Localisation of function' takes on further difficulties when we go from innate to learned mechanisms, for here each individual will differ according to what is 'stored' and also, perhaps, according to different developed 'strategies' adopted for thinking.

If the arguments (or at any rate the conclusions) suggested here are substantially correct, the changes produced by ablation or stimulation should be interpreted in terms of the changes to be expected in systems of various kinds. In a tightly coupled system it is impossible to specify a flow of information, or to say where any particular function is localised, for there is interaction throughout the system. Its performance may change when bits are removed, or stimulated, but the changes can only be understood in terms of the functional organisation of the whole system. In short, we need a model to interpret the changes, and this is where the engineer

should be able to help. The changes in behaviour associated with lesions might be important evidence for suggesting and testing models, and the models should serve to explain the effects.

Modern neurology started when large telephone exchanges were first being built. The 'telephone exchange analogy' has no doubt been useful in the study of the peripheral nervous system, but it is extremely misleading for tightly coupled systems, as the brain would seem to be. The input and output regions of the brain - the 'projection areas' - would be expected to be revealed fairly simply by suitable lesions or direct stimulation, for no doubt they lie in telephone-like pathways, as for the peripheral nerves, though even here there tend to be some cross connections which complicate the picture. The problem becomes acute where the nervous system is analysing or computing, for here the system cannot be like a telephone exchange. Histological examination suggests more or less random networks in large parts of the cortex (Scholl, 1957, *ref. 7*). Such systems are only beginning to be studied. Lashley's 'Equipotentiality' (that any part of the cortex is equivalent functionally to any other part, certain areas excepted) which he established for the rat brain (Lashley, 1929, 1950, *refs. 5,6*) would be expected for such a system. If part of a tightly coupled system is destroyed there is generally very little effect on its performance, except when it is pushed to its limits, for it is now a smaller system. This is just what Lashley found. It is interesting that Uttley's conditional probability computer consists in part of random circuits; some destruction here would not have any devastating effect on the function of this machine. (Uttley, 1955, *ref. 8*).

The behavioural effects of brain damage are of vital importance to the brain surgeon quite apart from their interpretations. They also provide fascinating data for insights into the function of the normal brain but only, I submit, if we look at the changes in the light of conceptual models of brain function. We only come to understand fully how car engines and radio sets work, even after reading about them, by noting their eccentricities and their ailments. We see why a certain 'symptom' is produced when we know how the normal system works, and we come to understand more fully how it works when we see and think about the symptoms of overheating, weak mixture, or whatever it may be. We could not possibly say what a plug in a car engine does simply by noting what happens when we remove it, unless we understand the general principles of internal combustion engines - we must have a model. The biologist has no 'Maker's Manuals', or any clear idea of what the purpose of many of the 'devices' he studies may be. He must guess the purpose, and put up for testing likely looking hypotheses of how it may function. In both these tasks, he should receive valuable help from the sympathetic engineer.

I wish to thank Mr. A. J. Watson, Dr. L. Weiskrantz and Professor O. L. Zangwill for valuable discussions, though I should not wish to commit them to the views expressed. I am indebted to Professor Zangwill for any knowledge I possess on these problems.

REFERENCES

1. BARLOW, H. B. Retinal Noise and Absolute Threshold. *J. opt. Soc. Amer.*, 1956, **46**, 634.
2. BARLOW, H. B. Incremental Thresholds at Low Intensities as Signal/Noise Discriminations. *J. Physiol.*, 1956, **136**, 469.
3. GREGORY, R. L. A Note on Summation Time of the Eye indicated by Signal/Noise Discrimination. *Quart. J. exp. Psychol.*, 1955, **7**, 147.
4. GREGORY, R. L. An experimental treatment of vision as an information source and noisy channel. in Information Theory: Third London Symposium 1955. ed. Colin Cherry. *London: Methuen.* (1956).
5. LASHLEY, K. S. Brain Mechanisms and Intelligence. *Chicago: University Press.* (1929).
6. LASHLEY, K. S. In Search of the Engram, Symposia of the Society for Experimental Biology IV. *England: Cambridge University Press.* (1950).
7. SCHOLL, D. A. The Organisation of the Cerebral Cortex. *London, Methuen.* (1957).
8. UTTLEY, A. M. The Conditional Probability of Signals in the Nervous System. *R.R.E. Memorandum No. 1109.* (1955).

DISCUSSION ON THE PAPER BY MR. R. L. GREGORY.

MR. J. T. ALLANSON: I want to take up two points which are not central to Mr. Gregory's argument. The first one is in relation to the systems which he has described as tightly coupled. Since this paper is largely concerned with the problem of what may be learned from engineering situations and engineering methods of analysis, I think it is relevant to point out that in many fields of engineering, tightly coupled systems do not obey two of the generalisations which are made in this paper. In the theory of electrical networks, for instance, it is possible to define quite precisely what one means by systems which are tightly coupled. By and large, such systems cannot satisfy at the same time, the generalisation on page 678 of possessing bizarre behaviour when part of the system is destroyed, and the generalisation on page 680 that if part of the tightly coupled system is destroyed there is generally very little effect.

I am doubtful if any system at all can satisfy both of these generalisations at one and the same time. If there is generally very little effect, I do not see how one can say that the effect is bizarre. I suggest that this phrase (highly-coupled systems) is not a particularly useful description or way of looking at the brain at all.

The second point centres on page 680 Mr. Gregory says if we take Dr. Uttley's probability computer and remove some of the circuits it would not have a devastating effect on the function of the machine. I think this depends entirely on the original state of the computer.

If there is a certain superfluity in the connections it is perfectly true. If one takes the connections as described for the precise version of the computer, in which they are not made at random, but there is *one* AB unit and *one* BC unit and so on, a removal of a percentage of these connections would result in bizarre effects rather than in a very slight modification without significance.

DR. A. J. ANGYAN: I was very glad to read Mr. Gregory's paper, and think I can agree with all his main points. I, too, shall try to demonstrate a model which is intended to be a representation of some kind of localisation of function in the nervous system.

I would like to know how Mr. Gregory differentiates between "essential" and "accidental" features when making a model. Are, for example, the "essential" features of a nerve cell not "accidental" when describing the

learning process of the organism in the more general sense, as compared with these terms when used to describe the anatomical and electrophysiological properties of the cell?

DR. W. S. McCULLOCH: I am one of those unfortunate men who began his study of neurophysiology in the days when practically everything had to be deduced from the effects of lesions. Eilhard von Domarus in his famous Ph. D. thesis at Yale (*ref. 1*) poked fun at us - I mean at neurophysiologists, thus:

> "Clinicians correlated the loss of specific responses with the destruction of particular portions of the nervous system and supposed the function lost to be that of the tissue lost, instead of realizing that all they knew was that the remaining functions were functions of the remaining structure. We might caricature this by saying that stereoscopic vision was the function of the right eye because a man who lost his right eye lost stereoscopic vision. The notion of inhibition, or neural shock, prevented a clear conception of what they did actually know. Now return to the caricature. Suppose that we had seen patients who had lost stereoscopic vision by loss of the right eye and attributed that function to that eye, and then had a patient who had lost a left eye; we might now explain his loss of stereoscopic vision by supposing that the destruction of his left eye, by inhibitorische Fernwirkung, prevented his right eye from functioning stereoscopically. And, as if this had not been enough, the gratuitous hypothesis of the vicarious assumption of function obfuscated all issues. We have discovered that the right eye gave stereoscopic vision and the left eye plain vision; but in our patient who had lost his left eye the right eye vicariously assumed the function of plain vision. This completes the caricature"..

In other words, the argument from lesions, unless we are careful, leads us into utter rubbish, and the great difficulty is that we have been compelled to think of new ways of defining functions. I have been trying to define functions in new ways because I am interested in the functional organization of the brain; but all that I can say is, that the function of any component is to be excited by those things which can excite it, and to be inhibited by those things which can inhibit it, and perhaps to drift a little in its properties with use. When we try to say more or speak of the function of a whole tissue or a whole area, we are very apt to talk nonsense.

REFERENCE

1. DOMARUS, EILHARD VON. The logical structure of mind, an inquiry into the philosophical foundation of psychology and psychiatry. Ph. D. thesis, Yale, (1934).

DR. M. M. MacKAY: I wonder whether we may not go too far in running away from the idea of localised function. I have in mind the rather analogous case of a human community, to which much that Mr. Gregory has said of the brain would seem to apply. The economist attributes a function to each man as a cog in the machine; but he also distinguishes functions such as that of an industry. The man is sharply localisable; the industry need not be, but often is - witness the power of strategic bombing.

I feel that this analogy can be useful and stimulating towards a more balanced view on the vexed question of 'function'. No economist pictures the country as parcelled into discrete 'areas'; but he gains great flexibility of thinking by the ability to picture more or less discrete entities with definite functions such as the cotton industry, the electronics industry, and so forth. I see no reason why evidence from comparative anatomy, from developmental studies, and above all from the action of drugs which are function-specific, should not give us valid functional maps (not necessarily having any simple geographical interpretation) with the same power to catalyse our thinking about cerebral organization.

DR. J. A. V. BATES: I think Mr. Gregory has brought out a useful point in stressing that there may be confusion in the use of the word "function". For many years people have pointed this out - soon after Broca published his observations on brain lesions affecting speech in 1861, Hughlings Jackson and Wundt were saying that to localise the parts of the brain which destroy speech and to localise the function of speech are two different things. So if this has been the obstacle to clear thinking that Mr. Gregory or Dr. McCulloch have made it out to be, I would say that there is no excuse. But I think the difficulty lies elsewhere, and in Mr. Gregory's terminology it is this: that if you have any complex organisation, there will be systems within it which are to be considered tightly coupled, and systems which are to be considered loosely coupled; if we study the effects of brain lesions how are we to know which sort of system we are dealing with? Probably the degree of coupling in sub-systems in a complex organ like the brain shows a hierarchy of grades, just as it does in the radio set which Mr. Gregory uses in analogy; if the set goes silent because a resistor has gone "open-circuit", he is correct to point out that it would be ridiculous to say the function of that resistor was to produce sound, but with the same set, if the same defect is found to be due to some change in the loudspeaker cone, it would not be ridiculous to say that the function of the loudspeaker cone was to produce the sound. We know that in the brain one can hit on examples of extremely tight coupling in this sense. For example, if one interferes with certain tracts in the brain stem and gets anaesthesia, one is not amiss in assigning to those tracts a particular function in conduction of impulses which give rise to the

sensation of pain. Other types of lesion produce altered sensation through interference with loosely coupled systems and one characteristic of such systems which the neurologist recognises is that such defects tend to recover. But one essential difficulty in studying brain function by means of lesions lies in deciding which sort of system you are dealing with. It would be instructive to learn if engineering has produced any general rules of theorems on this subject which could be of value in this type of investigation.

DR. H. B. BARLOW: I agree, of course, with a lot of what Mr. Gregory has said, but I am distressed by the general tenor of it, for this reason. Suppose nobody had ever done any ablation experiments, then after hearing his talk one would say, "Here are some experiments I need not do at all". But obviously a very great deal of valuable information about the brain has been obtained by ablation experiments. The fact that one can be confused by the results is true of any experiments. There are some cases, after all, where you do get quite clear-cut results. For example, if you hold a blind-folded cat so that the tops of its front paws touch the edge of a table, then it will lift them up and place them on the table. Now if a small region at the frontal pole of the cat's cortex is removed, then this placing reaction (on the contralateral side) is no longer performed: but if the whole of the cerebral cortex except this small region is ablated, then the placing reaction is retained. (*ref. 1*).

When you can get clear experimental results like this, the experiments are surely worth doing, even if the interpretation is not as simple as appears at first sight.

SIR FREDERICK BARTLETT, CHAIRMAN: There are two questions which I should like to ask Mr. Gregory to answer. Towards the end of his paper, on the last page, he says that if one is considering a biological system which is computing, or working in an analytical sort of way, we have to consider that system as containing a number of very large parts, and we have to have recourse to a notion something like, at any rate, that of Lashley's notion of equi-potentiality, where any part, within limits, has the same properties in response as any other part. So far as I can see, Lashley himself arrived at this principle through a very long inductive process, and I would like to know from Mr. Gregory if he considers that the distinction that he has drawn between the use in this field of an inductive process and a deductive process really does amount to very much -

*REFERENCE

1. BARD, P., Macleod's Physiology in Modern Medicine, 9th Edition, *Mosby, St. Louis*, (1941), 140.

whether, in fact, one has not got to use the experimental approach in order to get at any of these general properties of large systems.

The second question that I want to ask I can put very simply indeed. He says it is interesting that Dr. Uttley's conditional probability computer consists in part of random circuits. Does this mean that any model which has to be constructed and is useful to help in dealing with the central nervous system is all right so long as it contains a number of random circuits.

Finally, I wish he would say, for my information and for others, what sort of valuable help he wants from the sympathetic engineer.

MR. GREGORY (in reply): I do not know whether I can at all adequately deal with the points raised, but thank you very much for them. The first point raised by Mr. Allanson brings to light something pretty silly on my part. I did not wish to say that with a tightly coupled system one would *at the same time* get a bizarre change and no change at all when parts are removed. I think there is some confusion in my paper on this point. I should have said, perhaps, that (a) where the system is tightly coupled there would be only small output changes except when the system was being pushed to its limits, and (b) where there are serial processes without feed-back loops there will in general be rather large and bizarre changes in the output. I think that Mr. Allanson is correct in his second point - concerning removal of some of the circuits in Dr. Uttley's probability computer. It is perhaps interesting to note that in general lesions are more serious in adult than in young animals or people: could it be that there are fewer superfluous connections in the adult organism?

I think I do not fully understand Dr. Angyan's query about "accidental" and "essential" features of nerve cells. It seems to me that we might mean by "essential" those features which are of causal importance to the functioning of the system, while "accidental" are those features which are not causally important. A model should be helpful in indicating and emphasising the "essential" features of the system in this sense.

Dr. McCulloch seems to me to have put the main argument far more clearly than I have succeeded in doing. I am grateful for this support. It may well be that some of the early neurologists were less confused than many modern neurologists seem to be on this matter. Hughlings Jackson, for example, said very much the same sort of thing, though not in the context of engineering models of brain function.

I agree with Dr. Mackay that analogy with localised functions in human communities is useful. I have tried to develop this line of thought in a chapter in a forthcoming book (*ref. 1*). I think with him that the

REFERENCE

1. Current Problems in Animal Behaviour ed. by Zangwill & Thorp. *Cambridge University Press*. (In the press).

important point is that functional maps in general do not have simple geographical interpretations, either in communities, industry or machines. This is partly why simply looking at a system and removing "parts" is not sufficient for gaining an understanding of the manner of function of the organism.

I do not altogether follow Dr. Bates' first point. Perhaps the difficulty lies with the undue emphasis which I placed on "tightly coupled" systems. One might say that the sound comes from the loudspeaker cone, but I think that this is because it is the output terminal of the system. I tried to emphasise that in my view one *can* localise input and output terminals, and parallel paths, quite simply.

Dr. Bates seems to be using the term "tightly coupled" in a sense different from the way I was using it. Perhaps he means what might be called "closely matched". I meant not closely matched, but rather a system of servo loops tending to give stability to the system as a whole. I do not agree with Dr. Bates that interference with certain tracts of the brain stem with anaesthesia would *necessarily* imply that those tracts function to conduct impulses normally giving rise to pain. There may well be further considerations which together with this observation would lead to such a conclusion, but in itself it is surely not sufficient grounds for the conclusion. All we could say is that when these tracts are anaesthetised the system does not give rise to pain. It remains a question *why* it does not give rise to pain. *One* possible answer is that the tracts were conducting impulses representing pain information. But this is not the only possibility, and so it does not *follow* that they were in any direct manner important to giving rise to pain.

In respect to Dr. Barlow's point: the last thing I wanted to suggest was that ablation experiments are a waste of time. I tried to make this clear at the end of the paper. I was, however, concerned to point out that their interpretation seems almost impossible, (except in a few special cases) *in the absence of some functional model*, however tentative this may be. It seems to me that we can seldom say anything *directly* about what a given region does by observing changes in behaviour after ablation. The exception is where there are parallel paths, and this is the case for the visual projection areas, if they *are* visual projection areas.

Sir Frederic Bartlett points out that Lashley arrived at his principle of equipotentiality by lengthy inductive processes. It seems to me that this is so if we take the term "equipotentiality" to refer to the observed behaviour changes with various amounts of ablation, but not if we take it to refer to properties of the neural structure responsible for the behaviour. To infer something about the function of the brain from observed behaviour does not seem to me to be an induction from the data. It seems to be a hypothesis which may be suggested by the data, but which

cannot be *inferred* from it, either inductively or deductively. It is not always clear whether writers on equipotentiality are using the term as a short-hand label for the inductive generalisations, which Lashley made, or whether they are using it as an explanatory concept to refer to a supposed property of the brain: namely that any part is functionally equivalent to any other part. If it is a label then it is, so it seems to me, an *induction*, but if it refers to the brain then it is a hypothesis about the neural systems involved, and it is an attempt at *explanation*. I think we should be very careful to make it clear when such a term is being used as a label and when it is being used to denote an explanatory concept.

To suggest that the brain is a system which might on general grounds be expected to display impairment only when pushed near its limits, is to make any special postulate unnecessary. It thus seems to me that defenders of equipotentiality as an explanation of the facts should show that the brain is not such a system. Until this is done, any special postulate for the system is unnecessary. It is also a poor explanation, because it does not relate the case of the brain to other systems whose functional principles are known to us, but even suggests that in this particular it is unique, which may inhibit the development of explanatory models in terms of known systems.

I fear that I have raised (or fished up?) a red herring in putting emphasis on random circuits in this context.

Dr. Bates has largely answered the query as to what sort of help the sympathetic engineer might provide. He points out that an essential difficulty in studying brain function by means of lesions is to know which sort of system we are dealing with. It seems to me that the engineer can be of great help in suggesting just what to look for, and what sort of experiment to perform, in order to discover what kind of system it is. For example, once we know about servos and their properties under various conditions, we know that studies on tremor are likely to be important, and careful experiments are then carried out which would be thought silly, or even mad, in the absence of the alternative models suggested by analogy with other control systems. Again, in the context of engineer's systems Lashley's ablation results seem to have a simple explanation. In the absence of engineering knowledge special principles tend to be invoked by psychologists and physiologists. Since these do not have application beyond the biological examples considered, they have little explanatory power, and are often merely alternative labels for what needs explanation. An example of this would be Bergson's elan vital, which is utterly useless as an explanation, because it does not compare the unknown with anything else. It is no more than a label, but such labels can have a fascination; they tend to quell further questions which might lead to new experiments, and explanation in general terms.

SESSION 4A

PAPER 6

SOME QUESTIONS CONCERNING THE EXPLANATION OF LEARNING IN ANIMALS

by

A. J. WATSON

BIOGRAPHICAL NOTE

Mr. A. J. Watson was educated at Birkenhead School and Corpus Christi College, Oxford. At the latter, he read the Final Honours School of Psychology, Philosophy and Physiology. Subsequently, for one year, he was Senior Scholar of Christ Church, engaged in research in philosophy. From 1952-4, he was Junior Lecturer in Experimental Psychology at the Institute of Experimental Psychology at Oxford. Since 1954, he has been Lecturer in Experimental Psychology in the University of Cambridge.

SOME QUESTIONS CONCERNING THE EXPLANATION OF LEARNING IN ANIMALS

by

A. J. WATSON

SUMMARY

THIS paper is concerned with three problems concerning the nature of learning in animals. These are: (1) What is learned? Do animals learn to make certain patterns of muscular response to stimuli, do they learn to seek stimuli as targets, etc? (2) Is a reinforcement principle necessary to explain learning? Do animals only learn when some goal-signal, or substitute for a goal-signal, is subsequently received? (3) To what extent are animals capable of abstracting from the stimuli with which they are presented?

These questions are discussed with reference to samples of the relevant experimental evidence and to some psychological theories which have been developed to answer them.

IF I were asked actually to construct an artificial intelligence, I should be totally at a loss and I should be nearly as helpless if I attempted to understand a technical explanation of the mode of operation of such a device. Since I know nothing of the engineering problems and difficulties arising in this kind of work, it may well be doubted whether anything that I can say will be of any relevance to the questions with which this meeting is concerned. I can speak only of the behaviour exhibited by natural intelligences and, even here, only of that displayed by one species, namely the laboratory rat. This work may, however, be of some relevance in that this is a rather artificial animal.

There seem to be two ways in which experimental research of this kind may have a bearing on the study of "artificial intelligence". The first is in providing the data which is to be explained, or imitated by such a mechanism. There is obviously little point in elaborating a mechanism which

is not designed to imitate a reasonable portion of the known characteristics of behaviour in various circumstances. Although there may be serious doubts about the validity of many of the conclusions drawn from the experimental work on animal behaviour, and although the experiments, when repeated, often fail to show the degree of consistency which we could desire, they do serve to indicate that this behaviour, and the conditions controlling it are much more complex than common observation might suggest. So, uncertain though they may be, we must rely upon the results of such research for the facts which are to be explained. Of course, this consideration need not hold when we are concerned with the design of mechanisms which, although they provide outputs which one might wish to describe as "intelligent" are nevertheless not directed to imitate the behaviour of any actual animal. The suggestion here will be that since most functions that such "intelligences" will be designed to perform will find a limited analogy in some restricted aspect of animal behaviour, the study of the latter may prove fruitful in the execution of such a design. This is clearly possible, but I do not see how, a priori, one can say how far it will prove helpful in this way. I shall here consider the study of "artificial intelligence" as an attempt to design mechanisms which imitate, and therefore explain, the characteristics of behaviour actually shown by animals. For this purpose we clearly must have a statement of what it is that is to be imitated, and this is the first and main way in which biological work of this kind is relevant to the subjects under discussion here.

As a matter of logic, however, there is a second way in which such work is relevant. It is not the case that students of behaviour have selected variables for investigation at random. Neither has this work been directed only by generalization from already established facts. For, while part of it is suggested by empirical generalization from previous knowledge, there is also a substantial section of such research which has been inspired by theoretical considerations. These theories are not all of one logical kind; and, considered as scientific theories, they all tend to have certain logical peculiarities. But at least some of them have been claimed to be explanatory in that they tell us something of the nature of the mechanisms controlling behaviour. Now these explanations are not necessarily couched in physiological terms or in terms of any particular embodiment of the theory. Rather, what is claimed is that by observation of, and experiment upon, input-output relations of the organism it is possible to infer, and to test, hypotheses about the kind of function, or transformations, which must be carried out by the mechanism in order to produce the observed behaviour. That this is logically possible seems clear. But it is difficult to say, in advance, to what degree of detail it is possible to carry on this kind of theorizing in the absence of knowledge of the nature of the components actually forming the mechanism and of their engineering

possibilities. It is, of course, at this point that suggestions and analogies from the study of man-made control systems are likely to be of assistance in understanding animal behaviour.

I wish to present some conclusions and suggestions of this theoretical kind, together with a selection of the evidence which gives rise to them, and which they are directed, in this preliminary way, to explain. In so doing I shall be attempting to give one or two features of the specification, which any artificial intelligence would have to fulfil, which seem to me to be indicated by our present knowledge.

I. WHAT IS LEARNED?

Learning tends to be regarded as a central feature of intelligent behaviour, no doubt for good reasons. A machine which cannot learn, no matter what other feats of prodigious calculation or complex adaptation it may accomplish falls far short of displaying those characteristics of behaviour which, in animals, we call intelligence. It is at least debatable whether a machine which can learn does not show *all* the characteristics of intelligent behaviour, though possibly to a limited degree only in a quantitative sense. Now it had rather been assumed by many psychologists that when an animal learned, it learned to *do* something in a given situation - to beg when told to do so, to jump in a certain way beneath a fruit tree when hungry, to make a certain pattern of muscular movements when at a certain place on the way to a water-hole and so on. Ordinary language tends to make distinctions here. We speak of, for instance, knowing the way to a given place and knowing how to do something, e.g. how to ride a bicycle. But since such distinctions need have no implications about the nature of the mechanisms controlling behaviour it has often in the past been thought that the former, knowing a route, etc. was reducible, as far as the mechanism is concerned, to the latter, knowing what to do in certain circumstances. This, of course, was the natural consequence of attempting to explain learning in terms of an elaborated conditioned reflex mechanism. There are numerous familiar objections to an account of learning in these terms. It has been observed by many - for instance, by Hughes and Schlosberg *(ref. 11)* that the classical conditioned response is not a replica, even in attenuated form, of the unconditioned response. This, indeed, is also clear from Pavlov's original results *(ref. 18)*. What one tends to find is a different, though partially similar response which it seems natural to regard as of an anticipatory nature. Further, it has also been clear for many years that it is not helpful to regard all behaviour as a combination of reflex and conditioned reflex responses. When used in this way, the word "reflex" tends to be devoid of meaning. Certainly, its explanatory implications are lost. This difficulty, apparent in any study

as "instrumental conditioning" and indeed also apparent in common observation of behaviour, led some theorists to distinguish between "operant" and "reflex" behaviour. It must be supposed that a large proportion of the responses observed in, for instance, an experiment on learning are simply "emitted" by the animal. These are "operant responses" and learning consists in the selection of an appropriate group from an animal's "operant repetoire" in order to obtain some reward. Other theorists continued to regard behaviour in terms of reflexes, conditioned or unconditioned, but, in fact, they seem to have adopted little more than the word from Pavlov's and Watson's earlier work. Thus, although Hull *(refs. 12,13)* speaks of conditioned reflexes and responses, it is clear that he is talking, for the most part, about what others have called operant behaviour and his interest lies not in showing that there is anything particularly reflex about such behaviour but rather in attempting to systematize those features which govern the selection, during learning, of some responses rather than others from the animals' repetoire.

There remains, however, this legacy from the original "conditioned reflex" view of behaviour, that most theorists have assumed that what is learned is what to *do* in given circumstances. The emphasis is upon the action. An animal is thought to make a given response to a given stimulus; if certain subsequent events occur, then the condition of the mechanism must be in some way altered so that this response becomes linked to the stimulus in question. There has been, of course, a great deal of controversy about what are the essential conditions for any such linking to occur. But many theoretical views which conflict on this point agree in regarding learning as a "one stage" process of direct connection between input and output.

Now there seem to be good reasons for rejecting a "one-stage" account of learning of this kind, if it is to be interpreted literally. Several studies have shown, in various ways, that if animals do learn to make responses - that is, to give a certain muscular output, this is not all that they learn. Thus, it has been shown that if a rat is trained to run a maze, and its mode of progression is subsequently altered by, for instance, flooding the apparatus *(refs. 8,15)*, or by inflicting upon the animal cerebral lesions which prevent walking and moving in the usual way, the animal will still traverse the maze by whatever means is available to it, with few or no errors. In the second of these experiments, there is reported at least one animal which was rendered quite incapable of any form of walking, but which, by rolling and scrambling, still found its way through the maze without error. Here, then, it cannot only be that the animal had learned to produce certain definite kinds of muscular output in response to certain cues. It has also been shown, for instance by Gleitman *(ref. 9)* that animals will learn about a situation before they have made any particular response to it. In this experiment rats were placed in a small transparent cage. A

shock was administered through the floor of this cage and it was simultaneously moved from a point A to a point B in the experimental room. At B, the shock was switched off and the rat was released. After several such trials, the animals were tested in a T maze in which the crosspiece of the T ran along the path traversed by the cage. At the choice-point, therefore, the animals had a choice between going to that place at which the shock commenced or to that at which it ended and at which they escaped. The latter was chosen, although no *responses* had previously occurred during training. In general, the extent to which animals learn to make specific responses seems to vary with the experimental situation. In enclosed alley mazes, under certain conditions and after considerable practice, rats seem to rely, rather heavily, upon learning to make a sequence of definite movements in response to certain cues. It has frequently been noted that under these conditions they will run head-on into obstacles newly placed in the alleys, etc. *(ref. 4)*. On the other hand, in elevated mazes, during the earlier stages of training, the animals seem to learn to approach or avoid certain places or features of the environment, irrespective of the responses they have to make in order to do so.

It seems probable, then, that even if animals do learn to make specific responses to cues or stimuli this is not all that they learn. It is reasonable to suppose that they also learn the sequence of cues which should be sought in order to obtain some goal. It may be useful to emphasize the difference between these two views. Consider a discrimination experiment in which a jumping stand is used. Here the animal is taught to jump from a platform to one of two or more windows. A correct choice leads to a food reward on a platform behind the window and an incorrect choice to punishment. Which window is correct in any given trial is indicated by the position of the cues between which the animal is being trained to discriminate; and the position of these cues is so changed during a series of trials that each window is correct equally often. Now the response of jumping is very similar whichever window the animal jumps towards, for the windows are placed symmetrically about the stand from which the animal jumps and at the same distance from it. However, the animal must orientate itself slightly differently for the two jumps and, if we include these responses in the total response, it may be said that two different responses are made. If these are designated R_R and R_L, then we may contrast the views of the nature of the learning occurring here as follows:-

(a) If the animal learns to make different responses to different cues it it must learn

$$BL \longrightarrow R_L$$
$$WL \longrightarrow R_R$$
$$BR \longrightarrow R_R$$
$$WR \longrightarrow R_L$$

where L and R indicate the sides upon which the cues, say black and white, are placed. Suppose here that a jump to black is always rewarded and a jump to white is punished.

(b) If, however, the animal does not learn to make a specific response, but learns rather only to approach or avoid cues - R_{A_D} and R_{A_V}, then it must learn

$$B \longrightarrow R_{A_D}$$

$$W \longrightarrow R_{A_V}$$

It may be suggested that in both the above cases there is needless complication, for in order to discriminate successfully the animal need only, in the first case, learn two things - say, $BL \longrightarrow R_L$ and $BR \longrightarrow R_R$; and in the second case, it need learn only either $B \longrightarrow R_{A_D}$ *or* $W \longrightarrow R_{A_V}$. It may, however, be demonstrated that in fact animals learn both to jump to the positive stimulus and also to refrain from jumping to the negative cue.

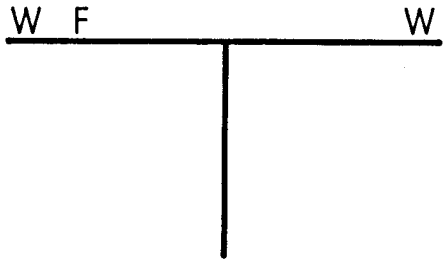
Both the above kinds of analysis have been proposed by psychologists. There are three differences between them which should be noted (a) the first system "connects" responses directly with stimuli and may be called a one-stage process, while the second system may more appropriately be called a "cue-seeking" device. (b) The first system has, in a sense, to learn more than the second - the second achieves an economy in learning (of the kind so far discussed) at the expense of complexity at the motor stage. (c) There is no account of how the second, a cue-seeking system, manages to do the correct thing after it has selected the cue which is to be approached. There are clearly, in principle, a number of possibilities open here. In the extreme case of inefficiency, the system could simply try out all the responses in its repetoire and continue to do so until it arrived at the one selected. Such a system would manifest no learning in its behaviour until it attained its goal. In the present case, it would be just as likely to jump to the incorrect window, or, indeed, to do anything else, as to jump to the correct window. And this behaviour would be repeated no matter how many training trials it was given. On the other hand, it would still have learned something and would demonstrate this learning when it did eventually respond correctly, for this would terminate its search. It is obvious that little or no behaviour is of this random kind. A second class of possibilities is that the mechanisms of moter control for approaching or avoiding cues, once selected, are built into the organism. This may in fact occur in the case of certain features of behaviour in certain species. There seem to be several kinds of control of this kind which could obtain. Thus, if an animal has to move from some position towards a cue it can see, it might be that it selects some response from its repetoire and then constantly corrects this in accordance with the way in which it finds itself travelling towards, or away from the goal. Or again, the system may be so constructed

that when a given cue has been selected as a goal, a certain more or less stereotyped response is made. In a fairly constant environment this may well be a reasonably effective procedure. It may be that such a response is simply made once, or that it is repeated until the goal is attained. Behaviour which seems partly to be similar to something of this kind has been reported by ethologists in their studies of instinctive behaviour. No doubt such systems as these, and many others, are employed in the control of the behaviour directed to the attainment of different goals and cues. But it seems most unlikely that such innate, or "built-in" mechanisms can account for all such behaviour, for a rat may be taught to perform any of an almost unlimited variety of contortions in order to obtain a reward, or some cue which leads to reward. It seems unreasonable to suppose that all these patterns of response are built into the mechanism. Moreover, animals, and certainly human subjects, show a gradual and often prolonged improvement of mechanical dexterity when performing many kinds of skilled tasks. Of course, some of this improvement may be attributed to learning of the kind mentioned above; that is, that the organism selects more effective cues to search for in the progress of the skill. But it seems reasonable to suppose that, in addition to any learning of this kind, the subject is also learning better ways of obtaining these cues or goals. Thus, the economy of learning in favour of the second system mentioned above, namely that which learned to approach or avoid certain cues, may be misleading, for it will probably often be the case that such learning must be supplemented by this second form of learning. The animal must learn not only which cue to jump at, but also how to jump at that window rather than anything else.

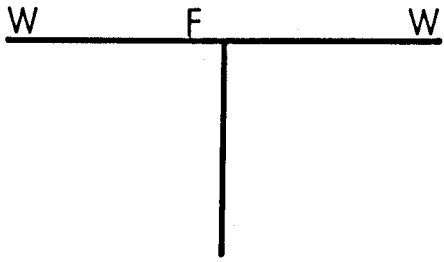
The fact, however, that these two forms of learning will commonly be found to be taking place simultaneously, and that it will often be very difficult to analyse experimental results in such a way that the two components may be assessed separately, should not lead us to reject a "two-stage" hypothesis of this kind. If correct, it would seem to imply that we should look for two distinct controlling systems, one which learns what to search for and the other what to do in order to arrive at what is searched for. This kind of distinction is contained in, for instance, Deutsch's theory of behaviour *(refs. 6, 7)*. This, although it is concerned to make hypotheses mainly about the nature of a "cue-learning" system, implies that the output of this system must be handed on to a system of moter control which will involve further learning. Certain advantages follow from this distinction. It becomes intelligible how animals can display both a retention of learning in circumstances in which no response has been practised, or in which the mode of response hitherto used is no longer suitable, and also an ability to perform complex manipulative movements with increasing efficiency. It may also be suggested that, with practice, an animal may in this way come to perform certain tasks with increasing efficiency in ways which seem, to some extent, to be similar to the behaviour that we actually observe.

Consider, for instance, a hungry rat being trained to run a straight alley, for a food reward. Several observations suggest that, as training progresses, the animal's response tends to become increasingly automatic and, as mentioned above, a point may be reached at which they will run full-tilt into a newly-interposed obstacle. Now presumably the greater the density of cues for which an animal searches, and therefore the more decisions he has to make in his progress down the alley, the longer it will take him to reach the food. And, further, a greater proportion of the system will be occupied in learning and controlling this behaviour. It would therefore be advantageous for an animal, in a case, such as this, to minimize the number of cues for which it seeks and to learn some more or less protracted response which leads it from one cue to the next. In this instance, we might expect a rat to learn to approach a cue at the beginning of the alley, and then to exhibit, as it were, a certain amount of locomotor activity, which will take it almost to the end of the alley. At this point a further cue will be registered which would end this section of activity and some other response would then be made. In general, we might expect that the density of cues which would be learned in such a task would be determined by the number of times the form of response has to be modified. But, clearly, the animal has to discover when and where such modifications will be required. In the course of learning, therefore, we should expect that initially a rat would learn a certain amount about a considerable number of cues - that is about the sequence of cues down the alley, and that his behaviour would be determined by all of these. As learning progresses, however, it may be that most of these cues are eliminated as he discovers that a single unchanging form of response will take him from one cue to another considerably removed from the first. Hence faster running times will be obtained, not only because the animal knows more about the route to food, because retracing and exploratory activity are reduced, etc., but also because the rat is able to pay attention to fewer cues and hence to maintain uninterrupted running over longer stretches of the maze.

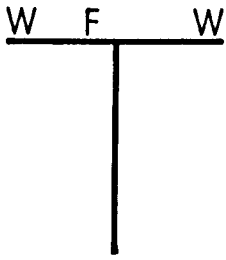
If any such view as this were correct, certain consequences follow. If during training on a task, the improvement in an animal's performance is to be attributed in part to the reduction in the density of cues which he utilizes to guide his response, we should expect that considerable limitations would be imposed upon the extent to which he learns about the environmental cues which he passes. There is some evidence which can be interpreted in this way. For instance, it has been found that rats tend to fail to learn the location of an irrelevant incentive when this is placed close to a relevant incentive, and also tend to be successful in this task when the two incentives are somewhat separated spatially *(refs. 16,25)*. Thus, a rat is trained, when thirsty, to run to a water reward which may be found at both ends of a simple T maze, and is by artificial techniques induced to go to each arm equally often. Food is present in one alley and



(i) Rats do not learn position of food.



(ii) Rats do learn position of food.



(iii) Rats do not learn position of food.

W — water
F — food

Fig.1(a)

not in the other, but since the animal is thirsty and not hungry it never eats this food. If the food is close to the water, and the animal is now made hungry, it will demonstrate, on the next trial, little or no preference for the arm of the maze in which the food is located. If, however, the water is at some distance from the choice point, and the food is close to the choice point, such a preference will be exhibited. Finally, if both incentives are close to the choice-point, with the water just beyond the food, again no preference is shown. *(fig. 1(a))*. Various explanations, none of which seem to be entirely satisfactory, have been proposed for these effects. A possible explanation, relevant here, is that during the period of training to run to water, the animal reduces to a minimum the number of cues which are used to guide the response. Suppose that, having made a choice, the animal requires to pick up a cue to inform it of the consequences of having made that response - i.e. to tell it which alley has been entered. It may then release a uniform running response which will take it down the length of the alley to the point at which a further cue is received which immediately precedes the water. At this point the form of response must again be modified. If such an account were correct then we should expect the animals to learn more about the food position when this was far from the water and close to the choice point, than under the other conditions. For here it may use the food as a cue during the training period. When, however, both incentives are close together, and near or far from the choice point, it may be able to make use only of those cues immediately preceding the water. We should then expect that if, under conditions where the location of food is not normally learned, the food, or the precise position of it, were used as a cue to the position of the water, and the animal was thus forced to use this cue in guiding his response, a test when hungry should show a preference for that arm of the maze in which the food is located. This result has, in fact, been found, under one condition *(ref. 1) (fig. 1(b))*. Many further experiments would, of course, be required to support such a hypothesis. But, it may serve as an example of the kind of suggestion under discussion and to illustrate the difference between the learning of what cues to seek and what responses to make in order to attain them.

A distinction of this kind, of course, raises difficulties in the interpretation of much data recorded in animal experiments. Consider the case of a Skinner box in which a hungry animal learns to press a lever in order to obtain a small food reward. If the above kind of interpretation is extended to this case, we should suppose that the rat learns to seek some cue consequent upon pressing the lever in some way, and, in addition, it learns what pattern of response to use in order to obtain, or "approach", this cue. If these two processes are taking place simultaneously, the learning curve obtained must be regarded as due to a combination of

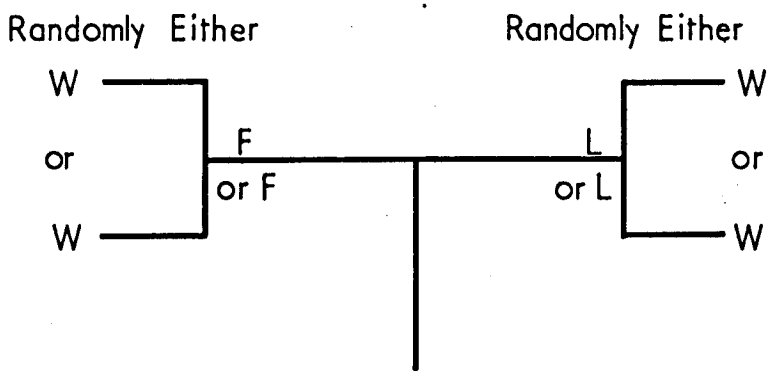


Fig.1(b)

these features. Again, if we now extinguish the response by ceasing to reward the behaviour, we must distinguish between a decrement in performance due to the fact that the animal ceases to search for this cue rather than others, and a decrement due to the rejection of this response as a means of obtaining this end. In principle, of course, these factors are separable. For if the only change in the situation during extinction is the failure of the food supply after the depression of the lever, then the cue for which the rat is searching, in making this response, will still be obtained. Extinction in this case will presumably principally involve the rejection of this, or some subsequent cue, as a stimulus to be sought en route to the food. Conversely, a situation might be devised in which extinction involved, at least initially, mainly rejection of a given response pattern as a means of obtaining a cue. Thus, if the cue for which an animal is searching consists of various stimuli which he receives when the lever is fully depressed, and the mechanical properties of the device are altered so that the lever cannot be depressed, extinction may, in the first place, be a matter of rejection of pattern of response. However,

although different situations can be devised theoretically to distinguish between these two kinds of extinction, in practice matters will be much less simple. Two points may be mentioned.

First it will usually be far from clear what cues an animal is searching for in such a situation. It is common, for instance, for the food-delivery mechanisms in Skinner boxes to emit a rather pronounced click as they operate and as the pellet of food drops into the food-cup. The sequence of events is, therefore, that the animal presses the lever, which is a short distance away from the food-cup, releases the lever and hears the click more or less simultaneously, and runs over to the food-cup to find the food. Now it is found *(ref. 3)* that, in extinction, many more unrewarded responses will be made if the empty food-delivery mechanism continues to operate, and thus produces a click, than if it is entirely disconnected, with the result that the depression of the lever is followed by silence. This is true even when no artificial delay in the presentation of the reward is introduced. Differing interpretations will be given of the extinction curves obtained in such experiments, according to what we suppose to be the cues for which the rat seeks. Suppose the animal has learned to seek some cue resulting from the depression of the lever - say the sensory feedback he receives when the lever encounters a stop - and has also learned that this cue is followed by the click, and that this, in turn, is followed by food. In the one case during extinction, he learns that the click is no longer followed by food, and in the other case that those cues derived from the depression of the lever are not followed by the click. In either case, we have an instance of extinction of cue learning and some explanation is required of why the one case results in more rapid extinction than the other. There seem to be one or two possible lines of attack on this problem, but they are not relevant here. It may, however, have been the case that the animal learns to depress the lever in order to obtain the click, and has also, as above, learned that this is followed by food. If this is what has been learned, then, in the one case, the animal learns, as before, that the click is no longer followed by food; while in the second case, he learns that a particular response is no longer followed by the click. Here then, when the click is present, we have extinction of cue learning, and when the click is not present we have, at least initially, extinction of response learning. The difference in rates of extinction might here be due to the different rates at which extinction of these two kinds occurs. It seems reasonable to expect that extinction of response would tend usually to occur rather faster than the extinction of cue learning for, if an animal has learned that one cue leads, by means of the execution of a certain response, to another which, say, in turn leads to a goal, and if this conjunction of cues now fails to occur, it seems likely that the most efficient strategy on the part of the animal will often be to vary its form of response in order to restore the conjunction before abandoning the first stimulus as a cue

leading to the goal. In any event, the interpretation we place upon the data obtained from extinction studies in Skinner boxes will clearly depend upon what cues we believe the animal to have learned as signs leading to the goal. And it will often be difficult to determine these.

A second possible reason for caution in interpreting this kind of data, if the distinction between cue-learning and response-learning is valid, should be mentioned. It is clear that the performance of an animal in any task is greatly influenced by its prevailing state of motivation and by such factors as the amount of incentive - e.g. food - with which its performance is rewarded. It was at one time held that these factors, especially the amount of reward, influenced not only the animal's current performance but also how much it learned on a given training trial. Thus, if two groups of animals were trained in a maze, one group receiving 1gm of food at the end of each run, and the other obtaining only 1/2 gm, it was thought that the first group should learn the maze in a smaller number of trials than were required by the second group. This view, at least in so simple a form, seems not now to be commonly accepted. And providing that the amount of reward is neither very small not very great and also that each individual animal finds only a single amount of reward at a single place in the maze, there seems to be no good reason to suppose that amount of reward influences rate of learning, if this is measured in terms of errors. On the other hand, amount of reward clearly does influence performance if this is measured, for instance, in terms of running speed *(ref. 5)*. Moreover, any sudden change in the size of the incentive also produces a rapid change in the running speed obtained. Thus, a group of animals which has been receiving 10 gm of food as a reward will, when their reward is halved, rapidly decrease their running speed to that of a group which has been receiving only 5 gm. Indeed, their speed usually drops below that of the control group. It seems possible, then, that changes of reward affect what we may, for present purposes, very roughly characterize as the "vigour" of a sequence of behaviour. If this is correct, then it seems possible that the curve obtained during, for instance, extinction in a Skinner box, may represent a combination of two quite different effects. The animal may, for instance, be learning that the click is no longer followed by food, and hence be tending to extinguish the tendency to seek the click. But, at the same time, the general "vigour" of the behaviour may be falling, due to the change from reward to no-reward, with the consequence that there will be a tendency for the lever to be depressed progressively less forcefully, and hence some signs of extinction will occur for this reason also.

Thus far, I have tried to suggest that there is some reason for supposing that learning in rats should be regarded as a two-stage process in which the animal learns (a) a sequence of cues which he should, in general terms "approach" or "avoid" in order to obtain certain goals; and

(b) that, for at least some of the behaviour which rats display, it seems likely that learning also occurs in those mechanisms which control the motor output that results in "approach" or "avoidance" of certain cues. It is further suggested that within such a two-stage process certain economies might be made in the time required to execute a task and in the mechanisms necessary for its execution, and that this leads to testable predictions. On the other hand, examples have been given in which the supposition that learning mechanisms should be thus distinguished renders the interpretation of simple learning and extinction data rather complex. This should be borne in mind in any attempt to draw from such data precise inferences about the characteristics of parts of the mechanisms involved in the total process.

II. THE CONDITIONS GOVERNING LEARNING

I wish now to turn to one major issue concerning the conditions governing learning. This discussion will be restricted to what has above been called cue-learning. It may be that the considerations which arise here also apply, at least in part, to response-learning. But it seems likely that this should be discussed separately. However, in so far as the issue with which I shall here be concerned has commonly been discussed within the framework of a "one-stage theory", of direct connection between stimuli and responses, reference to the animal's response is to some degree unavoidable.

The question at issue here is whether or not it is the case that learning only occurs when it is accompanied, or followed, by some "success signal" indicating the achievement of some goal; and, if this is the case, in what way does this goal signal function retrospectively to permit learning. It appears that any "one-stage" theory of learning in terms of connection of stimulus and response must necessarily suppose that learning is dependent upon some kind of goal signal, and it may be useful to state, in simple form, what may be regarded as the early classical account of learning in the case of, say, a rat learning a single-choice T-maze. Consider the first run of a hungry rat through such a maze, with a food reward at the end of one of the arms *(fig. 2)*. Suppose that a certain constellation of stimuli S impinge on the animal at the choice point, and that he has open to him only the two alternatives of turning left, R_L, or right, R_R. Suppose, first, that a correct choice is made and the rat turns left. Then some kind of temporary connection will be established, $S \longrightarrow R_L$. Shortly afterwards the animal finds food and it is supposed that this confirms - i.e. renders permanent, a fraction of the temporary connection. Secondly, suppose that an error is made on this first trial and that the rat turns right. Then $S \dashrightarrow R_R$ is temporarily established. If now the

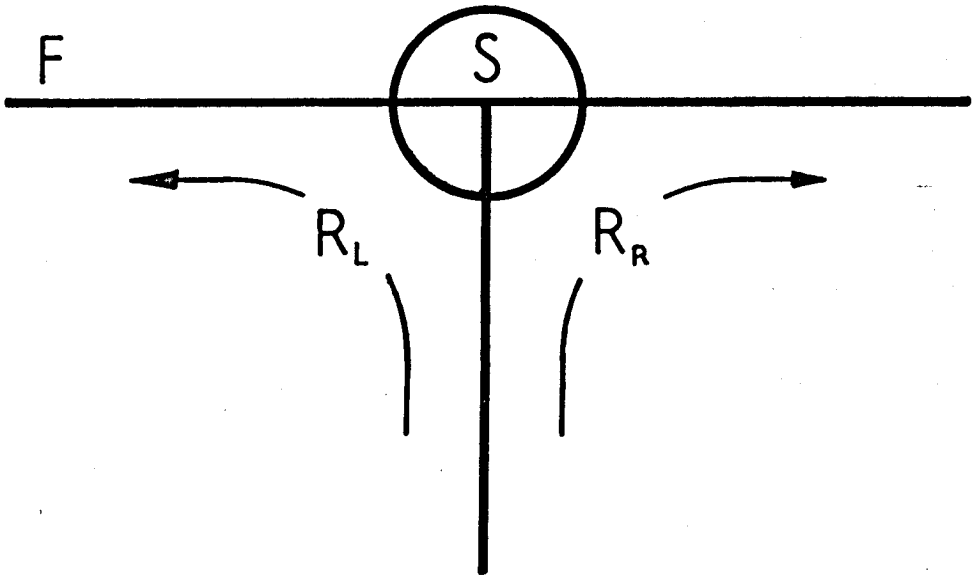


Fig.2.

animal proceeds along the right-hand alley, turns round at the end and traverses the length of the cross-piece of the maze, it will again find food. Again, a permanent increment of $S \longrightarrow R_R$ will be established. However, a longer time will have elapsed since the conjunction of stimulus and response took place and it is therefore supposed that the increment to $S \longrightarrow R_R$ will be less than that to $S \longrightarrow R_L$. On these grounds, simple instrumental learning, the elimination of errors, etc., can be explained. And this kind of account was the cone of the early views of workers like Thorndike and Hull *(ref. 26)*.

This is, of course, a "one-stage" view of learning and two features of any account of this kind should be noted here. First, for any kind of learning to occur there must be a goal-signal, and the rate of learning depends, inter-alia, upon the immediacy with which this is delivered. Second, there is no distinction here between learning and performance, in that if, after being trained in the way described, an animal finds itself again at the choice point, it will continue to turn left, if it is to do anything, even if it is no longer hungry. For what determines response at the choice point, and what, as it were, embodies the learning, are one and the same thing, namely the connection between S and R_L. Theories of this kind have, of course, been greatly elaborated so that some distinction

between learning and performance - the latter depending on motivation, incentives, and so on, can be made. But with these elaborations we need not be concerned in detail here, for the consequences I wish to discuss follow, for the most part, from the above basic assumptions.

(a) *Latent learning*

It would appear from these assumptions that, in the absence of a goal-signal, no learning should occur. Hence, if an animal is allowed to explore some environment, but is not rewarded, it should learn nothing during its wanderings. This matter is not as easy to test experimentally as it may, at first sight, seem. For there is considerable difficulty in deciding what may be rewarding, or constitute a goal-signal, for a rat. Certainly such rewards cannot be restricted to food, water, etc., and there is, for instance, reason to believe that being removed from the maze may be associated with a reduction of fear and hence may count as a reward. This kind of difficulty largely vitiates many of the early experiments on latent learning, i.e. learning without reward. Thus, it has been shown in several experiments *(refs. 2, 19)*, that if hungry rats are allowed to explore a maze without food reward, and such a reward is subsequently introduced at the end of the maze, they make a smaller number of errors in learning to run to this reward than do a control group of animals who have not previously explored the maze. It seems unlikely that the difference between the groups may be attributed solely to the fact that one group has become more "adjusted", in general, to the maze, and hence is less nervous and so learns more efficiently. Inspection of the animals behaviour during the pre-reward period, however, reveals that even at this stage of the experiment, a slight, but reliable, preference is detected for those paths which lead through the maze to the place at which they are removed and at which food will subsequently be placed. From this it follows that some slight reward must have been present at this stage of the experiment, which "confirmed" the appropriate S-R connections, and from this assumption the subsequent behaviour can be explained without too much difficulty. In this way, many experiments, which seem to show learning in the absence of any goal-signal, may be reconciled with the above kind of view.

It is possible, however, to demonstrate genuinely "latent" learning rather more convincingly. Several experiments are relevant here; two will be described for the purposes of illustration. An early experiment by Herb *(ref. 10)* followed the traditional procedure of latent learning experiments. One group of animals was trained to find food in a maze, while a second group was allowed to explore the maze without reward. The food was then introduced, and the performance of the two groups compared.

In this case, however, food was not placed at the end of the maze, but at the ends of the blind-alleys. The animals had, therefore, to learn to enter these in turn in order to obtain the reward. In this, as in preceding experiments, it was found that during the pre-reward period the animals exhibited a slight preference for that path through the maze which led to the end-box; that is, there was some tendency to eliminate the blind alleys. From this, as before, it follows that, in the case of this group of animals, the S-R connections associated with the through-path were being strengthened at the expense of those associated with the entry into blind-alleys. If, now, food is introduced into these blind alleys, these animals should, if anything, be at a disadvantage when compared with the group which has not had this preliminary exploratory experience. In fact, the converse result is found, namely that the animals who have been allowed to explore the maze need very few additional training trials, after the introduction of the reward, to reach the level of performance which the group without exploratory experience have taken many trials to develop. Hence, during this period of explanation these animals have learned something about those parts of the maze which they were tending to avoid, and this seems strongly to suggest learning in the absence of a goal-signal which functions in the manner prescribed by this kind of theory.

A similar difficulty arises in the case of "latent extinction". In this type of experiment *(ref. 20)*, animals are trained in a maze, until they are able to traverse it without error, in order to obtain a reward in the end goal-box. Now, according to the kind of theory under discussion, extinction of an S-R "connection" should occur only when the combination of S and R is no longer followed by a goal-signal. In these experiments, however, the animal is placed, after training, directly into the goal-box, which is now empty. They are restrained here for a short time and then removed. It is found that if, after several such trials, they are again tested by placing them once more at the beginning of the maze, they show much greater extinction of their previous training than does a control group of animals who have been similarly trained, but have not experienced the empty goal-box. In terms of S-R theory, this result presents some difficulty of explanation, for since the animals have never run the maze unrewarded, the conditions under which extinction should be found do not obtain. And although attempts have been made to extend the theory to cover this kind of finding, these modifications do not appear to be satisfactory.

It thus appears that there is definite evidence that learning may occur which is difficult to explain on the kind of theoretical view mentioned above. It seems that evidence of this kind tends to raise the following dilemma. If we retain an S→R, one-stage account of learning, we are faced with the difficulty that an observed preference in performance under one set of conditions may, or may not, be carried over to a new set of

conditions - e.g. when food is introduced in the non-preferred alleys. This kind of view has, however, the advantage that no problem is created about the extraction of information stored in the system. For, on this view, the mechanisms which learn and which dictate behaviour at the choice point are one and the same. If, on the other hand, we explain learning more in terms of a system which learns to seek a series of cues, then it may be argued that since at the end of each trial some goal-signal is attained, all the information, even about blind-alleys, gained during the trial will be retained. But, in this case, no account is given of how a system can select from this information in order to guide itself to that point at which the reward is placed. To this I shall return. But it may at this point be useful to mention a second source of experimental evidence concerning the dependence of learning on the receipt of some goal signal.

(b) *Irrelevant incentive learning*

The principle governing this form of investigation is that an animal is motivated to find some reward - e.g. water, at one or more places in its environment; and a second, irrelevant reward - e.g. food, which it is not seeking, is also present in this environment. Having been trained to find the first reward, the condition of motivation is changed so that the animal will seek the previously irrelevant reward, and it is tested to discover whether during the preceding learning period it has learned anything of the location of this reward. One group of investigations of this kind have been mentioned above.

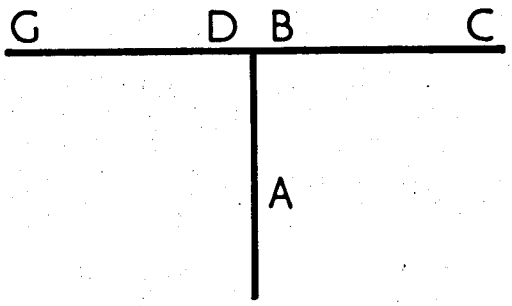
There is a considerable number of relevant experiments in this field. Here I shall mention only two in detail and then summarize some of the conditions which have been found to be important in determining to what extent learning of this kind may occur. Spence and Lippitt *(ref. 23)* trained thirsty rats in a single-choice maze. In one arm of this maze they found water and, in the other, they found food. The animals choices were forced, at the choice-point, by a system of changing blocks so that, during the training on any one day, the animals were forced to enter each alley equally often and in random order. Upon entering the arm containing water, they were allowed to drink for a short time. When they entered the food-arm, they were allowed to remain for a short time and were then removed. Since the animals were thirsty and not hungry, no food was eaten. Thus, only those trials to the water-side were followed by a goal-signal. Hence, if learning depends upon association with reward, we should expect no learning of the location of food; and, when tested hungry this prediction was sustained. The animals' choices on the first trial after the change of motivation were predominantly to the alley in which they had previously been rewarded when thirsty. In a second experiment, however, *(ref. 24)* evidence of irrelevant incentive learning was obtained. Again, the rats

were forced to enter each side of a single choice maze equally often. In this case, however, the animals were satiated and the incentive for their performance was return to cages containing other rats, placed at the end of each alley. Food and water, as irrelevant incentives, were placed in the arms of the maze. In these circumstances, when the rats were deprived either of food or of water, it was found that they tended to choose correctly - i.e. to choose that alley which contained the incentive relevant to their motivational condition.

Upon the basis of many such experiments it appears that some of the major conditions governing such learning are as follows: (1) The "relevant" motivation should not be too strong. Animals trained under 6 hours of water deprivation are more likely to learn the location of food than others trained under greater degrees of thirst. (2) The relevant and irrelevant incentives should be somewhat separated spatially. (3) If there is a major relevant incentive present at all in the situation, this should be found after each choice - that is, it should be found after those choices which lead to the alleys containing the irrelevant incentive as well as after those leading to the other alley. If these results are considered in conjunction with those derived from experiments on latent learning, it seems likely that we should conclude, first, that rats are capable of learning in the absence of any obvious goal-signal - at least in so far as these register the attainment of some object, such as food for a hungry animal, which is directly related to some primary state of biological need. And second, that if an animal is motivated in this way to any great degree, and is finding the object which satisfies this need, the extent to which he will learn about incidental features of his environment will be a function of the intensity of the prevailing drive and the extent to which these incidental features are relevant to the learning necessary for the attainment of the relevant incentive.

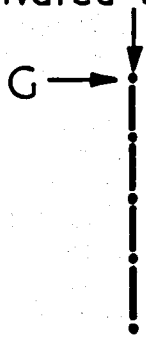
It is clear that such results, taken together, are incompatible with any "one-stage" S——>R connection account of the kind given above; and this is probably a difficulty inherent in this kind of theory. For, although various additions to, and modifications of, this theory have been suggested, all have broken down, in one way or another, in the face of the experimental evidence. If we suppose, however, that learning here consists, largely, not in the acquisition of S-R connections, but rather in the learning of a sequence of cues, then it seems that an explanation may be offered which is consistent with the experimental evidence. Perhaps the most developed theory of this kind is that proposed by Deutsch; and, without attempting to present this theory in detail some consideration of this general approach must now be given.

Deutsch proposes that, at least in the case of cue-learning, we should suppose that a mechanism is operating which stores not only the cues which



A B C D — enviromental cues
G — goal signal eg. taste of food

Activated by hunger



Initial state of system.

Suppose on the 1st. trial the system goes A into BC returns to DG.
State of system will then be thus:—

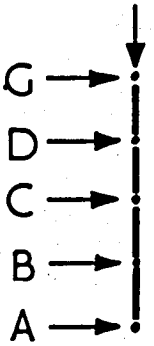


Fig.3.

an animal observes in moving about its environment, but also the order in which these cues are observed. Very roughly, it is proposed that we should regard the system as consisting of a series of chains of units, called links. Each chain is "activated", from one end, by the physiological factors associated with various states of motivation. Cues are attached to that chain, or chains, of links which are currently activated. A cue is first attached to the first, or highest, link in a chain which is free - that is which does not already have a cue attached to it. When a second cue follows the first, the first is moved one step down the chain and the second takes its place, and so on until either a goal-signal, which is to be regarded as permanently attached to one of the top links is received, or at least until a cue which was already connected to the chain is received. Thus in learning a simple T-maze, the state of affairs would be as in *fig.3*. This system will have acquired some information about the cues which lead to the goal. However, as mentioned above, there is always the difficulty, in the case of a cue-seeking device, of so using the stored information that errors will be eliminated. In this theory the difficulty is overcome, roughly speaking, by making two assumptions: (1) That the system will seek, or "approach", that cue which is lowest in an activated chain; and, (2), when a cue which is already connected to the chain, is received, that part of chain more remote from the goal-link then this cue is deprived of activation. Thus, in the case illustrated in *fig.3*, the system on its second trial will first seek and approach A; and at the choice-point, if it receives both B and D, that part of the chain below D will be switched off. Hence, it will not enter the blind-alley and approach B, but will rather proceed correctly to D and G. Hence the system will learn, and will display its learning.

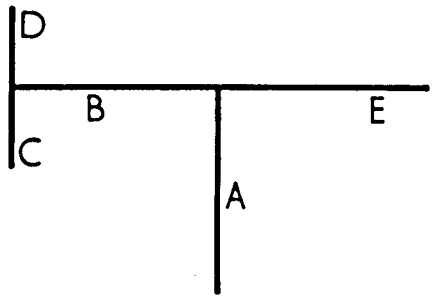
Now there seems to be little doubt that such a system can accommodate at least the majority of the results derived from experiments in ordinary learning situations. It should be noted that, on this account also, learning is dependent upon the reception of a goal-signal. For only the arrival of such a signal brings to an end the process of connecting the preceding cues progressively further down the chain and fixes the order permanently to the row of links. Since, however, this signal may be either a cue representing the attainment of the goal for which the animal was seeking, or any cue which has already been permanently attached to a chain, the kind of evidence quoted above presents no difficulty for the theory. For even if the maze is completely novel for the animal during its initial explanatory period, it must sooner or later encounter something familiar. And this will serve to preserve the preceding ordering of the cues as they were experienced in the maze. It is therefore difficult to see how to discover experimentally whether or not such a goal-signal, or substitute cue, is necessary in order that the performance of this group of animals

should be superior to that of a group without preliminary experience of the maze. It is, however, implied that on any trial during which learning occurs the animal should be seeking something - i.e. some chain, or chains, of links should be active. If it could be shown that animals who were satiated for all goals still explored a maze and, in so doing, learned something of the lay-out of the maze, this might bear against the theory. Unfortunately, the experimental evidence on this point, though considerable, is conflicting and it is as yet impossible to draw any definite conclusion.

We have then, in this theory, an example of an explanation of much of the experimental evidence on learning in rats, in terms of the learning of sequences of cues and requiring some form of goal-signal for such learning to take place. It seems to me, however, that it may be doubted whether such a goal-signal is necessary and, indeed, that if we take into account some of the modifications and extensions of the above system, the goal-signal comes to play a subsidiary role in the learning process.

The goal, or familiar signal in this theory is required primarily because the process is regarded in terms of the development, upon a fixed chain of links, of a fluctuating sequence of cues, and because the goal-signal itself is regarded as permanently attached to one of these links. It is necessary to preserve the order and to ensure that the cue observed immediately before the goal is connected to the link immediately succeeding that to which the goal is attached. In fact, in order to explain latent learning, it must be supposed that a series of such chains are established, with inter-connections between them through the mechanisms responsible for the reception of cues. This whole system must then be "fixed" - i.e. the cues must be permanently attached to the links they are then occupying. Fairly complex switching systems seem to be required to accomplish these functions and there must be great multiplication of the representation of cues. In the absence, however, of any knowledge of possible physiological identifications of the system, these should not be considered, at this stage, as important objections.

It does appear, however, that an alternative which does not depend upon fixed chains and a goal-signal, may be rather more simple. There seems no reason why a system should not be devised which stores information about the observed association of cues, either in terms of probabilities, or in some other form. Thus, suppose that, in *fig. 4*, the animal explores the maze and, in so doing, observes the cues in the sequence shown. On the basis of this information, the chance of E following A, B following E, and so on, is estimated. If, now, the animal is motivated, placed at D and rewarded, the chance of G following D is also estimated. When the system is now placed elsewhere in the maze, say at A, it combines the probabilities, for instance multiplicatively, of G following D, D following B and D following C, B following E and A, and so on. If, having observed A, it is faced with the choice of B or E, it will choose that cue to approach which is most strongly associated with, or has the greater probability of being followed by, the



Sequence of cues observed in preliminary exploration:—

AEBABCDBDBEB

The system will estimate the chance of

A followed by E
E followed by B
B followed by A
A followed by B
----- and so on.

And after feeding at D

D followed by G

Fig.4.

goal. It appears that a system of this kind would not require a goal-signal of any kind in order to preserve its previous learning. A goal-signal is required only, when it is motivated, to indicate to it which, of the environmental cues about which it has information, it should seek. This information is not stored on chains, and hence no "fixing" by a goal, or familiar signal, is required. The "chains" are rather created temporarily whenever the animal is motivated in an environment about which it has any information.

This is not the place to develop this view further. It is clear that this general suggestion is very close to that proposed in detail by Deutsch; in particular it takes from his theory the fundamental views that an animal

learns a sequence of cues to approach, and that, when it is set to find a goal, it selects from these cues, as targets to approach, working backwards from the goal. There are, however, differences in the way in which the information is recorded and extracted, of which the most important here is the role played in learning by a goal-signal.

I have tried, in this section, to present examples of experimental results concerning the dependence of learning upon the reception of a goal-signal; and it has been argued that these results are incompatible with a "one-stage" S-R theory, which, necessarily, requires such an indication of success. It appears that the evidence may more readily be explained in terms of a type of system which learns sequences of cues. And here it seems that a goal-signal may, or may not, be required, according to the design of the system.

III. CONCEPTS

Finally, and very briefly, I should like to raise a problem quite distinct from those discussed above. It has there been suggested that rats learn a sequence of cues, that they "approach" or "avoid" these according to their motivational requirements, and that a second stage of learning may be involved in developing those outputs which will result in approach and avoidance in various situations. It has been assumed throughout that there is no difficulty about the analysis of the "cues" to which the animal responds.

It is, of course, clear that we are very ignorant about the nature of the mechanisms which isolate these "cues". What kind of mechanism (if any) analyses, out of the visual input, the common element in all triangles, and so on? In such cases, however, it may be assumed that there are certain comparatively invariant aspects of the stimuli between which the animal must discriminate which, if they can be detected, will enable it to approach the correct cue. And various suggestions have been made about how this analysis might be done. Similar problems arise in connection with, for instance, relational discrimination. It is a matter of dispute whether, when rats are presented with figures of different sizes and are rewarded for choosing, say, the larger figure, they learn to choose a figure of that particular size, or whether they choose the larger figure. Evidence may be quoted in favour of both views. Spence has also shown that transposition - that is, choice between two new figures which is apparently based upon the size *relation* learned in the earlier training series - may in fact be explained upon the assumption that only learning to approach a given size, and avoid another, had taken place *(ref. 21)*. On the other hand, whatever may be true of rats, it is certainly the case that human adults can learn to respond upon the basis of the *relation* holding between the figures. It

is also clear that this cannot be explained on the basis of Spence's theory, since certain predictions from this theory are not held for this case, although they are confirmed by the results of certain animal experiments *(ref. 22)*. It has also been shown that very young children, when trained to choose the larger of two figures, cannot transpose correctly to two new figures when these are of very different sizes from those upon which they are trained. They cannot, when questioned, specify the relation upon which their response was based. At a slightly greater age, it is found that their performance changes so that they become able to transpose to such cases correctly and also to give, verbally, the relation upon which they are working *(ref. 14)*. In general, one has the impression that their behaviour changes, gradually, from being rather rat-like towards increasing similarity to adult abilities. And this appears to occur in conjunction with, rather than to be caused by, a development of their linguistic abilities.

One might arrange, then, three abilities in a series. We have first simple, absolute discrimination. There, for instance, we may have a rat, a child, or a man, trained to choose a black rather than a white card. Secondly, we have those cases in which some complex analysis is required of the input from each cue as, for instance, in shape discrimination. Thirdly, we have those cases in which it is required that response should be based not upon the properties of one stimulus alone, but rather upon the relations holding between the stimuli. It is clear that, after a certain age, humans can do this; it remains doubtful whether rats and very young children can, or whether, over a restricted range, they merely give the appearance of doing so. Now, this is a series in which the cue upon which the response is based becomes increasingly abstract, in the sense that the precise physical properties of the stimuli become progressively of less importance. It is possible to extend this series at least one stage further. It is certainly possible to train a human adult to choose, from a selection of stimuli, that one which differs from the rest. And, after they have discovered this relationship, they will be able to continue to select the "odd" one, whatever the nature of the group of stimuli with which they are faced. Here, the physical properties of the stimuli are of no importance whatever; all that is required is that there should be one stimulus differing from several other identical stimuli. Similarly, it will be possible to train adult subjects to select either of a pair of stimuli which differ from each other, and to reject both of a pair which are the same as each other, and vice-versa. So far as I am aware, it is not known how young children behave on these kinds of problem. But it is known that, after very considerable training, monkeys are able to solve them *(ref. 17)*. The evidence concerning the abilities of lower animals on this kind of task is rather conflicting. Success has been claimed for the rat and for the canary *(ref. 27)*; but, in attempting to train rats on the oddity problem,

using a rather different method, I could find no evidence whatever of any such ability, no matter how long the training was continued.

The inference which may be drawn from this selection of evidence is that there may be two kinds of "concepts". There is clearly a sense in which an animal which can discriminate between shapes has, for instance, a concept of triangularity. But this seems to be different from having the concept of "oddity" or "difference"; the latter are stimulus-neutral in a sense in which the former are not. It may also be the case that human beings and, to a lesser degree, monkeys can handle and work in terms of these abstract concepts, while lower animals cannot. Finally, it may be the case that the development of linguistic abilities may be dependent in part upon the presence of mechanisms which can work in terms of such abstract concepts. It is, however, difficult to see how language itself could be a necessary condition for these abilities. Now, whether we think in terms of two, or more, different kinds of concepts, or rather in terms of a scale of concepts of increasing degrees of abstraction, it seems likely that many of the activities in which man appears to be more intelligent than other animals depend upon the use of these very abstract concepts. It would be of great interest, therefore, to know what kind of mechanism could be devised which could be taught, for instance, to choose the odd stimulus from a group of stimuli, whatever the nature of the stimuli might be.

It has been my purpose, in this paper, to present three aspects of learning in animals which seem to me to be of importance if such behaviour is to be "imitated" mechanically. In the case of the rat, it has been suggested that we should regard learning as a two-stage process in which the animal learns, first, what cues to seek and, second, how to attain them. In the case of such "cue-learning", it has been argued that such learning can probably occur in the absence of a goal-signal of any kind. It is suggested that the mechanism records measures of the association between cues and that, a goal being set, it then determines which of the cues before it is most likely to lead to the goal, proceeds to this cue, repeats the procedure, and so on. This view is very similar to that proposed by Dr. J. A. Deutsch and it will be obvious that I have here drawn extensively upon his work. Finally, the question of the kinds of cues which organisms are able to use to direct their behaviour has been raised. It is suggested that the differences between men and rats in this respect may be responsible, to a very considerable degree, for the difference in their "intelligence".

It would, of course, be absurd to suppose that any views as general as these have been proved correct by experimental work with animals; it is the purpose of this paper only to raise and illustrate the kinds of problems of explanation which such work poses.

REFERENCES

1. ALEXANDER, M. F. M. R. Unpublished M.Sc. Thesis. University of Cambridge.
2. BLODGETT, H. C. The Effect of the Introduction of Reward upon the Maze Performance of Rats. *Univ. Calif. Publ. Psychol.*, 1929, **4**, 113. (Reported in e.g. Murn, Handbook of Psychological Research on the Rat.)
3. BUGELSKI, B. R. Extinction with and without sub-goal reinforcement. *J. Comp. Psychol.*, 1938, **26**, 121.
4. CARR, H. A., and WATSON, J. B. Orientation in the White Rat. *J. Comp. Neur. and Psychol.*, 1908, **18**, 27.
5. CRESPI, L. P. Quantitative variation of Incentive and Performance in the White Rat. *Am. J. Psychol.*, 1942, **55**, 467.
6. DEUTSCH, J. A. A new type of behaviour theory. *Brit. J. Psychol.*, 1953, **44**, 304.
7. DEUTSCH, J. A. A Theory of insight, reasoning and latent learning. *Brit. J. Psychol.*, 1956, **47**, 115.
8. EVANS, S. Flexibility of Established Habits. *J. Gen. Psychol.*, 1936, **14**, 177.
9. GLEITMAN, H. Place learning without Prior Performance. *J. Comp. & Physiol. Psychol.*, 1955, **48**, 77.
10. HERB, F. H. Latent learning - Non-Reward followed by Food in Blinds. *J. Comp. Psychol.*, 1940, **29**, 247.
11. HUGHES, B., and SCHLOSBERG, H. Conditioning in the White Rat. IV. The Conditioned Lid Reflex. *J. exp. Psychol.*, 1938, 23, 641.
12. HULL, C. L. Principles of behaviour: An introduction to behaviour theory. *New York: Appleton-Century Crafts.* (1943).
13. HULL, C. L. A behaviour system. *New Haven: Yale University Press.* (1952).
14. KUENNE, M. R. Experimental investigation of the relation of language to transposition behaviour in young children. *J. exp. Psychol.*, 1946, **36**, 471.
15. LASHLEY, K. S., and McCARTHY, D. A. The Survival of the Maze Habit After Cerebellum Injuries. *J. Comp. Psychol.*, 1926, **6**, 423.
16. McALLISTER, W. The spatial relation of irrelevant to relevant goal objects as a factor in simple selective learning. *J. Comp. & Physiol. Psychol.*, 1952, **45**, 531.
17. MOON, L. E. and HARLOW, H. F. Analysis of Oddity Learning by Rhesus Monkeys. *J. Comp. & Physiol. Psychol.*, 1955, **48**, 188.
18. PAVLOV, I. P. Conditioned Reflexes. *London: Oxford Univ. Press.* (1927).
19. REYNOLDS, B. A repetition of the Blodgett Experiment on Latent Learning. *J. exp. Psychol.*, 1945, **35**, 504.
20. SEWARD, J. P., and LEVY, N. Sign Learning as a factor in extinction. *J. exp. Psychol.*, 1949, **39**, 660.
21. SPENCE, K. W. The differential Response in Animals to Stimuli Varying within a Single Dimension. *Psychol. Rev.*, 1937, **46**, 88.

22. SPENCE, K. W. Failure of transposition in size discrimination of chimpanzees. *Am. J. Psychol.*, 1941, **54**, 223.
23. SPENCE, K. W., and LIPPITT, R. O. An experimental test of the sign - Gestalt Theory of trial and error learning. *J. exp. Psychol.*, 1946, **36**, 491.
24. SPENCE, K. W., BERGMANN, G., and LIPPITT, R. O. A study of simple learning under irrelevant motivational reward conditions. *J. exp. Psychol.*, 1950, **40**, 539.
25. THISTLETHWAITE, D. Conditions of irrelevant incentive learning. *J. Comp. & Physiol. Psychol.*, 1952, **45**, 517.
26. THORNDIKE, E. L. Animal intelligence: Experimental studies. *New York: Macmillan.* (1911).
27. WODINSKY, J., and BITTERMAN, M. E. The solution of Oddity Problems by the Rat. *Am. J. Psychol.*, 1953, **66**, 137.

DISCUSSION ON THE PAPER BY MR. A. J. WATSON

DR. E. R. F. CROSSMAN: I have found Mr. Watson's paper quite stimulating; I would like first to clear up two theoretical points and then to pose a question of how to relate experiment to theory, on which Mr. Watson might be able to help me.

Mr. Watson distinguishes between one- and two-stage learning, the former apparently being "stimulus-stimulus" and the latter "stimulus-response" learning. Now surely the former could be characterised as association, remembering or even conditional-probability computing, and would it not be better to reserve the term "learning" for the latter case where an observable change of behaviour occurs?

Secondly, in the latter or response-learning case, I cannot see how goal-signals can be eliminated. The selection of responses for greater adaptiveness requires information, which must come either from the environment or from within the animal, as to how adaptive a given response is. In other words something must signal the degree of approach to the goal.

My own experimental question concerns human learning of manual skills, as for instance learning to make cigars (Crossman, 1959). To begin with one cycle takes about 10 sec. on average, and thereafter the time comes down gradually to, say, 4 sec. Why does this happen? Closer observation shows great variability of cycle-time, which diminishes with practice, and it is fairly clear that what is happening is that the quicker methods of doing the job are being selected for more frequent performance at the expense of the slower ones. Learners do not get any better at the slower methods, they just do the quicker ones more often. I would like to ask Mr. Watson what kind of learning mechanism he thinks might give this result? I have thought of three possible answers. First there might be direct or secondary reinforcement from a primary drive; this seems difficult to believe. Secondly, the reinforcement might derive from the lesser physiological cost of quicker methods - the so-called "principle of least effort". Thirdly, one might suggest that the effect stems from the decay of short-term memory, for repeating an action involves remembering it. If the memory of what has been done were somehow "fixed" by the successful completion of a cycle, the decay during the cycle would mean that methods taking longer would be less well remembered and less likely to be repeated.

The first two of these postulate specifically learning mechanisms, whereas the third suggests a purely "mechanical" effect. Which does Mr. Watson think would be preferable to explore first?

MR. J. MORTON: There is a point in animal learning which corresponds to the problem Dr. Crossman has just raised. If you train a rat in an alternative pathway maze, it learns to choose the shorter path, although on Mr. Watson's original theory the conditional probability of reaching the goal is the same, unity, by either pathway. Therefore there must be some weighting of the probabilities. As with the human subjects this weighting could be by means of time or effort involved, with apparently the same result. Grice (1942, *ref. 3*) did this experiment varying the relative, not the absolute, differences between the paths, and his data suggested that in maze learning, at least, the effect of a given reinforcement on a response is a function of the amount of activity, -which is related to the distance - on the part of the animals, intervening between response and reinforcement, rather than the temporal delay per se. Other experiments, (e.g. Anderson 1933, *ref. 1*) with delayed reward, indicate that time can be used directly in learning, so it would seem that there is no single and general answer to the problem posed about possible mechanism for selecting alternative responses.

With reference to Mr. Watson's theoretical suggestions, I would like to know what explanation he puts forward for two experiments. Firstly one by Spence and Shipley (1934, *ref. 5*), which indicates that with a partly learned maze, the location of the goal has an effect such that blind alleys leading in its direction are much more difficult to eliminate than those leading in the other direction. The explanation normally given is that the animal is anticipating the turn; but the concept of 'a turn' does not seem to fit in particularly well with the computation of conditional-probabilities to the goal in stimulus - stimulus theory.

The second experiment is by Dashiel (1930, *ref. 2*) and Muhlhan and Stone (1949, *ref. 4*) using a checker-board maze with many equally correct paths to the goal. In the process of learning the rat selects preferentially two different types of passage, firstly the L-shaped routes, with a minimum of turns, and finally a zig-zag route with a maximum of turns. The paths are clearly the same length and one might imagine that the minimum-turn route involved less effort and enabled more speed. (I have seen no analysis of times for the experiment.) However, the animal finally chooses a path which stays closest to the direct route, again seemingly

REFERENCES

1. ANDERSON, A.C.: J. Comp. Ps., 1933, 13, 27.
2. DASHIELL, J. F. (1930): Comp. Ps. Monograph. No. 32 (1930). (Reported in Woodworth and Schlosberg Experimental Psychology, 1954)
3. GRICE, G. R.: J. Exp. Ps., 1942, 30, 475.
4. MUHLHAN, G. J., STONE, C. P.: J. Comp. Phys. Ps., 1949, 42, 17.
5. SPENCE, K. W., SHIPLEY, W.C.: (1934) J. Comp. Ps., 1934, 13, 423.

indicating a concept of direction, which does not fit in with a stimulus-stimulus theory.

The explanation is perhaps quite simply that the animal has available a set of different learning mechanisms available possibly arrayed in a hierarchy, and which switch in to control the animal's behaviour perhaps on the basis of simplicity of mechanism initially, and then in relation to the relative success of these mechanisms in similar situations.

DR. A. M. UTTLEY: Mr. Watson has spoken about stimulus-stimulus associations and stimulus-response associations. One should now speak of association lattices rather than something as simple as a paired association. In the very simple machine I have built to imitate trial-and-error learning, there is a key association between a double conjunction and a triple conjunction, between the stimulus-consequence conjunction and the stimulus-response-consequence conjunction. Watson himself has said that an association which fits facts better is that between stimulus-response and stimulus-response-consequence. Is this what he means by a two-stage association or does that phrase refer to an intermediate goal?

DR. R. L SHUEY: I should say that I am not a psychologist and will probably make some comments that are ridiculous to psychologists. In reading the paper, I got confused as to how one separates goals from motivation. For example, in many experiments, the analysis depends heavily on what the experimenter assumes to be the goals of the animal; when in fact, the animal's goals may be entirely different. I believe it is well known that monkeys will solve problems for no good reason. One might ask why he is doing it if he has no goal, but I am sure he has motivation.

Let me illustrate this with another example. I happen to work in a place where there are both long and short paths by which I can reach my home. I take the longer path, because I like the scenery. I do not know what a psychologist, observing me from above, would conclude from my behaviour. How can you separate it from my goals and my motivation?

DR. N. S. SUTHERLAND: I should like to ask Mr. Watson a question. I think it is really the question Dr. Uttley asked. Perhaps I could put it in a slightly different way. I am not clear why Mr. Watson wants two things - stimulus-stimulus connections and stimulus-response connections - because it seems to me that what you are doing with your system would be done on a system which computed, given a stimulus, the probability of some other stimulus for a given response. So what you would be learning is a series of connections. An animal does not simply learn to give a response to a stimulus, nor does it learn that stimulus connection; it learns that if

it makes a given response it will be followed by that stimulus. This may not be different from what you are saying - I am not sure.

MR. A. J. WATSON (in reply): These are rather diverse points. I shall discuss them in the order of their subject matter, rather than in that in which they were raised.

First, a verbal question. Dr. Crossman suggests that 'association', or some similar term should be employed to name what I have called stimulus-stimulus *learning*, and that the use of 'learning' should be restricted to name what I have called stimulus-response learning. The ground for this suggestion is that 'learning' is properly something which involves an observable change in behaviour, and stimulus-stimulus association may not involve such a change. I do not wish to argue about the word. I have used 'learning' to refer to any process whereby information is stored. However, it should be remembered that stimulus-stimulus association *may* involve changes of behaviour - if it did not do so sometimes we could never detect that it had occurred; and stimulus-response association may, equally, on some occasions not involve changes of behaviour - if, for instance, the motivational condition of the animal is not such as to lead it to seek for the goal which the response in question is instrumental in achieving. It seems to me, therefore, that the two kinds of association here have equal claims on the name 'learning'.

Second, Dr. Uttley points out that one should speak of association 'lattices' rather than of simple paired associations. This is certainly correct, and represents a considerable improvement upon some earlier theories of learning. I have here discussed the limiting case of pairs of associated items for the sake of simplicity. Dr. Uttley goes on to ask what is meant by 'two-stage' association. Is this something different from the system embodied in his learning machine, or not? My belief is that it is something different, in the following sense. Dr. Uttley's system will, of course, compute measures of association between stimuli which are presented to it in suitable temporal relations. It will also - or would with certain modifications which he has mentioned elsewhere, - compute a measure of association between a goal, or 'consequence', and a stimulus-response conjunction. Now my point is that it seems to me that the facts of animal learning and performance require for their explanation that even very short sequences of behaviour must be split up into a sequence of units. In each of these units the animal is trying to reach some stimulus or cue. There are, therefore, two sorts of goal. First, there is that goal which the whole sequence of behaviour is directed to achieve - e.g. food - and this goal dictates what intermediate cues the system shall seek. Second, these intermediate cues are goals for response at given moments during the sequence. Hence my 'two-stage' association is based on the view that actual

behaviour almost always involves what Dr. Uttley calls 'intermediate goals'. Thus, in order to imitate animal learning we require two separate systems, though both may be of the general kind suggested by Dr. Uttley. The first must select cues to seek, on the basis of stored information about the associations between them and with the final goal; and the second must learn associations between a given stimulus - the momentary 'intermediate goal' - and a given preceding stimulus - response conjunction. These must, however, be distinct stages in the system. I do not see how they both can be achieved by a single mechanism, though Dr. Uttley's mechanism may perform *either* the one *or* the other.

This point is closely related to that raised in Dr. Sutherland's question. I think it is suggested here that a 'two-stage' system is unnecessary, in that all that is required is a mechanism which, having been set stimulus A as a goal, computes the probability of achieving this goal given that stimulus B and response R are present. If this probability is sufficiently high, and if B is in fact present, then R will occur. This is substantially what Dr. Uttley's system will do and an amalgamation of my two stages is thus obtained. I have tried to suggest that there are reasons for thinking that this unitary mechanism will not provide an adequate imitation of animal behaviour. These reasons, if not compelling, seem to me persuasive. One of them is that if response R cannot in fact be made, or if no particular R has been learned to effect the transition from A to B, still it appears that the animal does not have to learn afresh what to do when it is faced with A and is seeking B. Performance may sometimes not be highly efficient, but it is often better than chance expectation.

This brings me to an ambiguity in the notion of a 'goal' which is brought to light by Dr. Crossman's point that stimulus-response learning *must* involve a goal-signal. For, otherwise, how could selection of response for greater efficiency take place? My point here is that we must distinguish between two functions of a given input into the system. The first is purely 'informational' in that the system simply records the fact that this input has occurred. The second function arises when this input is one which the system has been set to seek, either as an 'intermediate', or as a final, goal. The question is, under what conditions does the system store permanently the arrival of this input and compute its' association with other inputs? It may be that such permanent storage and computation arises *whenever* the input arrives, or it may be that this only happens if the system is actually seeking this input - i.e. when it functions as a goal, or 'success' signal. I tend to the view that the first answer is correct, in the case of stimulus-stimulus learning. I should hesitate to express an opinion about stimulus-response learning at this time. But it seems *possible* that in this case also the consequences of making a response

are permanently stored even if these consequences are not 'goals' for the system at the time.

Dr. Shuey has raised another point in connection with this discussion about the function of 'goals' in learning. It is, he points out, all very well to speculate at length about the nature of the system which exhibits goal seeking behaviour, but he doubts whether, in any particular case of animal behaviour, we can specify what these goals are with sufficient accuracy. What is the goal for a monkey who solves problems apparently for no other reward than that obtained by reaching the solution? And what is the nature of the goal-input into Dr. Shuey when, in defiance of the principle of least effort, he takes the longer, but more pleasant, walk home? I should like to be the last to claim that it was always easy to specify the goals for any sequence of behaviour; and sometimes it may be quite beyond the range of practical possibilities. There are two points I should like to make here. First, there are some cases, usually in experimental settings, where it *is* fairly obvious what the goal for a sequence of behaviour is. It seems perfectly clear that the goal of a hungry rat, in familiar surroundings, is something closely associated with food. Whether it is the *taste* of the food, or some other consequence of ingestion, is another question which is at the moment receiving a good deal of attention. But the point here is that there are at least some cases where we can say with a good measure of accuracy, and little doubt, what the goal in question is. Second, not only are there many cases in which we are in fact in great doubt about what the goal governing behaviour is, there are also cases in which it may be doubted whether the goal-setting mechanisms are of the same *kind* as those involved in food-seeking and similar activities. I suspect that Dr. Shuey's examples are of this latter kind. We might, for the moment, group all this latter kind of behaviour together and call it exploration, although further analysis will probably show that we have several distinct kinds of activity here. The question then becomes, what is the nature of goal-setting and recording mechanisms in the case of exploratory behaviour, and how do they interact with the rest of the system to produce the observed behaviour? This is a large and most interesting question, and cannot be discussed now. I should like to say only that it is not clear to me that the goal-setting mechanisms involved in exploratory behaviour can be of the same kind as those involved in other types of directed behaviour. It may rather be that we have here the operation of a rather different kind of mechanism which comes into play when the system is not set to find one of the 'primary' goals. It seems to me that the further analysis of this matter is one of the most important tasks in the study of behaviour.

I turn now to the points raised by Mr. Morton. He argues that my insistence upon the stimulus-stimulus nature of much animal learning, and

upon the function of intermediate goals, fails to give an accurate picture of this behaviour. He gives three experimental examples which one might have expected a theory of this kind to explain, but which, he says, it in fact does not explain.

The first point is that animals will learn to take the shorter of two alternative paths to food. I do not agree that this is not to be predicted in terms of the kind of theory I have been discussing. For, if we suppose that there are more stimulus 'steps' in the longer than in the shorter path, then, at the choice point, the probability of reaching the goal will be greater if the first stimulus down the shorter path is approached than if the first stimulus down the longer path is approached. This prediction breaks down under two conditions. First if the animal has already been trained on the longer path to such an extent that the probability of reaching the goal, given that the first stimulus of this path has been observed, is unity. Second, if the time required to reach the goal by either route is so short that no intermediate goal is required and only one stimulus 'step' is involved. I do not think that there is sufficient evidence to warrant a statement about what animals actually do in these circumstances. It is, however, very unfortunate that Grice found that the ability to choose the shorter path was related to actual physical *distances* involved, but not to the time required to traverse them. As Mr. Morton points out, it has been shown in other experiments that animals *will*, in other situations, choose a path which is shorter in terms of time; and I should like to see further experiments of the kind carried out by Grice before developing hypotheses about weighting. But I agree that this is a reasonable complication which we may be forced to admit.

It seems to me likely that the two further experiments which Mr. Morton mentions should be interpreted in terms of the use of long-distance cues. I should suppose that animals would always prefer long-distance cues or intermediate goals, since this will involve fewer stimulus 'steps' before reaching the final goal. However, quite apart from such factors as sensory acuity, etc., the use of such cues will be limited by the kind of response which the animal must make to reach them. If this is extremely complex, it may be more efficient for the system to adopt a larger number of intermediate goals, each requiring simpler response learning for its attainment. Thus, in Spence and Shipleys' experiment, I should suppose that there was a long-distance cue in the region of the food. Here the response is comparatively simple - a right turn - and hence there will be a tendency to anticipate and make this response too soon. In Dashiell mazes, on the other hand, the use of a long-distance cue in the region of the goal requires the learning of a long and complex response sequence. Hence it seems reasonable that initially the animals adopt a route with more stimulus 'steps' in it, but which requires less complex response

learning. With further training they may be expected to use the more direct route which the adoption of the long-distance cue permits.

Finally, there remains Dr. Crossman's question about the cigar manufacturers. Why and how do they manage to improve their efficiency? There is little doubt that rats represent a much less complicated system than do people, even when the latter are engaged in repetitive tasks. I am therefore extremely dubious about the value of any inferences concerning the explanation of human behaviour which are derived from theoretical considerations arising out of the study of animals. For what it is worth, I suggest that we distinguish the 'why' and the 'how' questions involved here. The first question is *why* (in the sense of motivation) the subjects are seeking to create cigars as fast as possible. The second question is, given that they are set to achieve the goal of rapid cigar manufacture, what is the nature of the system such that faster methods can be selected and slower ones omitted. In making this suggestion, I am implying that Dr. Crossman's third suggestion is not correct. For it seems to me unlikely that this more efficient behaviour would develop unless the rapid production was a goal which the system was seeking. Hence I doubt whether any 'mechanical' explanation of the kind Dr. Crossman suggests is correct. In very general terms, I should be inclined to think that Dr. Crossman's first suggestion, that some direct reinforcement is involved, should be explored in answer to the first, or 'why', question. He finds this explanation difficult to believe. But I think it seems implausible only if it is taken to provide an answer both to the 'why' and to the 'how' question, and if the reinforcement is supposed to be primary. I agree that the production of cigars in 4 secs. is unlikely to be a 'final' goal which human beings are set to seek. Clearly, it will be an intermediate goal. But what the final goal is, the attainment of which is furthered by attainment of this intermediate goal, is a question involving motivational considerations of a kind which render inference from animal to human behaviour especially suspect. Direct investigation would be required. In answering the 'how' question, the only suggestion I can make is that the behaviour will be so organised that the smallest number of reaction times will be required – that is, the number of intermediate goals will be reduced to a minimum.

These are not very satisfactory answers, and they reflect my uncertainty about the whole matter. But I am afraid that this is the best I can do at the moment.

SESSION 4A

PAPER 7

INFORMATION, REDUNDANCY AND DECAY OF THE MEMORY TRACE

by

DR. JOHN BROWN

BIOGRAPHICAL NOTE

John Brown is an experimental psychologist and obtained both his first degree and his Ph.D. at Cambridge University. In 1953 he went as a lecturer in psychology to Bristol University and in 1954 to his present post at Birkbeck College, University of London. He is interested in the development of quantitative approaches to psychological problems, with particular reference to memory. He is the Honorary Secretary of the Experimental Psychology Group.

INFORMATION, REDUNDANCY AND DECAY OF THE MEMORY TRACE

by

DR. JOHN BROWN

SUMMARY

A new form of an old theory of memory will be elaborated. The essence of the theory is that the memory trace is subject to decay (decrease in the signal-to-noise ratio) but that the effect of decay depends (1) on the amount of initial redundancy in the trace, (2) on the coding of information in the trace, (3) on the information available from other traces. With appropriate assumptions, the theory seems able to provide possible explanations of many of the rather complex facts of human learning and forgetting. The paper will include a discussion of the nature of the memory span.

COMMUNICATION theory, as developed by engineers, has one obvious contribution to make to the study of memory. This is to supply a measure of amount of information which has considerable theoretical advantages. For the measure provides a rational basis for comparing memory for different types of material and it also takes account of errors in remembering in a logically satisfying way. It has indeed already been used in a number of memory studies. But communication theory may have a further contribution to make. It may supply the concepts in terms of which a fruitful theory of memory can be formulated. (Hints to this effect are given by Hick, 1955, *ref.* 9). The concepts I have in mind are concepts such as noise, signal-to-noise ratio, redundancy, channel capacity and, of course, the concept of information itself. These concepts are of especial importance to a theory of memory built round the idea of a memory trace subject to decay and it is such a theory I wish to propose.

The idea of a memory trace subject to decay has traditionally been regarded by psychologists as barren and unprofitable. It has seemed to throw little or no light on the complex facts of learning and forgetting. Some of the salient facts are (1) The time course of forgetting is not

fixed but variable. Thus it is more rapid if further similar material is learned in the interval before recall, a phenomenon known as 'retroactive inhibition'. It is also affected by whether learning trials are spaced or massed. Moreover, the passage of time sometimes results not in further forgetting but in the *recovery* of recall. (2) We not only forget, we also frequently distort, quite unwittingly, what we think we remember. It has not been clear how simple trace decay could lead to an unambiguous output which does not correspond with the input. (3) In learning, items at the beginning and end of a sequence are learned more quickly than items in the middle - 'the serial position effect'. (4) In general, less time needs to be spent on learning if practice is given at intervals rather than continuously, i.e. spaced learning tends to be more economical than massed. The most widely accepted interpretation of these phenomena is in terms of various kinds of inhibition which are supposed to be generated and to dissipate in a lawful manner during learning, retention and recall: a good summary of this approach is given by Osgood, 1953 *(ref. 14)*.

Although psychologists in general have been reluctant to use the idea of trace-decay to explain forgetting, those who have specifically considered memory over short intervals ('immediate' memory) have usually been more willing to postulate a decay process. Forgetting of what is learned from a single presentation of information can be extremely rapid. A delay before recall of only 4 seconds can reduce recall from 93% to 41% provided rehearsal of the information during the delay period is prevented (Brown, 1958 *ref. 5*). Rehearsal is prevented by presenting a rapid sequence of stimuli to which the subject is required to make immediate responses. Incidentally, the nature of the activity used to prevent rehearsal does not seem to influence the outcome (Brown, 1955, 1958 *refs. 4, 5*). This makes it reasonable to believe that the activity merely prevents rehearsal and that it does not directly interfere with the memory trace. The rapidity of forgetting in immediate memory is also shown in the fact that significant forgetting can be shown to occur even during the period of recall itself (Brown 1954, *ref. 3* and Conrad and Hille, 1958, *ref. 7*). Such facts (see also Broadbent, 1957 *ref. 2*) strongly suggest that decay of the memory trace is a plausible explanation of forgetting in immediate memory. If this is so, it is natural to consider whether all or most forgetting is not also due to this factor.

OUTLINE OF THE THEORY

The theory I want to propose runs as follows: It is assumed that when a memory trace is established, it usually has some internal redundancy. In other words, the trace is established with more features than are necessary to represent the information which the trace is required to store.

The trace is assumed to decay, rapidly at first and then more slowly. Some decay is compatible with reliable recall, owing to the initial redundancy. The critical amount is not fixed. It depends on the amount of initial redundancy and on the coding, which determines the form the redundancy takes. It also depends on whether other traces can supply relevant information (i.e. on redundancy in the trace system as a whole) and on any information supplied from the environment (i.e. on the richness of cues to recall). If the initial redundancy is sufficiently high, or if redundancy is increased by further learning trials, then the trace will still be adequate when decay has become slow. Long-term retention is then possible. Finally, it is assumed that the amount of initial redundancy is limited by the amount of storage space available when the trace is established.

CONCEPTS

1. *Storage space.* Any modifiable feature of a neurone anywhere in the brain forms part of the brain's storage space (provided the functional arrangement is such that it can in fact be used for information storage). Thus the word 'space' here is used in a generalized sense and does not imply that there is a localized area in the brain in which information is stored. Storage space may be either *general* or *specific.* It is general if the brain can use it for the storage of any kind of information: it is specific if it can be used for the storage of only one kind of information. Limited specificity or limited generality is also possible. For example, certain storage space might be specific to visual information but available for the storage of all kinds of visual information.

2. *The memory trace.* The theory is not concerned with the memory trace as a physiological entity but only with its functional properties. The expression 'memory trace' will refer to that portion of the storage space which is used to record a particular event or message. Often it will be convenient to speak as though a single memory trace corresponds to each event. But since decay of the trace will be postulated, the time at which each part or aspect of an event is recorded may be significant, especially if the time interval before recall is short. Thus, when it is necessary for the purposes of a particular discussion, an event (for example, a spoken sentence) which is not recorded simultaneously will be considered to establish a *series* of memory traces.

3. *Capacity of the trace.* The storage space occupied by the trace will have some definite channel capacity which is the capacity of the trace. If the trace decays, the capacity will not be fixed but will depend on the time for which it is used to store information. If it is used to store information for one second its capacity might be 100 bits: whereas its capacity if it is used to store information for one hour might be only 1 bit.

4. *Residual information in the trace.* By this will be meant the amount of the original information which can still be extracted from the trace at any stage in its decay. Under suitable conditions, it can be calculated from Shannon's formula for transmission over a noisy channel (Shannon, 1949, *ref. 16*).

5. *Decay of the trace.* By decay of the trace is meant a fall in the signal-to-noise ratio, i.e. a fall in the correlation between the initial and final states of the trace due to random disturbances. This produces a fall in the capacity of the trace and it is this consequent fall in capacity with which we shall be concerned. It should be pointed out that change, as such, in the trace does not necessarily involve decay in the defined sense. For example, if a trace starts as a neural reverberation and ends as a pattern of synaptic changes, the initial state may be uniquely deducible from the final state in which case no decay will have taken place.

6. *Redundancy in the trace.* When the trace is formed, more storage space may be used than is necessary for the immediate representation of the information to be stored. The trace then has redundancy. The absolute redundancy is the difference between the residual information in the trace and the information capacity of the trace: the relative redundancy is this difference divided by the information capacity. A trace which starts life with redundancy may, for a time, withstand decay without loss of information: redundancy will be lost instead (see later). The amount of redundancy in the trace does not necessarily depend on the amount of redundancy in the input to the sense organs. A redundant input might be recorded so as to eliminate the redundancy (indeed Attneave, 1954, *(ref. 1)* has suggested that the function of the perceptual machinery is to strip away redundancy from the input) and a non-redundant input might well be coded redundantly into a memory trace.

7. *Trace system redundancy.* Since we do not live in a random world, the events stored in different traces will be related to one another. As a result, different traces will often carry in part of the same information so that the trace system *as such* contains redundancy (i.e. independently of redundancy within individual traces). The existence of this redundancy means that the trace system can often supply some of the information lost through decay of an individual trace: recall of the event which this trace represents will then be facilitated. An analogy will help to show this. Suppose the word AND is written in chalk on a blackboard and the letter N is progressively smudged. Some smudging of the N will leave it legible: it starts with internal redundancy so that some 'decay' does not matter. Further smudging, however, will make it difficult to read unless attention is paid to the other two letters of the word. These letters help us to guess the smudged letter and this is because AND is a probable letter combination, i.e. because the letters carry common information. In this

analogy, the word AND is the analogue of the trace system and the individual letters are analogues of individual traces. It should be stated that not all trace system redundancy is useful in this way but only redundancy due to different traces carrying common information (a similar point applies to redundancy within the individual trace).

8. *Coding.* The coding concerns how the information is represented in the trace and how much storage space it occupies. Consequently it determines both the amount of redundancy and how it is utilized. It is these two factors which determine for how long the trace can withstand decay without loss of information. It is desirable to distinguish two limits to coding efficiency. Shannon has shown that it is theoretically possible to code so that the information passed through a channel is negligibly below the channel capacity (Shannon, 1949, *ref. 16*). In practice, however, constraints may exist - such as not coding over infinitely long sequences - which are such that even the best possible code falls appreciably short of the ideal. The best attainable code will be called an optimal code.

Some of the concepts will now be illustrated. Consider a 2-state storage element. The rate of transmission, R, will be maximized when the two states are used with equal probability. If the element is 'noiseless', then the capacity will be 1 bit. If there is a probability, p, independent of the initial state of the element, that after time t it will have changed its state, then the maximum value of R will depend on the value of p. Curve A in *fig. 1* shows the actual form of this dependence: it is calculated by application of Shannon's formula for information transmitted. The value of R so calculated can be regarded as the capacity of the storage element. Suppose now that two such elements are available for the storage of each 'bit' of information. Curve B in *fig. 1* shows how the capacity of the pair of elements falls with p (p is assumed to have the same value for each element and the two elements are assumed to be independent): curve B is simply curve A doubled at each point. *Figure 1* shows that, with ideal coding, no loss of information would occur until p became greater than 0.1. In practice, however, an ideal code is impossible. If 00 signifies that the two elements are both in a certain state and 11 signifies that they are in the opposite state, then to use 00 and 11 as the two possibilities provides an optimal way of coding one bit of information so as to give protection against noise. Curve C shows the value of R when such optimal coding is adopted. It will be seen that it falls appreciably short of ideal coding.

ASSUMPTIONS

1. *Decay of the trace.* It is assumed that there is decay in the capacity of the trace, due to an increase in purely random noise, and that this decay is at first rapid and then slow. For example, the decay function might

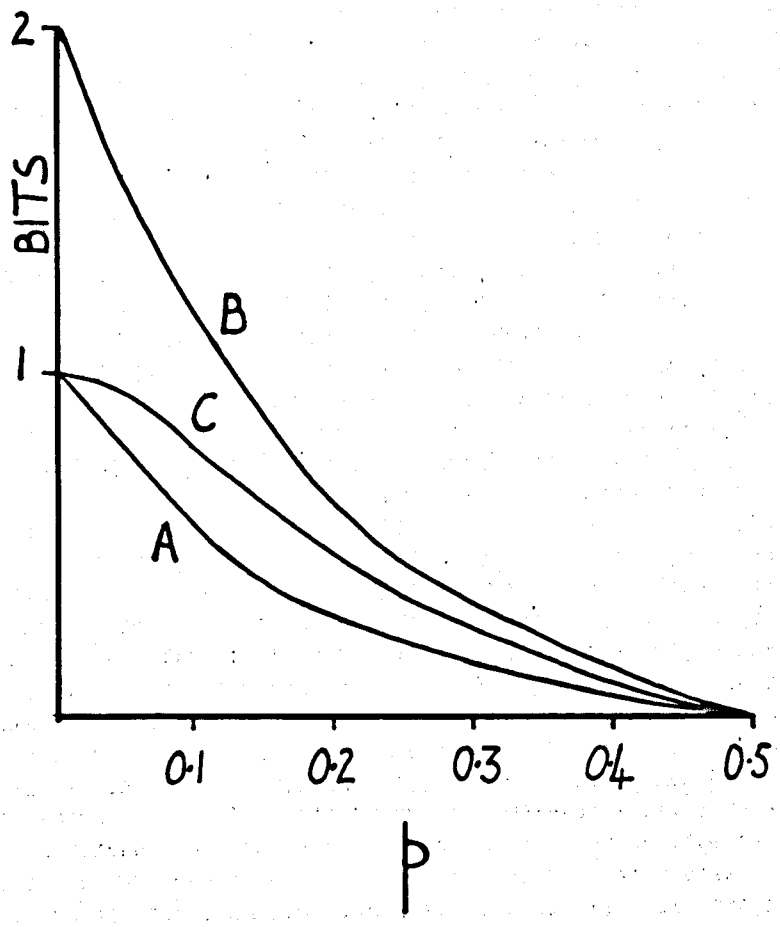


Fig 1. Relations between capacity, error probability and information transmission with optimal coding - see text.

prove to consist of two differently weighted exponential decays, the one initially dominant with a short time constant and the other with a long time constant. The exact function will have to be inferred from behavioural studies, unless physiology produces some direct evidence. No single study will suffice to show what it is since the time course of decay is only indirectly related to the time course of forgetting: this is discussed later.

2. *Availability of storage space.* Decay of the trace would not lead to forgetting if everything were recorded with sufficient redundancy. But redundancy uses up storage space and, since the brain is of limited size, there must be constraints on the amount of storage space that is used at any one time. One possibility would be that all information is recorded with a

small constant amount of redundancy. This would mean, however, that important information was forgotten as quickly as unimportant information. It is more likely that the amount of redundancy can be varied but that there is an overall limit on the amount of storage space which can be used at any one time. This would lead to competition both between simultaneous messages and between successive messages not separated by an interval of rest. Two rather different possibilities have to be considered. One is that there is a limited amount of general storage space available at any one time. The other is that there is a limited amount of specific storage space available at any one time. Provided this space were not fully specific, competition could still occur but only between messages drawn from the same subset of the total ensemble of possible messages - such, perhaps, as the subset consisting of all possible visual messages.

3. *Storage of order-information.* By order-information is meant the information in the spatial or temporal arrangement of stimuli. For example, the order information in a randomly arranged sequence of different items is $\log_2 n!$ bits. It will be assumed that order-information tends to be recorded with relatively little redundancy and that it is therefore easily forgotten. This assumption implies that storage space in the brain is at least partially specific. For if it were general, low redundancy coding of order-information would not be inevitable. A word should be said about the meaningfulness of separating out the order-information from the total information. It is impossible to remember order-information alone. Nevertheless, it *is* possible to remember what the items were (the what-information), without remembering their order. It is also possible to remember the order-information without remembering all the what-information. Thus it would suffice to remember that item 1 came from class α, item 2 from class β etc., without remembering which members of each class were involved.

4. *The effect of repetition.* If information is presented to a person on a number of occasions, this may strengthen learning in either of two ways. Firstly, each repetition of the information may establish a new memory trace. This would strengthen learning, despite decay of each individual trace, provided information extraction from the traces so established were efficient. Secondly, repetition may reconstitute the changes which form the trace which may then become more lasting so that the capacity of the trace decays more slowly. This would have the advantage that no additional storage space were used. It seems not unlikely that repetition has both these effects.

5. *Use of stored information.* Because certain information is stored in the trace or trace-system it does not follow that it is available in recall. Assumptions are therefore necessary about the efficiency with which stored information is extracted. It will be assumed that information extraction from the trace or traces of something one is trying to recall is usually efficient. This is the simplest assumption and it is not unplausible except

for certain pathological conditions. (If repression is a genuine phenomenon, it might be interpreted as motivated inefficiency of information extraction. Even then, it could still be held that inefficient information extraction is the exception rather than the rule). However, it will not be assumed that information extraction from the trace-system as a whole is always efficient, since to extract such information may require a search process and complex recoding. Consequently although information which has been lost from an individual trace may still be present in the trace system, recall may fail because this information is not readily available.

6. *The coding system.* The nature of the coding system is an extremely important part of any theory of memory. All the foregoing assumptions (except the first) indirectly concern this system. A few further points of a general nature should be made. If coding is to be efficient, it must be adapted to the probabilities of events. Although the nervous system may treat some events as more probable than others ab initio, most probabilities must be learned. Efficient coding therefore demands that the coding system changes as the individual gains experience of the world in which he lives (cf. Oldfield, 1954, *ref. 13*). Alterations in the coding system must themselves be stored at least in the sense that the decoding of later established traces is partly dependent on earlier established traces. Thus the trace system may not only supply lost information from an individual trace but may also determine how it is decoded. Little can be said about the nature of the coding changes. Clearly they are often bound up with learning to use new words. The effect of these changes is to reduce trace system redundancy which (while it can be useful) would otherwise involve excessive consumption of the brain's limited storage space.

SOME APPLICATIONS OF THE THEORY

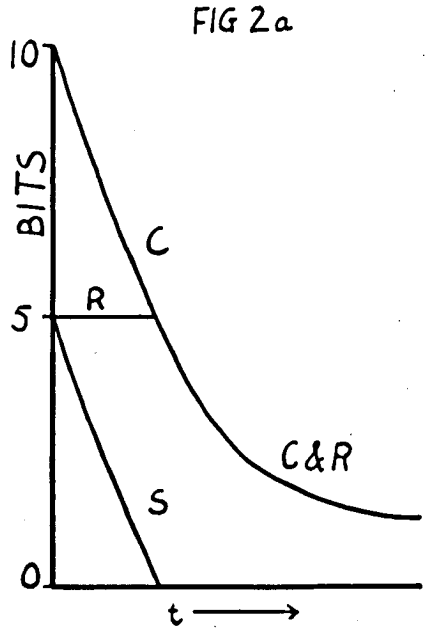
The basic problem is what determines the time course of forgetting. The time course of forgetting does not necessarily follow the time course of decay. Failure to appreciate this has been a main cause for the unpopularity of the trace-decay hypothesis. Among the factors which will influence the time course of forgetting are (i) the amount of redundancy, (ii) the coding, (iii) the supply of information from the trace system, (iv) pro- and retro-active inhibition. Some hypothetical examples may help to clarify possible relationships between trace-decay, forgetting, redundancy and coding. Suppose 5 bits of information are coded into a trace of 10 bits initial capacity and that the capacity decays with time t as shown in curve C in *fig. 2 (a)*. If coding is ideal no forgetting will take place until the capacity falls below 5 bits and retention will then follow the fall in capacity. This is shown in curve R. Curve S shows how the

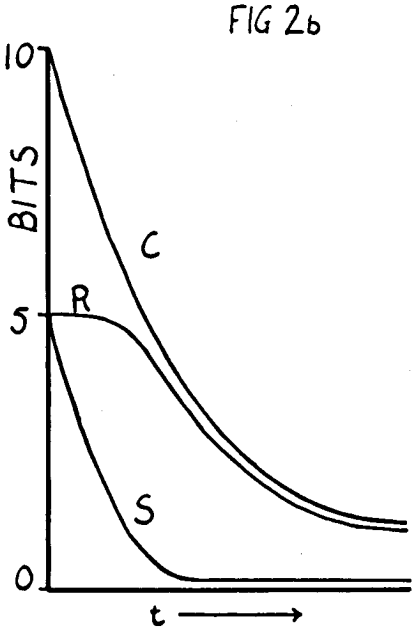
redundancy falls. In practice, however, coding will not be ideal, even if it is optimal, and relationships such as those shown in *fig. 2 (b)* are therefore more probable. Curves C, R and S again refer to capacity, retention (residual information in the trace) and redundancy respectively. Even here, the discrepancy between optimal and ideal coding may be optimistically small if the particular case previously considered is any guide (see *fig. 1*). *Figure 2 (c)* shows what might happen if different parts of a message were coded with different degrees of redundancy: a plateau in the curve of forgetting occurs when residual information in the trace is still protected by redundancy. Finally *fig. 2 (d)* exhibits the condition for relatively long-term retention. Only 2 bits of information have been coded into the trace and, by the time the redundancy has been lost, trace decay has become slow.

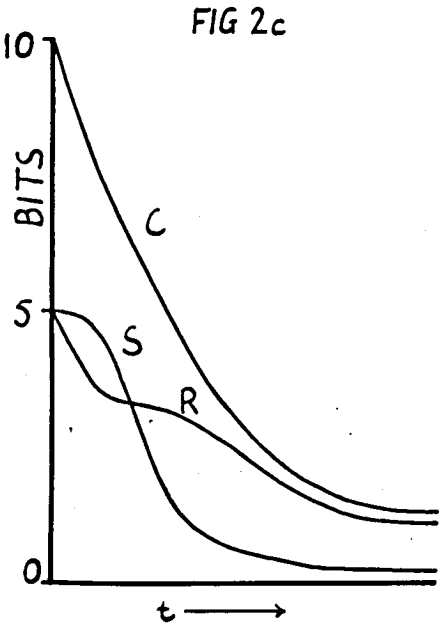
In *figs. 2 (a), (b), (c)* and *(d)*, the possible effect of trace system redundancy has been ignored. When information is lost through decay of an individual trace, it may be available from the trace system. This will happen especially often with order-information, if order-information is especially liable to be lost (assumption 3). For example, memory of the order of the words in a quotation may depend on information in the trace system concerning probable word orders. Extraction of trace-system information may often be a complex process. This might explain many cases of recovery of recall which have been attributed to release of inhibition: it is an open question whether the patient who succeeds in recalling a childhood event after six weeks on a psychoanalyst's couch does so through release of repression or through use of trace system redundancy.

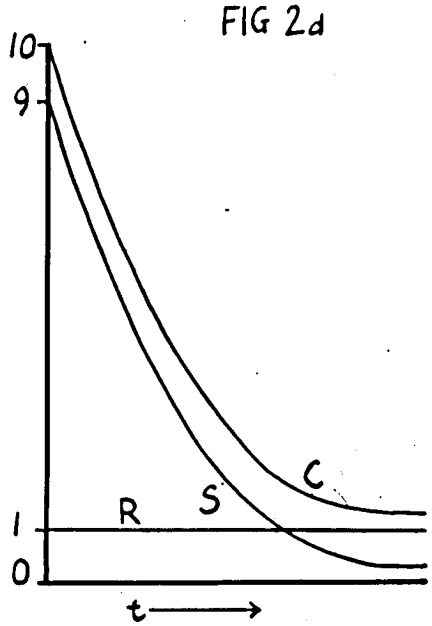
Another of the factors which influence forgetting is pro- and retroactive inhibition (P.I. and R.I.). If two similar things are learned in succession, forgetting of either tends to be accelerated. It has been suggested (by Underwood, 1957, *ref. 17*, for example) that all forgetting is due to R.I. and P.I., with the implication that trace-decay need not be postulated. But R.I. and P.I. are names for effects rather than causation processes and it may be *because* of trace decay that R.I. and P.I. occur. For if a number of things are learned at different times, trace decay may make the subject forget which things were learned when. Thus if two lists of nonsense syllables are learned in succession (as in the typical experiment on R.I. and P.I.), forgetting of either list may be accelerated by confusion between them. To say that forgetting is due to R.I. and P.I. may be another way of saying that it is due to loss of order-information through trace-decay. It seems unlikely however that only order-information is lost through trace decay.

We may now consider how evidence concerning the time course of the decay of the capacity of the trace can be obtained, since forgetting does not automatically reflect this decay accurately. The general principle is to maximize the residual information, R, in the trace for all values of t









Figs 2(a), 2(b), 2(c) and 2(d). Possible relations between trace capacity C, residual information R and redundancy S.

(where t is the age of the trace). Fortunately, there is no reason to think that the conditions which maximize R for one value of t will not also maximize R for all values of t. This is because, provided the 'noise' in the trace is random, optimal coding will always be such that the points used in the code hyperspace are as far from one another as possible. Techniques for trying to maximize information transfer in psychological studies have been discussed by Quastler (1956, *ref.15)*. A complication has to be borne in mind. The amount of storage space the trace occupies and hence its capacity is assumed to be variable (assumption 3). In practice, attempting to maximize R is likely to include maximization of the amount of space which the trace occupies. If some storage space is specific, two types of maximization are possible. In type A, as much space as possible of a specific kind would be used: in type B, as much space as possible of any kind would be used. For type B, different kinds of information would be included in the message to be stored. If there is specific storage space in the brain, type B maximization should lead to a higher value of R.

The assumption (assumption 2) that the amount of storage space used for a given message is variable provides a possible interpretation of a number of phenomena.

Firstly, intentional learning is generally superior to incidental learning. This may be partly because more storage space is used during intentional learning so that more redundancy and better retention is possible. Brown (1954, *ref.3*) obtained quite striking evidence that intention-to-learn can have some such direct effect on the strength of learning. The experiment showed that the subject was able to choose which of two simultaneous sequences he would learn best even though he was forced to perceive *both* sequences. (Presentation of the sequences was very rapid so that rehearsal effects were eliminated). Secondly, if the effect of repetition is partly due to the establishment of further traces, it will be to this extent limited by the amount of storage space available. If the amount available at any one time is limited, then less space will be available after a period of intentional learning than after a period of rest. More space will therefore be available, on average, during spaced learning trials than during massed learning trials. Spaced trials can therefore establish traces with greater redundancy than massed trials, so that fewer learning trials tend to be necessary.

There is also a factor which will work against quicker learning which, if it does not outweigh the first factor, will at least ensure that trials must not be too widely spaced. This is that, with spaced-trials, learning will depend on traces which are on average older and more decayed. Once the criterion of learning has been reached, however, this factor will make learning more durable, since older traces decay more slowly (assumption 1).

Thirdly, the average amount of storage space available per item will fall with the number of items to be learned, if the amount of storage space available at any one time is limited. This may explain why it becomes

disproportionately more difficult to learn as the amount to be learned is increased. There is also another factor. As the length of a sequence increases, the order-information becomes an increasing proportion of the total information. If order-information is especially difficult (assumption 3), this will contribute to the disproportionate increase in difficulty with length.

Fourthly, the serial position effect in learning (i.e. the quicker learning of the ends of a sequence of items as compared with the middle) may reflect variation in the use of storage space. Thus the ends of the sequence may be taken as reference points for the order in the sequence and may be coded with greater redundancy for this reason. It is significant that the serial position effect is to some extent subject to voluntary control: Kay & Poulton (1951 *ref.10*) found that the *middle* of a sequence becomes well remembered if the subjects know that they may have to start recall in the middle. It seems possible that the difficulty of remembering the order depends on the distance, in number of items, from the nearest reference point. (A somewhat similar suggestion was made by Crossman, 1955, *ref.8*).

All the facts mentioned in the introduction have now been discussed except distortions in remembering. These have sometimes been ascribed to corresponding distortions in the trace (Koffka, 1935, *ref.11*). On the present theory, a distortion is liable to occur whenever a trace has become inadequate through decay. Recall then becomes reconstructive in character, i.e. dependent on use of information from the trace-system. This may lead to an unambiguous output but it cannot guarantee absence of error.

In conclusion, the theory will be applied to the problem of the so-called span of immediate memory. A subject can reproduce a sequence of items after only a single presentation, provided it is sufficiently short. The maximum length of sequence for which this is possible is known as the span of immediate memory. The most puzzling fact about the span is that, for a random sequence, it is almost independent of the information content of the individual items. Its actual value usually lies in the range seven, plus or minus two (see Miller, 1956, *ref.12*). This might suggest that there is a special mechanism for short term retention with about seven compartments (degrees of freedom, logons). A storage model with compartments has been suggested by Quastler (1956, *ref.15*), although he was concerned not with the memory span as such but with the storage of information assimilated from a single brief exposure of visual stimuli. Whether or not such a model is necessary to account for storage associated with the process of perception, an alternative theory is possible for the memory span (Brown, 1958, *ref.5*).

It is easy to see that, if there is trace decay, then there must be a limit to the length of sequence which can be reproduced successfully after a single presentation. For the traces established by the early part of the

sequence are decaying while the later part is being presented: conversely, the traces of the later part are decaying while the subject is attempting to recall the early part. Thus if the sequence is too long, decay of some of the traces will reach a point where information is lost, either about the items themselves (what-information) or about their order (order-information). Unless such information loss can be made good from the trace-system, correct reproduction will then be impossible. The relative importance of the what-information and the order-information in determining the size of the span will depend on the particular sequence. Since order-information is assumed to be coded with only a small amount of redundancy (assumption 3), it follows that the order-information will limit the span whenever loss of order-information cannot be made good from the trace-system, as when the order is random. This will explain why the information content of the items has been found to have little effect on the size of the span, since random order of the items has normally been used. It is interesting that the span is often as high as 30 for words in a meaningful sequence. For in this case, implicit knowledge of grammar and of other constraints on the way words follow one another reduces the amount of order-information which the traces established by the sequence have to supply. If the order of the items in a sequence was entirely predetermined (by selecting successive items from different classes in a fixed order), the size of the span would then be determined entirely by the what-information.

It might appear to be a consequence of this theory that the span should be extremely sensitive to the rate of presentation of the sequence. If the rate of presentation is halved, for example, the delay before recall will be doubled, at any rate for the first item of the sequence (it will be less than doubled for later items, unless the rate of recall is as slow as the rate of presentation). But at least two factors will tend to stabilize the span. One is that the order-information falls more than proportionately when the length of the sequence is reduced. Another is that, when the rate of presentation is slow, rehearsal of earlier parts of the sequence will be possible during presentation of later parts.

PLAUSIBILITY OF THE THEORY

The theory concerns the overall functional properties of the physiological mechanisms which underlie the phenomena of memory. It is not concerned with the actual embodiment of these mechanisms in the nervous system. But this latter problem cannot be entirely ignored. It would be foolish to postulate functional properties without regard to whether they are physiologically probable. At the same time it is well to recognize

that what now seems physiologically either probable or improbable may not continue to do so as our knowledge of the nervous system improves.

The basic assumption of decay in the memory trace is perfectly plausible. Indeed, in the search for the physiological changes responsible for learning, the problem so far has been to discover any changes in the properties of the nerve cell which are of sufficient permanence to form a possible basis for memory (Burns, 1958, *ref. 6*).

The assumption that order-information is stored with less redundancy than what-information is also not unplausible. For example, if stimulus A is followed by stimulus B, this may well lead to the establishment of a memory trace of A, a memory trace of B and also a memory trace of the event 'A then B'. The what-information would then be represented twice over, i.e. redundantly relative to the order-information. This type of reduplication of the what-information is a feature of the model for learning and classification in the nervous system proposed by Uttley (1955, *ref. 18*). Reduplication of the order-information can also occur in his model but reduplication of the what-information will always be more extensive.

The assumption that storage space is general, or at least of limited specificity, is liable to seem rather less plausible, especially to anyone who is attracted by Uttley's model. For in this model all storage space is entirely specific, since storage is achieved by each unit counting if an input specific to that unit occurs. An A-unit counts if A occurs, a B-unit counts if B occurs and so on. If storage space is to be general, it seems to be necessary to have storage space which is not mixed up with the recognition system itself, as it is in Uttley's model. There is one strong argument for the existence of general storage space in the brain. In the course of a lifetime we remember, and forget, a very wide variety of events. Moreover, this wide variety is only a small sample of the possible events we might experience and remember. If storage space is specific it is difficult to understand how the brain is large enough to contain enough storage space, since only a small fraction of it will be appropriate for the events that befall any particular individual. If, on the other hand, storage space is general, all or most of it can in fact be used. Indeed it could also be *reused*, if necessary, although no assumption to this effect has been made in the theory as it stands. If storage space is general, it is not impossible that the fall in learning ability with age is due to the individual beginning to run out of storage space.

The theory as outlined in this paper is only tentative. Nevertheless it seems to offer a promising approach to the problems of learning and forgetting. One of its virtues is that it leads, potentially, to quantitive hypotheses. To my mind, it does show that decay of the memory trace is not such a poor theory of forgetting as psychologists have held it to be. In conclusion I would like to thank all those who have contributed to the development of the theory by discussing it with me.

REFERENCES

1. ATTNEAVE, F. Some informational aspects of visual perception. *Psychol. Rev.*, 1954, **61**, 183.
2. BROADBENT, D. E. A mechanical model for human attention and immediate memory. *Psychol. Rev.* 1954; **64**, 205.
3. BROWN, J. The nature of set to learn and of intra-material interference in immediate memory. *Quart. J. exp. Psychol.*, 1954, **6**, 141.
4. BROWN, J. Unpublished Ph.D. dissertation. Cambridge (1955).
5. BROWN, J. Some tests of the decay theory of immediate memory. *Quart. J. exp. Psychol.*, 1958, **10**, 12.
6. BURNS, B. D. The mammalian cerebral cortex. *Arnold, London* (1958).
7. CONRAD, R. & HILLE, B. A. The decay theory of immediate memory and paced recall. *Canad. J. Psychol.*, 1958, **12**, 1.
8. CROSSMAN, E. R. F. W. The measurement of discriminability. *Quart. J. exp. Psychol.*, 1955, **7**, 176.
9. HICK, W. E. Letter to the Editor. *Brit. J. Psychol.*, 1955, **46**, 64.
10. KAY, H. & POULTON, E. C. Anticipation in memorizing. *Brit. J. Psychol.*, 1951, **63**, 81.
11. KOFFKA, K. Principles of Gestalt Psychology. *Harcourt, Brace, New York* (1935).
12. MILLER, G. A. The magical number seven, plus or minus two. *Psychol. Rev.*, 1956, **63**, 81.
13. OLDFIELD, R. C. Memory mechanisms and the theory of schemata. *Brit. J. Psychol.*, 1954, **45**, 14.
14. OSGOOD, C. E. Method and theory in experimental psychology. New York: *Oxford University Press* (1953).
15. QUASTLER, H. Information Theory (edited by E. C. Cherry). *Butterworth, London* (1956).
16. SHANNON, C. E. & WEAVER, W. The mathematical theory of communication. *University of Illinois Press*, 1949.
17. UNDERWOOD, B. J. Interference and forgetting. *Psychol. Rev.*, 1957, **64**, 49.
18. UTTLEY, A. M. The conditional probability of signals in the nervous system. *Radar Research Establishment. (R.R.E.) Memo.* No. 1109 (1955).

DISCUSSION ON THE PAPER BY DR. J. BROWN

DR. E. R. F. CROSSMAN: I read Dr. Brown's paper with great interest and agree with most of it, but I do feel that a clear distinction should be drawn between two types of memory which have rather different properties. The first may be called the "primary memory image", which stores unclassified sensory information, and the second is the memory for perceived material, which stores the classified results of perception.

To underline the difference, I should like to describe a recent experiment of mine on the memory for raw sounds. Words chosen from a small list were played at halfspeed on a tape recorder and subjects were asked to identify them. If they failed, after 20 and 40 sec. the range of choice was successively narrowed down. The fact that subjects could identify the sounds after delay, shows that they remembered something unmeaningful, and their reports show that they remembered the sound pattern itself, which seems to decay rapidly.

The second type of memory involves using previously established classificatory responses, or Gestalten, and here the pure decay-curve is much less applicable. In the particular case of the short-term memory for items such as digits, binary choices, points of the compass etc. I support Dr. Brown' in stressing the importance of order-information. I measured the span of immediate memory for different numbers of alternatives; there is a marked drop in span, n, as the vocabulary size, N, increases. However the total information rises with vocabulary size, whereas one might expect a constant "information-capacity". Now, by adding to the calculated selective information which is $n \log_2 N$, another term $\log_2 n!$ for the order-information, the result is more or less constant, at 25-35 bits according to the individual. What does Dr. Brown think of this result?

DR. A. J. ANGYAN: In connection with this very interesting and well conceived paper, may I be allowed to say something about my own analogical interpretation of these facts.

First of all may I say in defence of the Pavlovian theory of conditioning that even if the assumption of Dr. Brown should be correct, we should compare the amount of information carried by ordinary commonsense or psychological terms and the way of talking about observations on facts of behaviour with those of the Pavlovian conditioning practice and theory. It is apparent that the latter uses terms allowing considerably less redundancy than the psychological terms.

I should also like to point out that perhaps his assumptions about the constant storage content could be completed in another way. I tried to point out with my scheme that we do not have to deal with this type of decay of the memory traces, but only with one decrementally distributed over a whole life span of an individual. By means of a cyclical internal scanning system to be supposed in the brain, the thought mechanisms are distributed in two parts. The information has a greater degree of redundancy at the beginning of every cycle of searching for a cue, and is regulated by positive feedback, whereas at the other end of the cycle its feedbacks work in reverse and establish an inhibitory process which leads towards the sleeping state. The inhibition allows a redistribution of the what-mechanisms, as correctly supposed by Dr. Brown. I am inclined to interpret the commonly known memory decay curves in this way. This redistribution is randomly disturbed and the redundancy of information carried with it is increased at the next arousal or facilitation part of a cycle, except in normal speech, the redundancy of which is less variable. With this assumption we have a better representation of the known possibility that in spite of an existing general rule of a slow decay of memory processes, if we extinguish some what-information, we may recall it again when the same basic cycle is allowed to go on. On the basis of neurophysiological and developmental-physiological observations, I would suggest a mechanism of that kind ensuring a longer term, but more or less variable storage content for the animal brain and am interested to know whether he would agree with it.

If you would allow me to say some words about an animal model, the functional and structural mechanisms, and their dichotomy in the sense of the paper may be seen more clearly.

We make some observations on planarians (flatworms): by combining light stimuli with feeding one can reverse their originally phototactic or phototropic reactions. The starting point of our observations is that the negative phototropism of the flatworms, counted each minute on the light side of the container, the other half of which is dark, traces a straight line of a constant angle in a semi-logarithmic coordinate system. By conditioning light with food, the direction of this line can be exactly reversed. If we cut the animals in half they regenerate in a few days completely, from both parts. During regeneration, the head part of the animal gives a curve, oscillating between the dark and light parts of the container, whereas the phototropism of the tail part shows a delay, slower but straightforwardly negative phototropism. By testing the above mentioned conditioned reflex established before the transection, we found that responses to specific cues, specific intensity (what-contents) or to the sites of stimulation were retained only by the head part, during and after regeneration. Animals regenerating form the tail part show no recall of

these responses and must be retrained. In tests with phototropism, the latter retrace the time relations and the sequence of the formerly conditioned feeding experiments in their speed of movements, but in spite of that food is only found by them given a specific "cue".

It seems to me that an oscillating what-mechanism, and a straightforward near cue-seeking mechanism, with different time characteristics, might be to some extent, separately represented as functional principles of different structural organisation levels of even the simpler nervous systems. This would strongly support the view that a very basic general dichotomy on which memory functions are built is represented by them.

DR. A. M. ANDREW: I would like to quibble with Dr. Brown's use of Information Theory. My remarks probably do not affect his general conclusions, but I think it should be pointed out that Shannon's Theorem no. 11 cannot be made to apply to messages of finite length simply by assuming a smaller value for the channel capacity. Shannon proves that information can be transmitted over a noisy channel with an arbitrarily small frequency of errors when the message length is infinite. For messages of finite length, however, the frequency of errors cannot be made arbitrarily small. For example, if each message consisted of ten binary digits and the noise was such that each digit had one chance in ten of being incorrectly received, no coding system whatever would allow transmission of any information with better than 1 in 10^{10} probability of error. This is a very low probability, but the point is that it is necessarily a finite quantity.

I am assuming that the noise affects parts of the message randomly. If the noise were subject to some unnatural restriction, such as the requirement that it affects not more than one binary digit in each group of seven, a Hamming-type code could be applied to give zero probability of error. Noise in the nervous system is presumably not of this special kind, however.

The difference which my quibble makes to Dr. Brown's discussion is that the decay of information in a memory trace cannot be represented by his *fig. 2(b)* if R is taken to be an amount of information stored with zero (or arbitrarily-small) probability of error. To be rigorous, any curve R must be associated with a particular value for the probability of error.

SIR FREDRICK BARTLETT, CHAIRMAN: There is just one thing I want to ask Dr. Brown, and it is on the amount of stress he wants to lay on the particular difficulty which he points out about order information. I think what he has had to say about it was very interesting, but I would like to know whether he would think that the difficulty people have with orders - he suggests all kinds, but that is not true - is due entirely to the fact

that, as he records on page 732 "order information tends to be recorded with relatively little redundancy". Is that the only reason why order is very often quickly forgotten? It just happens that the only case where I have had a firsthand opportunity of investigating a prodigious memory, was entirely to do with an order of numbers. This was a lady who was a grandchild of the famous Dr. Bidder, the mathematician. When she had been a very small child, and unable at that time to understand anything about relations of numbers, she had been given by her grandfather to learn a tremendous long list of numbers arranged in a random manner in a book, and all the pages of the book, of course, were also numbered. I met her when she was a lady of about 60. The evidence was perfectly clear, and I think absolutely irrefutable, that she had not seen these numbered lists at any time since the fairly early age of 6 or 7. You could give her a page number and a number in the list, and then say "What was the next number? What was the seventh number after this?" Or take any other kind of order in the whole thing that you liked, either after or before, and she would accurately give you the exact and correct answer. In her case, would you say the order information was stored with a great deal of redundancy, or what? There is one other thing. It is perfectly true that word orders are very often exceedingly difficult to deal with in a direct manner, but if you have movement order, an order in dance, or a song order - an auditory type of thing - and if you connect the words with the movements or the music then it appears that order is not a difficult but an exceedingly easy arrangement to deal with. I have no doubt, that, within the general framework of Dr. Brown's system, he could account for this. My difficulty about his system is that it seems to have such a lot of variable conditions, which operate in such a lot of variable ways that it is almost impossible to find anything that could not fit into the system fairly well.

DR. J. BROWN (in reply): Dr. Crossman suggests that there are two types of memory store with rather different properties - one which deals with unclassified sensory information and one which uses previously established classificatory responses. He claims that there is rapid decay in the first type but not in the second. I agree that unclassified information is forgotten more quickly than classified. But I think this can be explained without postulating two types of memory store. On my theory, rate of loss of information depends partly on the amount of redundancy and partly on the appropriateness of the coding. A sensory input invariably contains a lot of irrelevant information which is discarded in the process of classification. Thus, when storage space is limited, this makes it possible to code the wanted information with much greater redundancy *after* classification than before. Moreover the process of classification is also likely to improve the appropriateness of the coding. Dr. Crossman's results on the

memory span are very interesting but he does something very peculiar with information theory. He calculates first the selective information content which is $n \log N$ and then adds to it a term $\log n!$ for the order-information. But the selective information already contains the order-information as well as what I have called the what-information (the latter in Dr. Crossman's experiment would be $n \log N$ minus $\log n!$) In private he has put forward a rationale for counting the order-information twice but it does not seem to me to be reasonable to do so. (It should be noted that the order-information is only $\log n!$ when no repetitions in the sequence of n items are permitted and that in this case the total information is $\log {}_N P_r$ for which $n \log N$ is a satisfactory approximation when N is much larger than n). Dr. Angyan's ideas sound interesting but I do not grasp them sufficiently well to want to comment. Dr. Andrew had the impression that I had failed to recognize that Shannon's Theorem no. 11 cannot be made to apply to messages of finite length. However on p.735 of my paper I did say "constraints may exist - such as not coding over infinitely long sequences - which are such that even the best possible code falls appreciably short of the ideal". It was because of this that curve R in *fig. 2(b)* was drawn to be below curve C at every point. I agree with Dr. Andrew that R does not necessarily represent the amount of information stored with zero (or arbitrarily small) probability of error and this is in accord with the facts of memory. On Sir Frederick Bartlett's points, this remarkable Miss or Mrs. Bidder is certainly a worry to anyone who claims that there is decay of memory traces. However, she had gone over the digits a large number of times, though she had not been tested for quite a number of years, and I do suggest in my paper that it is quite possible to have permanent memories built up, because if you can have the trace still redundant when decay has become slow, then you have the condition for long term retention.

SIR FREDERICK BARTLETT, CHAIRMAN: The trouble is that that ought to be built-up much more often and by many more people.

DR. J. BROWN: Most people do not spend their time learning random numbers. I think it is possible that throughout life the amount of storage space which is available gradually decreases - it is greatest in the young child, and becomes less later on. This may be masked by another important fact, which is that, as you grow older, you learn more and more efficient ways of recoding. Incidentally, one of the difficulties about doing experiments on order information is that the subject may start recoding order information into what-information. The most dramatic demonstration of this is one, I think, by Miller, where he trained university students to recode binary digits into decimal digits and then, instead of having a memory span

of 8 or 10 of these binary digits, they had memory spans of 40 or 50. Sir Frederick is worried about the predictability of the system. I think that what one has to do is to take very simple situations where as few variables as possible enter, and then do the experiments under those conditions. After all, even Newton's Laws of Motion are only easy to apply directly to some very simplified physical systems, and in practice it is often impossible to gather together all the information that you need in order to apply them.

SESSION 4B

IMPLICATIONS FOR INDUSTRY

Chairman: THE RT. HON. THE EARL OF HALSBURY, NRDC, London

Chairman: MR. J. MERRIMAN, HM Treasury, London

SESSION 4B

PAPER 1

AUTOMATION IN THE LEGAL WORLD

by

DR. LUCIEN MEHL

BIOGRAPHICAL NOTE

Dr. Lucien Mehl, born 1918 in Paris, studied at the University, Paris where he obtained his degrees in Philosophy and Law, and a Diploma of Advanced Studies in Political Economy and at the National School of Administration.

He is now 'Maitre des Requetes' to the Council of State and Director of external training at the National School of Administration.

He is a member of the International Fiscal Association, the International Cybernetics Association and the French Operational Research Society.

He has published a number of articles on administrative science, law, cybernetics and operational research.

AUTOMATION IN THE LEGAL WORLD

FROM THE MACHINE PROCESSING OF LEGAL INFORMATION TO THE "LAW MACHINE"

by

DR. LUCIEN MEHL

INTRODUCTION

I. It may seem an ambitious step to try to apply mechanization or automation to the legal sciences. However, a machine for processing information can be an effective aid in searching for sources of legal information, in developing legal argument, in preparing the decision of the administrator or judge, and finally in checking the coherence of solutions arrived at.

(a) Introducing mechanization in a field of this kind is a particularly complex task, and imposes heavy obligations. In the first place, much preliminary work is needed for introducing automation in legal affairs, and so much work can only be decided upon if it is found to be of definite use. Secondly, such an undertaking is not without its risks; the jurist may lose direct contact with the sources of law and no longer have the benefit of the intellectual activity involved in searching for information. Lastly, as a result of mechanization of this kind, thought may itself become inflexible, diminishing creative power and innovative effort.

Nowadays, however, machine processing of information is becoming essential; "Homo sapiens" is in fact exposed to the risk of being overwhelmed by the vast accumulation of knowledge. It is becoming increasingly difficult to gain access to the sources of ideas, and the researcher wastes valuable time and often intensive mental effort in detailed and unprofitable research, never being sure whether his investigations will be fruitful, or whether he will not by-pass the essential information. Moreover, it happens that writers doing research in the same field of knowledge are unaware of one another's work; and besides this, the difficulty of finding the information required makes the researchers specialize still more. They find it hard to link up the different disciplines, because they are generally doomed to remain in ignorance of everything outside their own customary field of investigation.

Even within a well-defined field of knowledge such as law, the researcher is frustrated in advance by the vast accumulation of information sources. In legal matters, the number of laws and regulations and the scope of jurisprudence are growing on an alarming scale, and everyone is complaining about the situation - administrators and judges, as well as those dealt with under the law.

Thus the researcher, like the man of action, runs the risk of being confined to partial views, just at a time when works of synthesis are becoming more and more necessary in the modern world. Indeed, it is no mere chance that during recent years, the new methods proposed to theoreticians as well as to persons engaged in practical affairs all show a concern for synthesis (cf. the use of models, analogous argument in cybernetics, operational research, etc.)

It is, therefore, desirable to mechanize information retrieval, which must be speeded up, made more complete, more reliable and lead to synthesis in documentation.

The spare time created by such mechanization can then be devoted to research proper, to true scientific thought.

(b) It is likewise possible to some extent to mechanize processes of reasoning in the social sciences, especially in regard to the legal aspects. The aim must be not only to make available to these sciences the powerful tool of mathematics (especially its new aspects), but also to perfect and systematize logical argument, at least for problems whose solutions can be derived unambiguously from the data (thus this would not rule out synthetic control by humans, because the solution to a legal problem may depend upon extra-rational factors, involving the whole of human experience).

II. However, if information retrieval, and indeed logical argument, are to be mechanized, the problem of mechanizing logic must be solved first. The dream of Raymond Lulle, who mentions this possibility in his "Ars Magna"; of Descartes, who investigated the general processes of reasoning in his "Discourse on Method"; of Leibnitz, with his idea of a universal characteristic, the mechanization of logic is possible today, and has even been realized in certain respects, as demonstrated by M. Couffignal in his work "Les Machines a Penser" (Thought Machines). In fact, on the one hand logic has itself made definite advances: we now know enough about the laws of thought, and we have a better knowledge of the analysis of reasoning. Boolean algebra offers us a convenient system of symbols in this respect. On the other hand, it is possible to combine it with binary notation, which is admirably suited for analytical purposes.

In addition, this progress in knowledge has been accompanied by improvement in techniques. Not only have our methods of classification and selection been improved: punched card systems and - to a still greater extent -

modern computers, already offer us the elements of solutions. The latter, in particular, have a great capacity for storing information with relatively quick access; flexible programmes can be fed to them, including not only directions connected with mathematical operations, but logical instructions as well.

The problem of making a "law machine" certainly involves a technical aspect. It will be necessary to find the type of machine capable of fulfilling this function, to determine the essential features of such a machine. However, any machine suitable for making selections will generally be suitable to a greater or lesser extent. The problem is thus essentially a theoretical and logical one. For solving it, we require more highly-evolved analysis of legal concepts than that to which we are accustomed, conducted in a different spirit, in some cases. It invites us to define new legal concepts which will combine easily and unequivocally.

III. One may imagine two basic types of law machine:

(1) the documentary or information machine, or - in more familiar terms - the machine for finding the precedent (or relevant text),

(2) the consultation machine; less properly, the "judgment machine".

There is no fundamental difference between these two types of machine; the difference is one of degree rather than of essence. The consultation machine will give an exact answer to the question put to it, whereas the information machine will only supply a set of items of information bearing on the problem. Conceptual and relational analysis is more acute in the consultation machine; its structure is more delicate, the network of information is more finely woven.

In addition, this machine may be called upon to verify the logical coherence of the legal provisions of laws or conventions.

Finally, the analytical work needed for making these machines may supply the essential elements for developing a machine for translating legal texts.

The present study will, however, be confined to the first two types of law machine.

I. MACHINE RETRIEVAL OF LEGAL INFORMATION (*)

1. *The Usefulness of an Information Machine*

Machine information retrieval is thus the first stage in mechanizing juridical activities, involving searching for the relevant text, the jurisdictional precedent or doctrinal studies.

(*) Here we shall speak of "automatisation" or "mechanization" rather than of "automation" in the strict sense of the term. The latter really implies continuous processes, without human intervention, and is more a dream of the future than an immediate prospect.

The primary advantage of such mechanization is to remedy the difficulties arising from the multiplicity of legal sources; the law may be international, national or local; it may be expressed in treaties, laws, decrees, regulations, orders in the legal systems of law courts and administrative courts, in the principles laid down by governing bodies (circulars and instructions), and in those of writers, (treatises and reviews).

If it covers a sufficiently wide field, mechanization can also be an aid in collecting easily, and without error or omission, the items of information bearing on legal situations, for solving which a knowledge of precepts falling within different branches of the law is required. For example, assessment of the legal situation of a company director leads to investigations into commercial law, civil (or social security) law and fiscal law. In this connection, a machine could make comparisons which would not have occurred to man, reveal incoherencies or contradictions which would not otherwise have been disclosed, and lastly initiate original solutions, so as to advance legal science and equitable applications of the law.

2. Principles for Machine Information Retrieval

We already know that the mechanization of information retrieval involves a problem of developing concepts and relationships, much more than being a technical question. It requires a series of analyses and syntheses to be carried out within the mass of legal information.

(a) First of all, the legal information has to be set in order. The ideal way of doing this is to codify(ø) the sources of the law, whether they be legislative or statutory texts or texts of jurisprudence. This is not absolutely essential, but in any case it is necessary to set the information in order, following the general principle that rationalization must precede mechanization. One will then set about establishing the basic concepts and their relationships, which can be expressed by means of a juridical algebra which will facilitate the process of encoding.(*)

(b) The basic concepts are elementary legal notions, the combination of which will enable all possible situations to be defined, in principle.

(1) Determining these elementary concepts is obviously the most delicate part of the preliminary work for machine processing legal information.

However, we find an initial step towards this in the legal vocabulary, certain expressions of which constitute the tables of law compendiums. These expressions, which are intended to make reference to the tables or indexes easier, are sometimes called "key-words" or "guide-words".

(ø) "Codify" here applies the collection and integration into a single document, of texts bearing on a branch of the law, presented according to a rational scheme called a "code".

(*) By "encoding", we mean here the transcription of legal information into a conventional script which can be used on the machine.

In general, however, the key-words - established on an empirical basis - cannot be regarded as true basic concepts. Indeed, the legal vocabulary is often ambiguous. This will be found to apply to the word "droit" itself, which in French can mean legal science, a faculty or prerogative or again, certain dues. In view of this diversity of meaning of the word "droit", one is inclined to reject it as a basic concept without any further investigation. The word "acte" is equally ambiguous, meaning either the legal action (actum juris), or the legal document (instrumentum juris).

In addition, the language of the law is burdened with synonyms, such as "offer" and "pollicitation", and furthermore these synonyms may cover nuances which are sometimes imprecise and of doubtful use. For example: limitation, expiration, foreclosure, prescription; or again, annul, repeal, rescind. The legal vocabulary sometimes becomes paradoxical. In French, the members of a "société" are "associés", but the members of an "association" are "sociétaires". It will be seen already that it is essential to give the words a single proper meaning, and eliminate synonyms.

(ii) Besides, the usual legal vocabulary, in particular, is too complex and too rich. A single word covers a vast amount of information and cannot generally be regarded as an elementary concept. In most cases, the elementary concepts cannot be defined by a single word or phrase.

Moreover, experience and reasoning show that the number of elementary concepts is relatively small, and that with a small number of well-chosen concepts, it is possible to cover all institutions and situations. Such reduction of the basic concepts to an elementary form will enable the simplicity, speed in operation and efficiency of the machine to be increased.

This idea may be expressed more specifically by reference to the exponential law of information. The data, notions, situations or problems capable of being expressed in basic concepts, affirmed or denied, increase according to a dual exponential function, whereas the concepts themselves increase in arithmetic progression.

Thus with two basic concepts, affirmed or denied, 4 logical combinations can be constructed: 2^2, or, in binary notation, 00,01,10,11. These combinations can themselves be linked up to form 16 logical functions: 2^{2^2}, or, in binary terms:

	00	01	10	11
0	0	0	0	0
1	0	0	0	1
2	0	0	1	0
3	0	0	1	1
4	0	1	0	0
5	0	1	0	1
6	0	1	1	0
7	0	1	1	1
8	1	0	0	0
9	1	0	0	1
10	1	0	1	0
11	1	0	1	1
12	1	1	0	0
13	1	1	0	1
14	1	1	1	0
15	1	1	1	1

Let us consider, for example, the two binary concepts, man-woman (H.F.) and single-married (C.M.) There are 4 possible combinations:

FC	HC	FM	HM
00	10	01	11

The logical function *man* becomes:

0 1 0 1

and the logical function *woman:*

1 0 1 0

The aggregate of single people is represented by the function:

1 1 0 0

The aggregate of bachelors and married women becomes:

0 1 1 0,

etc.

We thus finally get a form of super-encoding in binary notation, which will enable *all* the logical functions obtained from linking up the concept combinations to be realized (whereas the mere encoding of combinations offers only limited possibilities).

With 3, 4, 5, 6 basic concepts, we get the following table:

Basic concepts	*Logical combinations*	*Logical functions*
3	$2^3 = 8$	$2^8 = 256$
4	$2^4 = 16$	$2^{16} = 65,536$
5	$2^5 = 32$	$2^{32} \approx 4.10^9$
6	$2^6 = 64$	$2^{64} \approx 16.10^{18}$

One can see the rate at which the number of logical functions increases when the number of basic concepts is increased by a few units.

(c) Unfortunately, the method of developing the concepts is rather a laborious one, requiring systematic exploration and analysis of literature in the branch of law concerned. First of all, one assembles a number of key-words or ideas for guidance; then by successive approximations one can obtain concepts which become more and more exact and simple as the material is arranged in order.

It must be emphasized that "elementary concept" does not mean "rudimentary concept". The elementary concept sometimes expresses an extremely pregnant and significant idea, and it very often bears close affinity with its cognate words. In the field of chemistry, it may be compared with the atom, which is no less complex than the molecule, of which it is a part.

1. Sometimes the breakdown is direct "Lease" can easily be replaced by "contract for the hire of property" - a grouping of simpler terms; "articles" by "deed of partnership", etc. However, the first terms will frequently not exist (i.e. suitable words will not be available in the current language). It may be useful to express them by symbols, because if expressed in current language, they may be rather long. In fiscal law, for example, the word "contribuable" (taxpayer) which has "redevable" and "assujetti" as synonyms, is not necessarily a basic concept and it may be advantageous to use the following notion: a person or body corporate, subject to a fiscal obligation, either as a payee (receiving an income for example) or as a payer (paying a salary, for example). Moreover, it can be seen to be a binary concept. Regarding taxes on income, the binary concept "value received or given", or "appreciation or depreciation" of assets (since in French fiscal law, accrued values are in principle subject to tax) should be used in preference to "profit", "benefit", "income".

In the same way, the word "marriage" is not a basic concept in civil law. There seem to be grounds for retaining the concept of *"union"* which in association with others will evoke marriage, company mergers, communal groups. One might also find that there is a concept *"division"* (divorce, separation, splitting of municipal area, secession of territories, etc.) and a concept *"cessation"* (decease, cessation of an undertaking, winding-up of companies, etc.)

2. Moreover, it must be emphasised that the idea of an elementary concept is only a relative one. The analysis will have to be more detailed if great precision is desired, when an attempt is made to apply mechanization to a very extensive field. For the information machine, *slight* analysis will suffice, whereas the consultation machine is more exacting. In any case, the concepts must always be chosen in terms of the constituent elements of the problem under consideration, the question posed, and not in terms of the solution or answer deemed to be the unknown.

3. The advantage of binary concepts is that they are not only suitable for encoding; they may conveniently be combined with other series of concepts which do not have to be repeated in the memory of the machine, as well. For example, for the concept "person subject to a fiscal obligation" one will have only a single list of the different persons or bodies corporate, whether they be payees or payers. In the case of the concept "value given or received" there will only be one list for these different values, the lists themselves being series of elementary or compound concepts (It will sometimes be necessary to use ternary, quaternary concepts, etc.*)

(d) In the information machine, the combinations of, and relationships between, concepts are very simple. Whereas in a consultation machine all the logical functions have to be used, especially implication, two functions suffice here: conjunction or logical product of concepts, expressed in current language by the word "and", and disjunction, expressed in current language by the word "or". Conjunction and disjunction suffice to bring about the combinations and relationships of concepts defining the various legal situations.

3. *Designing the Machine*

(a) Any machine capable of selection can be adapted for machine information searching.

1. The first stage of the information machine is the card-index. Sets of information, encoded if necessary, are recorded on the cards. The card-index is an improvement on the book, the code or the table. It can easily be brought up to date by withdrawing or adding items, and the classification can be changed if necessary. In this way, the *mobility* of the basic information is ensured.

2. The second stage consists of using laterally-punched cards (selection by rods) or cards with general punching (visual selection). The card-index can then be consulted without any strict order for operating concepts being imposed: selection thus becomes *commutative*, and is speedier as well. In addition, there is no reason in principle for bothering to reclassify any cards which may be put back loose in the card-index.

3. Lastly, electro-mechanical processes (punched card systems) or electronic processes (computers) can be used. Selection thus becomes *automatic* and we can then speak of an "information machine".

---

* See Part II.

(b) The contents of the elementary document will vary according to the process used.

1. First of all, a card can be made out for each reference unit (clause of an act, legal decision, doctrinal study). The punch-holes in the card, or the recording on the magnetic tape correspond to the essential characteristics, i.e. to the basic concepts applied in the reference unit concerned.

With the manual, visual, and even punched-card selection systems, it is possible to have the legislative text or legal award, etc., recorded on the basic document, card or microfilm, allowing just a summary to be used. An electronic computer, on the other hand, will normally only give references (printed in plain language), and it will be necessary to refer to codes, digests or card-indexes.

2. One can thus apply the method of using one card for each fundamental idea or basic concept. The reference which covers the idea or concept concerned is defined by its co-ordinates and marked on the card by a small perforation. To find the reference (text or award) corresponding to a set of concepts, the cards relating to these concepts are superimposed. The references can be seen in the form of points of light. The advantage of this system is that no strict systematic arrangement of the concepts or complex codification is needed (Selecto system).

(c) The choice of technique will depend upon the extent of the legal field concerned and upon the requirements of the service, account being taken of the information capacity and selective speed of the various items of equipment.

1. The hand-sorted cards with lateral perforations allow 100 (4 x 25) elementary items of information to be recorded, which represents only about 6 basic concepts (with 128 perforations it would be 7). The number of concepts can be raised to 36 if they can be split up into 12 independent series of 3 concepts.

The visual-selection cards have a greater capacity (e.g. Filmorex, 20 x 28 = 560; Kodak 42 x 70 = 2,940), and the scope of the Selecto system seems to be of the same order. Adoption of card-indexes of this kind is in itself a definite advance.

The ordinary punched card-indexes have an intermediate capacity of 800 (80 x 10), but they have speedier selection.

As for the capacity of the great electronic units, it is practically unlimited.

2. The speed of selection must be taken into account. With hand-sorting, 12,000 cards can be dealt with per hour; 36,000 with visual selection and 60,000 with punched card selection. The selective speed of computers is beyond comparison with the speed of these systems (900,000 - 1,200,000 characters per minute).

3. It can thus be seen that if the amount of material to be recorded is large (and this is the case with legal material, owing to the accumulations from the past), automatic processes are needed.

The punched card system is not itself very satisfactory, as demonstrated by the test carried out in the Supreme Court of New Jersey (U.S.A.): Selection covered 180,000 cards on jurisprudence and the waiting period (maximum) was up to three hours, definitely longer than traditional procedures take.

Remedial measures can no doubt be found, such as dividing the card-index into relatively homogeneous and independent sections, which in some cases will enable the selections to be confined to a single section (but then the benefit of fully automatic operation and the possibility of comparison and cross-checking between the different sections is lost). We thus find that if it is desired to speed up searching for the required information, electronic processes must be adopted when the equivalent of 100,000 punched cards is exceeded. Exploration of the memory is thus infinitely more rapid, particularly when such exploration may not be exhaustive, if the memory is partially "addressable".*

(d) Mechanization of legal information thus leads to a certain amount of centralisation.

With manual or visual processes, this centralisation is confined to the study of concepts. This is done by a single team of specialists, which constitutes an "intellectual investment"; the cards can then easily be reproduced and distributed among subscribers.

With an electronic computer, however, it is the work of consultation itself which is centralised. Owing to the cost of such a machine, automation of legal information can only be undertaken on a national level. The machine could be installed at a legal information centre linked by teleprinter, or some other convenient means, to Parliament, to the principal courts and the major administrations. Capacity use would thus be ensured, and it would certainly be profitable, in view of the working time made available.

4. *Advantages of the Legal Information Machine in Greater Detail*

In the further discussion below, we shall assume that we have at our disposal a keyboard machine (digital); in effect, a computer.

(a) On the keyboard of the machine a concept will occur once only, whereas in the analytical table of a digest the key-words must be repeated.

Let us consider company law, for example. Under each of the headings, formation, dissolution, prorogation, etc., we shall find the following sub-sections in the analytical table:

private company, limited partnership, limited liability company, joint-stock company, etc.

But each of these sub-headings will occur as a keyword in its alphabetical or logical place in the table, the key-words themselves becoming sub-headings. One will see the following, for example:

---

* The electrostatic memory recently invented by the French engineer Edgar Nazare seems to meet the requirements of the law machine (memory based on a register system and binary notation, addressable and practically unlimited).

Limited liability company

formation

dissolution

prorogation etc.

It will be seen that, if the number of key-words is m, the complete table will contain $m!$ lines (assuming that all the permutations are valid). If the table is divided up only into headings and sub-headings, it will still contain $m(m-1)$ lines.

A keyboard of elementary concepts would thus allow a considerable amount of simplification.

(b) A second advantage of the machine lies in the fact that for the purposes of consultation the concepts may generally be taken in any order, especially when the question posed consists simple of a conjunction of concepts. If it is a keyboard machine, no specific sequence for striking the keys is required. With the digest tables which cannot give all the permutations, on the other hand, a certain sequence must be followed, involving the risk of wrong selection. In other words, with the information machine, the operation of selection is made commutative, and the machine reduces the maze to be searched.

Furthermore, by using fewer concepts than the number defining the problem posed, one obtains information on more general situations. By substituting one concept for another, information on closely-related situations can be obtained.

(c) A certain amount of training will undoubtedly be required for using the machine. The operator will have to know the unnamed concepts and the way to combine them. Only a minimum of preliminary training will be needed, however. Moreover, a table of definitions for the unnamed concepts can be supplied with the machine. A dictionary for translating the common concepts into basic concept combinations may be conceived as well. The word "dictionary" brings out well the idea that the jurist using the machine will have to learn a new language.

Furthermore, the effort of defining and using new concepts will lead to progress in legal science, owing to the heuristic value of analysis and synthesis carried out according to procedures different from the traditional methods in law.

Thus Mr. Aurel David - who has done work on the foundations and symbolization of civil law - was able to reduce all the contracts to sale and hire. Whence we get a more precise conception of what is truly human in man. His capacity for work, his intellectual power, is no part of his actual person, since under certain conditions it may be made the object of a contract *(ref. 3)*. Moreover, the present writer has reached the same

conclusion in analyzing in fiscal matters the notion of income, which in every case comes from one source: capital, which may be either a material good, or the physical strength or aptitudes of man.

II. AUTOMATION OF LEGAL ARGUMENT

One may conceive of a machine capable of providing an exact answer to a problem put to it, and not merely a set of information on the problem. Development of such a machine required more detailed analysis of the concepts and the application of relationships more complex than those of the information machine.

1. *Principles*

(a) Automation of legal argument can be achieved in two forms:

1. One may merely mechanize a limited process: the machine will then provide a decision within a highly specialized field of law.

One may, for example, imagine automatic invoicing with the aid of punched card techniques consisting chiefly of the calculation of taxes on turnover applicable to the various products sold, according to their nature and intended purpose or destination.

2. One may also conceive - more ambitiously - a consultation machine which will answer any question put to it over a vast field of law.

Such a machine will obviously be more complex than the previous one, but according to the exponential law of information, its complexity will increase much more slowly than the volume of legal information which it can handle. This means that a machine covering the whole field of law would be simpler and less cumbersome than a series of machines handling separate legal sectors. Moreover, such a machine would be more efficient than all the others together, because it would provide complete and general solutions and would enable interesting comparisons to be made. Whatever the size of the machine, however, the theoretical solutions are the same.

(b) *Elements of the Solution to the Problem.* 1. The consultation machine will first of all require more exact use of specific concepts than the information machine. The latter supplies, in fact, documents or - what amounts to the same thing - references to documents, whereas the consultation machine must provide the solution to a problem. The concepts must thus combine with each other according to a strictly logical system.

2. Moreover, whereas the information machine only uses conjunction and disjunction, the consultation machine needs all the logical functions: affirmation, negation, conjunction, disjunction, equivalence, implication.

3. Let us, therefore, study how these functions can be expressed and handled in binary notation*.

Let A and B be two concepts. Expressing negation by a line above the symbol, we may have, as we have already seen, the four following situations, according as the concepts are both absent from the combination concerned, one of them is present, or both are present.

$$\overline{AB} \qquad A\overline{B} \qquad \overline{A}B \qquad AB$$

Representing negation by 0 and affirmation by 1, these combinations can be expressed as follows:-

00 10 01 11

Giving each of these combinations value 0 or 1, according as one of the concepts in them is denied or affirmed, we get a series of logical functions. In particular, A is expressed by

0 1 0 1

and B by

0 0 1 1

The disjunction of A and of B, which we shall express as "A*u*B", is defined in binary notation by associating the value 0 with the simultaneous absence of A and B and the value 1 with the other combinations:

A	0	1	0	1
B	0	0	1	1
A*u*B	0	1	1	1

The conjunction AB, the case where A and B are both verified at one and the same time, will obviously be expressed by

A	0	1	0	1
B	0	0	1	1
AB	0	0	0	1

Every time we have to combine logical functions, even of an order more than 2 (i.e. derived from more than two concepts) we will know that the function *"disjunction"* being 0111, the resultant logical function is deduced from the functions for combination, by writing 1 when 1 occurs

* Cf. Louis Couffignal, "Les Machines á Penser" *(ref.2)*. This work gives a precise account of the mechanization of logic, from which we have taken inspiration here.

in any of the functions in the row concerned, and 0 if all the functions in the row include a 0.

It is easy to verify that the process of "*conjunction*" is very simple, as well: the resultant function is deduced from the functions to be combined, by writing 1 when there is a 1 in *all* the functions in the row concerned, and 0 in the contrary case.

The process of implication is more complex. If implication is expressed as $\subset$, we may write:

$$A \subset B = \bar{A} \cup B$$

If, in binary terms, A is written as 0101, $\bar{A}$ is 1010 and B 0011, hence $A \cup B = 1010 \cup 0011 = 1011$.

The function of implication is thus 1011 in binary notation. It means that, in comparing the various rows of the implying function and of the implied function, we must adopt the value 1 to determine the resultant function, except if there is a 1 in the row of the implying function and a 0 in the corresponding row of the implied function, in conformity with the following:

A	0	1	0	1
B	0	0	1	1
	1	0	1	1

If the resultant function only contains a series of 1's, it means that the implication is verified. If it contains at least one 0, it is not verified (see example above).

4. The concepts and relationships can thus be translated into the binary language of the machine, but to arrive at this result it will be necessary to arrange an intermediate stage and use an intermediate language (or rather, script) between the legal language, which remains a human language, and that of the machine. This intermediate script, which must define symbols, will be, as it were, a legal algebra.

Indeed, the need for a system of symbols is manifested every time one wishes to introduce strict logic into a branch of knowledge (e.g. mathematics or chemistry). Symbolism ensures accuracy in expressing basic data and speed in notation. It enables reliable reasoning to be conducted, avoiding any ambiguity in the combination of concepts. But it remains intelligible, whereas the machine binary script is a disconcerting abstraction.

2. *Concrete Example.*

To illustrate our proposition, we shall give an example of mechanization applied to a system of taxation on turnover.

This exercise in fiscal algebra will no doubt appear rudimentary. The concepts involved would admittedly appear to be inadequate for the requirements demanded from the basic concepts, in a more profound analysis. Furthermore, the example is relatively simple, as the concepts in question, affirmed or denied, are 4 in number.

Nevertheless, this example gives an intimation of what might be the way in which a juridical machine will work.

(a) *Account of the System and Equations.* We shall assume that in a given country a system of taxation on turnover is in force, the rules of which are as follows:

The system is cumulative, so that every transaction involves taxation of the overall value of the product.

Any supply of goods by a taxpayer to a person or (corporate) body, whether a taxpayer or not, involves taxation, if it is made within the country, at a principal tax (V_1) the general rate of which is t_0.

The sales of retailers (D), however, are subject to tax at a reduced rate (t-), if they are current products (Pc), but if the same retailers sell luxury goods (P1) the tax is levied at the higher rate (t_+), never charged when the sales are made by a wholesaler (G), whether he be manufacturer or trader.

The manufacturers (F) are subject to the principal tax under the same conditions as the traders (C). However, if they make sales direct to consumers (manufacturing retailers, FD) they pay an extra tax (V_2) at a rate t', in addition to the principal tax at the general or higher rate.

Lastly, exports are tax-free, but if the goods are exported by a trader (i.e. a non-manufacturing person), he gets a refund (R) of the tax levied at the previous stage, as well.

If the items are represented as follows:

$\longrightarrow$	sales
$-\!\!+\!\!\longrightarrow$	exports
S	tax
$=\!\!=$	the relationship of a necessary and adequate condition,

then conjunction of the concepts being expressed by juxtaposition of the symbols, and negation by a line above the symbol, we get

(1) $G \longrightarrow = SV_1\ t_o$

(2) $CD \xrightarrow{P_c} = SV_1\ t_-^t$

(3) $CD \xrightarrow{P1} = SV_1\ t_+$

(4) $FD \xrightarrow{P_c} = (SV_1 t_o)\ (SV_2 t')$

(5) $FD \xrightarrow{P_1} = (SV_1 t_+)\ (SV_2 t')$

(6) $C \longrightarrow\!\!\!\!+ = SR$

(7) $F \longrightarrow\!\!\!\!+ = \overline{S}\ \overline{R}$

It will be seen from this example that it is possible to represent a relatively complex system with a small number of logical equations (7).

(b) *Transcription into Binary Script.* Transcription of these equations into "intelligible" binary script for the machine is a relatively simple matter.

It is easy to confirm that the terms on the left-hand side of the equation only concern four basic binary concepts.

Let us give the arbitary value 0 to one of the series, and the arbitary value 1 to the other.

0	C	D $\longrightarrow\!\!\!\!+$ P1
1	F	G $\longrightarrow$ P_c

These four concepts can provide 2^{2^4} logical combinations in which they are successively affirmed or denied. Their associated binary numbers are as follows:

F	0101	0101	0101	0101	C	1010	1010	1010	1010
G	0011	0011	0011	0011	D	1100	1100	1100	1100
$\longrightarrow$	0000	1111	0000	1111	$\longrightarrow\!\!\!\!+$	1111	0000	1111	0000
P_c	0000	0000	1111	1111	P1	1111	1111	0000	0000

(c) *Example of the Machine in Operation.* (1) Let us form the logical function which expresses the causes of taxation at the extra tax t'. If u expresses disjunction we get the following in symbolic notation:

$$SV_2t' = (FD \xrightarrow{P_c}) \; u \; (FD \xrightarrow{P1}) = FD \longrightarrow ,$$

then effecting conjunction of F, D and

F	0101	0101	0101	0101
D	1100	1100	1100	1100
$\longrightarrow$	0000	1111	0000	1111
FD $\longrightarrow$	0000	0100	0000	0100

(2) Using $\overline{}$ for the implication function, we now put the question

$$FD \xrightarrow{P_c} \overline{} \; SV_2t' \quad ?$$

We first of all carry out conjunction of the terms of the first part:

F	0101	0101	0101	0101
D	1100	1100	1100	1100
$\longrightarrow$	0000	1111	0000	1111
P_c	0000	0000	1111	1111
FD $\xrightarrow{P_c}$	0000	0000	0000	0100

To find out whether the first part implies the second, we shall combine it with the second part by means of the implication function (1011).

FD $\xrightarrow{P_c}$	0000	0000	0000	0100
SV_2t'	0000	0100	0000	0100

The resulting function is:

	1111	1111	1111	1111
since	0	1	0	1
	0	0	1	1
	↓	↓	↓	↓
	1	0	1	1

That is, the relationship:

$$FD \xrightarrow{P_c} \overline{\overline{\quad}} SV_2t' \text{ is a true one}$$

On the other hand,

$$CD \xrightarrow{P_c} \overline{\overline{\quad}} SV_2t' \text{ is false, because}$$

with CD $\xrightarrow{P_c}$	0000	0000	0000	1000
SV_2t'	0000	0100	0000	0100

the resultant function is:

1111 1111 1111 0111

One can see, moreover, that when the machine gives a negative answer it indicates at the same time, by the position of the zeros, *why* the answer is negative. In fact, it is easy to confirm that in order to have a 1 in the thirteenth position, positions 13 and 14 of the first function must be reversed, i.e. manufacturing retailers are involved. The machine thus gives *motivated* decisions.

It will also be noted that if one omits to insert a condition in the question posed, the machine will point out this omission as well. In other words, not only does it motivate its answers; a dialogue may arise between it and the person consulting it, as well.

Of course, in this example, which comprises only 4 basic concepts, the results are almost immediately apparent, which would not be the case with a greater number of concepts. With 20 basic concepts, for example, the machine would provide much less obvious answers and would make its deductions quicker than a human being.

(d) *Supplementary Remarks.* (1) In the foregoing example, the basic concepts are binary. It might be objected that the numerical symbolism adopted is inadequate, if, for example, the concept constitutes an odd aggregate.

In practice, however, one can always reduce a complex concept to a system of conjunction and disjunction of binary concepts.

We shall consider an example:
Let us imagine that we have to transcribe the following three concepts into binary notation:

Manufacturers	F
Farmers	Ag
Traders	N ,

representing the various parties subject to a given tax:

we shall make

$$A_1 = F \cup Ag$$

$$A_2 = F \cup N$$

These concepts, obtained by disjunction, also have a concrete meaning: A_1 represents the producers (agriculture and industry) and A_2 the merchants (commercants) in the legal sense(*) (who in French law include traders and manufacturers, but not farmers).

We then write:

$$A_1 = F \cup Ag = 0101 \text{ producer}$$

$$\overline{A_1} = \overline{F \cup Ag} = 1010 \text{ non-producer}$$

$$A_2 = F \cup N = 0011 \text{ merchant}$$

$$\overline{A_2} = \overline{F \cup N} = 1100 \text{ non-merchant}$$

and then get:

$$F = A_1 A_2 = 0001$$

$$Ag = A_1 \overline{A_2} = 0100$$

$$N = A_2 \overline{A_1} = 0010$$

To this list we may add the taxpayer:

$$A = F \cup Ag \cup N = 0111$$

and the non-taxpayer:

$$\overline{A} = \overline{A}_1 \overline{A}_2 = 1000$$

It may even be useful to draw up a list of all 16 possible combinations which have a concrete meaning.

---

* This is why we have chosen here the word traders (negociants) to designate merchants in the usual sense of the term (sellers, not manufacturers).

0000	$A.\overline{A}$	No-one	1111	$A \cup \overline{A}$	Everyone
1000	$\overline{A}$	Non-taxpayers	0111	$A=F \cup N \cup Ag$	All taxpayers
0100	Ag	Farmers	1011	$\overline{Ag}=F \cup N \cup \overline{A}$	Everyone except farmers
1100	$\overline{F \cup n}$	Non-merchants	0011	$F \cup N$	merchants
0010	N	Traders	1101	$\overline{N}$	Everyone except traders
1010	$\overline{F \cup Ag}$	Non-producers	0101	$F \cup Ag$	Producers
0110	$Ag \cup N$	Taxpayers non-manufacturers	1001	$Ag \cup N = \overline{A} \cup F$	Non-taxpayers or manufacturers
1110	$\overline{F}$	non-manufacturers	0001	F	Manufacturers

We thus see what a large number of different combinations can be obtained from two concepts which are affirmed or denied.

(2) One may also note that the order in which the concepts are chosen is immaterial; but this of course assumes that by "concept" we mean a notion with a well-defined function in a system of relationships. If, for example, we write

$$\overline{A} \longrightarrow N$$

(a non-taxpayer selling to a trader),

$\overline{A}$ and N have not the same function as in

$$N \longrightarrow \overline{A}$$

In other words, a notion such as "trader" may cover several concepts (in linguistics, this is the phenomenon of declension). If one wishes to reduce the number of concepts, some order must be assigned to them (as in French or English syntax, whereas in Latin the word-order is optional; at any rate, largely so).

It seems, moreover, that the electrostatic memory to which we have already referred, enables the concepts to be taken in varying order, which may lead to simplification.

More generally, it seems possible to conceive of machines programmed in such a way that the meaning of the words and propositions depends - as in natural language - upon the general context, which would constitute a further advance.

(3) It will likewise be advantageous to divide the field of concepts into separate groups wherever possible. This will also lead to simplification of the machine's communication network, and will enable meaningless combinations to be eliminated, as well.

However, the necessary precautions should be taken so as not to deprive the machine of combination possibilities which might prove significant and useful.

3. *Designing the Machine*

We shall not dwell upon the design processes, the principle of which is the same as for the information machine. Improved equipment is required, however, and this rules out manual systems. Moreover, if mechanization of an independent process can be done by using a punched card system, a consultation machine - to be really efficient - needs an electronic computer with a memory which is largely "addressable".

CONCLUSION

The limits to Mechanization

Whether for the information machine or for the consultation machine, there are obviously limits to mechanization bound up with conceptual difficulties and with the existence of extra-logical factors. The machine is at the most suited to pursue an argument. It is incapable of evaluating data, and "a fortiori" of elaborating the principles of law.

(I) However complete the checking of legal material may be, however subtle the analyses carried out, however elaborate the classifications, it is probable that certain special cases, certain marginal situations will escape the designers of the machine, as demonstrated by the (apparent) paradoxes of the logic of aggregates.

Furthermore, this difficulty is not peculiar to jurisprudence. It occurs, for example, in the natural sciences (classification of species) and even in mathematics, where it is sometimes necessary to discover - for the expression of a function (and particularly for transitions to the limit) - formulations which are at once more exact, more general and more significant.

Besides this, it may happen in law that in certain specific cases, mere application of the principle of law leads to results which are manifestly absurd or iniquitous. It is then for the judge to seek out the underlying significance of the principle, the exact limits to its field of application and, if possible, more exact and adequate expression of the latter.

He will no doubt sometimes have scruples about deviating in this way from the application of the principle: the danger lies in giving way to a pragmatism which may be suspect. However, making judgments often involves departure from strict logic, which is proper to the machine.

(II) To judge is likewise not merely a process of applying the principles of logic. It involves use of all our resources, the totality of human experience, processes which are not wholly conscious but nevertheless valid. It is precisely because we do not know the source and mode of development of the moral principles, legal standards, aesthetic canons that we rely, for making decisions and resolutions, on the comparison of consciences, in the spontaneity of tastes and inclinations *(ref. 4)*.

Furthermore, there is no reason to suppose that these obscure processes of human thought correspond to inferior spiritual states. The contrary is very much the case: a desire to confine one's self to purely rational measures, supposing that this would be possible, would mean condemning one's self to impoverishment of thought, to dessication of the spirit.

Moreover, if the humanities are in some respects tending to take inspiration from the methods of the exact sciences, these latter are paying more and more attention to closed systems, the seat of complex equilibria, whereas in bygone days science analyzed open systems above all. Now it was in the field of the biological sciences and the humanities that the importance of the general and mutual interaction of numerous complex factors was first felt.

Thus although the juridical machine is suited to conduct legal argument, it is incapable of evaluating facts. This task falls to man, because the factual world often defies pure (rational) analysis. Finally, although the machine may be able to suggest solutions to us, it cannot formulate precepts. Elaborating the principles of law is for man to undertake.

A juridical machine can thus only be an aid to the jurist and not a substitute for him.

We shall have no "electronic judges" in the world to come, any more than we shall have a machine to rule us.

REFERENCES

1. BELEVITCH, V. Langage des Machines, Langage Humain. *Ed. de l'Office de Publicite, Brussels.* (1957)
2. COUFFIGNAL, L. Les Machines a Penser. *Les Editions de Minuit, Paris,* (1952).
3. DAVID, A. La Structure de la Personne Humaine. *Presses Universitaires de France.*
4. DAVID, A. Reflexions pour un Nouveau Schema de l' Homme. *Cybernetica, Revue de l'Association internationale de Cybernetique,* 1958, I, 39.
5. MEHL, L. Congress of Flemish Engineers, Antwerp. (1956).
6. MEHL, L. International Congress of Administrative Sciences (OPATIJA Conference, Yugoslavia. 1957).

7. MEHL, L. International Institute of Administrative Sciences (Paris, 1958).
8. PEAKES, G. L., KENT, A. and PERRY, J. W. Progress Report on Chemical Literature Retrieval. *Intersc. Publ. London.* (1957).
9. PERRY, J. W., KENT, A. and BERRY, M. Machine Literature Searching. *New York.* (1956).
10. PERRY, J. W. and KENT, A. Documentation and Information Retrieval. *The Press of West. Res. University and Intersc. Publ., Cleveland, Ohio.* (1957).
11. De GROLIER, E. Quelques problèmes de codification posés par l'usage des machine en vue de la recherche de l'information et de la traduction des documents. Second Congress of Cybernetics, Namur, 1958.

DISCUSSION ON THE PAPER BY DR. L. MEHL

PROF. Y. BAR-HILLEL: I am sorry that I missed Dr. Mehl's oral presentation and I shall now make my comments on his preprinted paper. I hope there was no strong divergencies between the material presented on these two occasions.

First, I would like to draw a still sharper distinction than Dr. Mehl did between the two types of information-providing systems; the one whose output is a list of documents, a set of abstracts, a set of documents, or anything that stands in a one-to-one relationship to them, and the other that provides a sequence of statements as an answer to a question. Not only is Dr. Mehl's distinction not drawn sharply enough but he also believes that these two systems show a close relationship to each other. I do not think so. For the last three years, I have tried to impress people with the fact that these two systems are quite differently organized and that there exist no a priori reasons to believe that they have anything of importance in common. *(ref.1)* To my knowledge, there exists in England no mechanized system of any generality that provides answer to questions. Attempts to create such systems have been made in the United States, and perhaps also elsewhere, but I know of none that were really successful.

The two types of information-providing systems have so often not been told apart or, even when their differences were recognized, were regarded as having a closely related structure, as with Dr. Mehl. One reason is that the usual procedure of getting an answer to a factual question (of which no one in the neighbourhood remembers the answer by heart) is to first determine one or more documents (reference books, catalogues, research papers, etc.) that have a good chance of containing the answer and then to scan through these documents - preferably with the help of an index - in order to find the answer. The *first* stage of this two-stage procedure of providing answers is then indeed identical with the *only* stage of the provision of a reference list of documents that are (presumably) relevant to some research problem. However, jumping from this obvious fact to the conclusion that therefore a mechanized one-stage answer-providing system must be structurally similar to a one-stage reference-providing system is a clear *non sequitur*.

In the customary two-stage answer-providing system, the first stage can be mechanized to a certain extent, as I mentioned in my symposium paper as can, of course, the structurally identical only stage of the reference-

Ref.1 on page 787.

providing system. I tried to show there that there is no serious chance of mechanizing the second stage. Total mechanization of an answer-providing system is therefore incomparably more difficult than the mechanization of a reference-providing system. It is not just a historical accident then that hardly anyone has so far tried to develop mechanized answer-providing systems and that those who have tried did not get very far. Dr. Mehl gives some indication how such systems could be developed, but his indications are surely not explicit enough to tell how one should really do it.

I personally do not think that you will see any serious developments in this direction in the near future. For some very restricted fields one could probably do something, though not very much, because the answer to a specific question will in general not just be stored in the memory; most of the time you will have to deduce it from what is stored, and unfortunately machines that are able to deduce are hardly in existence so far. The logic which Dr. Mehl is mentioning in his paper, namely propositional calculus or Boolean algebra, is only a very small part of the logic which is needed for real deduction. There is no indication that Dr. Mehl is aware of this simple fact.

In addition, in order to serve as a medium of deduction, ordinary languages would have to be normalized, a problem which Dr. Mehl realizes without seeing through all its formidableness. No serious attempts exist so far to provide such a normalization for any field of appreciable extent. If Dr. Mehl believes that this can be done in the foreseeable future, I would want to most strongly insist that there exist no justification for this belief. The problem of transforming ordinary language into a formalised language system whose underlying logic is at least an applied first-order functional calculus, has hardly been scratched so far, and nothing indicates that the enormous problems involved are on the verge of being solved. (The only relevant study of which I am aware is a report by Miss Thyllis Williams).

Let me also remind you of a fact well known to logicians that for language systems of the kind mentioned there exists no mechanical method to tell whether a certain sentence does or does not follow from a certain class of other sentences, though such a method exists, of course, for a system whose underlying logic is the propositional calculus. But then, as said above, such systems would be totally inadequate.

The examples worked out by Dr. Mehl in connection with this problem ingenious though they are, show the difficulties rather than a possible way out. If you go carefully through all the intricate symbolism, you will notice that the argument is totally *ad hoc*. No indication at all is given how one could derive, from this example, any generalisation of how to treat a similar subject. Similar examples are treated in first-year courses in

in logic. But surely there exists no unique way, or rather no way at all, for generalizing from such examples. So altogether, I am afraid that the hopes that information retrieval systems combined with logical deduction machines could be of help in solving problems in legal science are very premature, to put it very mildly.

MR. R. BENJAMIN: Prof. Bar-Hillel had to apologise that he had read the paper but not heard the lecture. I am afraid I have got to make the reverse apology; I have heard the talk but I have had no opportunity to read the paper.

Dr. Mehl points out that an essential preliminary to any mechanised literature search, in this field, is a rationalisation of the literature filing and indexing system, and that this rationalisation will be a major help by itself, whether or not mechanisation follows it. Referring to the possibilities of using electronic computers in the interpretation or possibly also in the drafting of legal documents, I feel the same problem would arise, and it would be a very severe one. From my *very* limited knowledge, it appears that legal documents use as few individual statements or sentences as possible, and make those as long and involved as possible. Computer language requires a very large number of very simple and concise statements; thus there is a very large problem in translating between these two languages. Indeed to a layman the rationalisation involved in putting some of the legal documents into the form of simple statements suitable for a computer, might be a very big help in itself - as well as being a formidable task.

MR. E. A. NEWMAN: I would suggest that, in certain respects, Prof. Bar-Hillel's criticisms of Dr. Mehl's paper are fallacious. Information retrieval is a subtle subject, and I shall attempt to make my point by use of an analogy. For my purpose there is sufficient analogy between solving a legal problem - or any other problem for that matter - and solving a generalised jig-saw.

Assume a store, containing all the pieces of every jig-saw puzzle that ever was. Assume that starting with one piece we have to make a complete picture.

What we would like is to find a label in the piece we have, in the form of a route instruction to get us to a store location which would prove to have just the remaining parts of our picture - and these already fitted together. The next best thing to this is for the piece we have to lead us to one piece that fits to it, and in turn leads to another piece, and so on. We are very much worse off if the clue on the jig-saw piece we have leads us to a location containing just the remaining pieces of the puzzle as a random arrangement, for then we have the difficult task of finding

in what order the bits fit together. If besides the jig-saw pieces we need, the store location contains pieces of many other jig-saws all mixed up - then we have a collosal task ahead.

The purpose of a library retrieval system is to find for us the information - the jig-saw pieces - we need, correctly ordered and marshalled. The jig-saw piece we have is often just a set of impulses in our central nervous system - which make us speak, or move, or make suitable noises. The resulting speech movements or noises should lead us directly to a store location containing just the information we require suitably marshalled. Nothing short of this is entirely adequate.

To achieve this the storage system must contain a parcel of correctly marshalled information for every problem we are ever going to wish to solve - to forecast these requirements is truly a Herculean task. Each parcel must be in a separate store location so organised within the store that the clue in our mind leads directly to it. In other words the information retrieval system has to fit all the problems we wish to solve, and the way we are going to ask for the information. Given foreknowledge this could be done. No store location would contain a book, but rather the relevant information extracted from many books.

In practice we use a common language coding system that takes us to several books and within the books use a common language index system to take us to items. We converse with a librarian - who to some extent knows what is in his books - and with his help convert our question into one that fits the system. But one thing is sure; to be of any use the coding systems we adopt must be related in some way to the problems we wish to solve, and the form our questions will take. When Prof. Bar-Hillel suggests that the indexing system one uses is quite different and unrelated to the question one asks - that is nonsense; in so far as it is so the index system is no good.

Further, as Dr. Mehl says, because in legal matters it is possible to some degree to anticipate the problems to be solved, and the form questions will take - to that degree it is possible to make a good index system - and in so far as it will relieve the librarian of some question matching - thus far the retrieval system can be automatic.

MR. J. W. FREEBODY: I think there must be something fallacious in Mr. Newman's own argument because if his hypothesis is true how is it possible for one to solve the type of problem which requires some originality of thought for its solution?

DR. I. C. PRICE (written contribution): This is a most interesting and clearly written paper. The development of techniques for searching for items of information and combining them logically would seem to have applications to much wider fields than law: for instance, scientific literature

searching, literary and textual criticism, and the mechanisation of office administration.

The use of large computing machines for legal purposes raises a number of interesting questions. If in a modern society the statutes, regulations, by-laws and legal precedents have become such a tangle that large computers are now required to deal adequately with regular legal business, the appropriate cure would appear to be common sense and self-control in the legislature rather than computers in the judiciary. But the use of formal logic in clarifying legal concepts and detecting inconsistencies is another matter. It could be argued that the desirability of avoiding ambiguities and inconsistencies, whether in new statutes or in case law, is such that formal logic ought to be used wherever practicable: and it may well be that the complication of the problem of applying formal logic to real legal problems would be such as to require the use of computers.

One field which might be used for a "pilot" investigation is the detection of inconsistencies in the rules of an association. This task would have two advantages, in that the field of concepts used is much narrower than that in the common law, and the law is already codified. The task of detecting inconsistencies in association rules is not entirely a trivial one: flaws are sometimes discovered in complicated rules, even when great care has been taken in drawing the rules up. An example occurred in the election rules of the Cambridge Union in 1956.

When the technique which the author has applied to his "tax" example is applied to a practical situation that might arise in the application of the rules of a society, one limitation shows up very seriously. This is the size of the storage required to describe a situation with n concepts. The storage of the description of each situation requires 2^n bits. Quite a simple problem can involve 30 concepts, which means that each storage location in the legal machine would contain about 10^9 bits, which is about 10^4 times larger than the whole immediate access store of a large modern electronic computer. So I doubt if the author's method of reducing legal argument to Boolean arithmetic will ever be used in practice. It might be better, though more difficult, to dry to develop a method of mechanical reasoning using Boolean algebra rather than Boolean arithmetic.

DR. L. MEHL (in reply): In spite of the precautions I hoped I had taken in my paper and speech, I find that my designs of juridical machines give rise to serious objections. However, I thought I had been prudent in my statements. It seems I was not. But now, I cannot go back and my duty is to face all the questions arising out of my ambitious designs!

The first question, examined by Prof. Bar-Hillel, is the distinction between the information retrieval machine and the machine for legal arguments: I persist in thinking there is no difference in a sense, because

the second machine, the machine for legal arguments, has not the pretension to create information. This machine only transforms the data according to logical rules. Of course, the analysis of the concepts and of the relations between them is more complicated in the second machine and more precise, but I do not believe there is really any essential difference between this machine and the one for information retrieval.

Indeed, this statement is not universally valid. But it is true in the area of legal problems, because the solution of these problems consists generally in finding a similar case foreseen by the written law or stated by a precedent award. In these conditions, it is possible, in legal matters, as Mr. Newman said, to anticipate, in a great number of situations, the problem to be solved. If the case is not expressedly foreseen, then the machine can only help the jurist find the solution; but it is unable to give it. I insisted on that point in my paper. In other words, as Mr. Freebody noticed, the machine is not able to solve a type of problem which requires some originality of thought for its solution. It is quite obvious also that, although the machine is, in principle, suited to conduct legal argument, it is incapable of evaluating facts.

May I add, with great respect to Prof. Bar-Hillel, that I am aware that Boolean algebra is only a part of the logic needed for real deduction. I know that logical problems are not all mechanizable, at least in the present state of our knowledge. This question was especially studied by Markow, Novikow (indecidability in group theory) and more recently by Jean Porte, (France, C.N.R.S.).

But it seems reasonable to begin with the simpler questions where Boolean algebra is available. I question also the affirmation that my example is totally *ad hoc*. The same system of formulation may be efficiently applied to other questions. Concerning generalisation of the system, I explained for example how a complex concept can be reduced to a system of conjunctions and disjunctions of binary concepts. The system here proposed is applied to turnover taxes, but it is also available for all taxes, and I examined also the questions of income-tax.

The system requires transformation of ordinary language into a formalized language, a formidable task said Prof. Bar-Hillel and Mr. Benjamin. Certainly it is, if we formalize the whole human language. But here the purpose is rationalizing juridical language, that is to say a technical language, and to divide the difficulties. I propose to begin with tax terminology (because tax law is very precise).

There is still an important problem concerning the storage of information, as Dr. I. C. Price explained in his interesting written contribution. He points out that, even with a reduced number of basic concepts, each storage location would contain a number of bits which can be a hundred times larger than the whole immediate access store of a modern computer.

Dr. Price suggests the use of Boolean algebra rather than Boolean arithmetic; I agree with him. I explained also in my paper that it would be advantageous - I might have said necessary - to divide the field of concepts into separate independent groups in order to lead to simplification of the machine communication network and to eliminate meaningless combinations.

May I conclude by expressing some surprise concerning some of the objections that have been made. There are numerous attempts and designs, and some realisations in mechanical information retrieval (for example in scientific literature, in patent office, etc.) The only originality of my paper, concerning the first machine, is to try facilitating information retrieval in the legal sphere by the same processes. If I feared anything, in presenting my paper, it was being trivial, and stating the obvious. As to the second machine, I agree with Prof. Bar-Hillel that it is not a prospect for the immediate future. But that is true only if we are thinking of a machine able to treat all juridical questions; in a narrow sector of juridical questions, it is possible to build such a machine, and particularly for tax questions, because tax questions are a very logical part of law, and in fact, such a machine, in a restricted meaning now exists. I gave in my paper the example of automatic invoicing. It is possible - and I believe the system exists in fact in certain companies - it is possible, when the invoice is automatically made, to include also the automatic calculations of turnover taxes, for example. The problem of knowing what is the rate for such and such a product, according to its nature, to its origin, to its destination, is a logical one, and computers are able to solve it.

I am, of course, conscious that the machine for legal argument which I propose is a very elementary one, but I think it is not a bad method to divide difficulties and to begin at the beginning.

I have tried to explain my thoughts in English: this is not very easy for me, so will you excuse me if I have not been very clear. Thank you.

REFERENCE

1. BAR-HILLEL, Y. A logician's reaction to recent theorizing on information search systems, *American Documentation*, 1957, 8, 103.

SESSION 4B

PAPER 2

THE MECHANIZATION OF LITERATURE SEARCHING

---

by

PROF. Y. BAR-HILLEL

SESSION 4B

PAPER 2

THE MECHANIZATION OF LITERATURE SEARCHING

by

PROF. Y. BAR-HILLEL

BIOGRAPHICAL NOTE

Yehoshua Bar-Hillel, born in Vienna in 1915, graduated from high school in Berlin, 1933, then went to Israel, where he got his M.A. in 1938, with Philosophy as major, Mathematics and Chemistry as minors. He went to U.S.A. in 1950 first to the University of Chicago, then to Harvard, then to M.I.T., as Research Associate in the Research Laboratory of Electronics in 1951-53 and also in the winter 1955/56. From 1953 he Lectured in Philosophy at the Hebrew University in Jerusalem and became Associate Professor in 1957. Since 1957 he has also taught in the Department of History and Philosophy of Science.

Major interests - symbolic logic, semantics, philosophy of communication. Principal topics of research:- foundations of mathematics, logic of ordinary languages, machine translation, and mechanization of literature searching. Joint author with Professor A. A. Fraenkel of "Foundations of Set Theory", to be published by the North-Holland Publishing Company in the series "Studies in Logic".

THE MECHANIZATION OF LITERATURE SEARCHING*

by

PROF. Y. BAR-HILLEL

SUMMARY

"FOUR sources of inefficiencies in the process of literature searching are briefly described. An "ideal" solution is outlined as a frame of reference and its shortcomings discussed. Mechanization of abstracting and indexing is rejected as impractical for the foreseeable future. The only stage in the whole process where mechanization is practically feasible at the moment is the procurement of a list of all the documents from a given document collection which fulfill a Boolean function over the subsets of the universal index-set of the collection. Abstracting and indexing are more complex intellectually than translating, and their complete mechanization therefore less likely than completely automatic machine translation."

MOST scientists and technologists, when engaged in research on some more or less definite subject, regard the reading of relevant literature as an integral part of their effort. It seems that only very few outstanding scholars are in a position nowadays to achieve important results without consultation of previously published material. It is well known that this consultation of the relevant literature is in danger of being inefficient in many different ways. First, and most important, the research worker might overlook some relevant publications. Secondly, he might spend too much time on reading irrelevant material. Thirdly, he might have to waste much time in order to arrive at the list of possibly (or probably) relevant reading material. Fourthly, time and money will have to be spent in getting hold of copies of the pertinent literature, whether through the gathering of an adequate library and document collection or through the obtainment, by various means, of copies of those books and other documents which are required for each special occasion; and there are very many ways in which this process is often inefficient, and seemingly tends to become even more so.

---

* This work was supported by a Grant-in-Aid from the National Science Foundation.

It is only natural that, with the advent of machinery which proved itself capable of performing operations which were for a long time regarded as peculiar to human brains, people began asking themselves whether these inefficiencies could not perhaps, partly or wholly, be overcome through the use of appropriately adapted machinery. I shall not dwell on those aspects of our problem that deal with providing copies of the recommended reading material. I shall rather concentrate on the three first aspects of the literature search problem.

In order to get a good grasp of the various aspects of our problem, it might be advisable to start with its "ideal" solution. The solution which immediately comes to one's mind consists in pushing, for every given research problem, an appropriate button, or combination of buttons, as a result of which action one would immediately be presented with a reference list, all items of which would contain material pertinent to the problem and such that no document containing pertinent material would be missed.

It requires but little reflection to discover that this ideal suffers not only from the defects common to all human ideals but is, in addition, at every stage, so full of ambiguities, vaguenesses and obscurities as to be close to contradiction. It is already an intolerable simplification to assume that all documents of the world can be classified, relative to a given problem, into two mutually exclusive and simultaneously exhaustive classes: Relevant and Irrelevant. Two decisive features are disregarded by such a conception: first, that relevancy is a comparative rather than a qualitative concept, a more-or-less rather than an all-or-none affair; secondly, that a document of little relevancy in the eyes of X might well be highly relevant in the eyes of Y, for reasons that are too obvious to need elaboration. In other, more technical words, relevancy, in the sense pertinent for our context, is a *pragmatical* rather than *semantical* concept. Even if a literature search system could be devised that would be "ideal" for X, it might well be far from ideal for Y, and there is no way out of the dialectics of this situation with regard to any system designed to serve more than one person. There is no need to advance further reasons to show that the mentioned ideal is unattainable *on principle*, not through human failure. Hence it is a false and deceptive ideal and getting rid of it is a necessary condition for a clearer view of the situation.

Let us therefore set our aim on a more practical target and be satisfied with a system that just works more satisfactorily than the existing ones, on the average, for a given group of users. Having eliminated the problem of setting up a universal system that would be optimal for every prospective user as a pseudo-problem based upon a pipe-dream, we are now confronted with the real problem of how to cope with the fact that the group of users of a given system is in general not stationary. A system which is efficient at a given time may become inefficient not only through changes in the body of literature to be searched but also through changes in the body of searchers.

This is, of course, commonplace, well-known to everybody working in the field at times, but apparently forgotten or repressed at other times, hence worth being stressed again.

Our "ideal" system can do us one more pedagogical service. How exactly is one to go about selecting the appropriate buttons to be pushed? Or rather, how does one set up a set of buttons so that, through selection of an appropriate subset of these buttons, the list of relevant references will be presented? This is our real problem, of course, after the utopian universal machine has been scaled down to a mechanism adapted to the needs of a given group of users.

We are now back to the bread-and-butter problem of improving extant literature searching methods and, during our preliminary discussion we seem to have entirely lost sight of the issue indicated by the first word of the title of this paper, "mechanization". Where does mechanization enter the picture now?

Some 10 years ago, electronic computers made their sensational début and proved themselves able to solve computational problems at speeds that were many degrees of order higher than those attainable by humans. Hence they were able to solve problems by sheer speed, even if the instructions given to the machine were far from being the most efficient ones, so that no humans were able to solve them by following these instructions, simply because their life was too short for such an endeavour. This striking feature of high-speed computers induced many thinkers to speculate about invoking their help in many other situations, where the methods usually employed in solving the problems arising in them were inefficient. Here, so one thought, was an opportunity of improving inefficient methods-not through the tedious process of analyzing the causes of their shortcomings and of finding out, perhaps by trial and error, which changes were apt to better their performance - but rather by performing the same operations, inefficient for a human, at such a high speed that they would thereby become efficient, at least more efficient than existing methods involving human beings.

To give an illustration, which I did not invent: surely, the most thorough method of searching the literature for a given research problem would be to read all the documents of the world or - to make it slightly less utopian - the whole collection of books and other documents in the Library of Congress. This method is clearly inefficient but - so it was argued in all seriousness - only because the speed of human reading is so small. Coding this library onto an appropriate storage medium and having a sufficient number of reading heads scanning this medium, the whole contents of the library could be scanned through in a few minutes, perhaps even seconds, and then ... And then what? At this stage, some thinkers went overboard. Intoxicated by the success of the electronic computers - this is the only explanation I can give of the phenomenon - they thought that somehow

the machine would be able to decide during this scanning procedure - occasionally called "reading", very suggestively but also very, very misleadingly (notice the term "reading head") - which of the scanned documents were relevant for the research problem at hand. Notice that the research worker himself could indeed have done just this - disregarding the question of the languages in which these documents were published. As a matter of fact, had we been able to increase the speed of our own reading (with understanding!) a billion-fold, this might have been a solution (certainly not, even in this imaginary case, the solution) of the literature search problem. Monitored by the slogan, "Whatever a human can do, an appropriate machine can do, too", one can see the seductive power of the argument. Nevertheless, there is scarcely need for pointing out the enormous fallacy committed here. Scanning is not reading with understanding. There are no serious proposals in view how to instruct the scanning device to select the relevant documents. Though this seems now to have been generally recognized, albeit with considerable reluctance, only slightly more sophisticated speculations have recently been hitting the headlines.

Let us now discuss some of the recent proposals to mechanize parts of the literature search process. Since this process is composed of many partial processes, it is a definite possibility that one or more of these stages could indeed become more effective through mechanization.

It might be worthwhile to work backwards. The last stage of a literature search consists in reading the documents which were recommended in the preceding stage, and which should be all the documents that contain material of relevance to the investigator's problem, if the previous stages were fully effective, and in making notes in your head or on paper for further processing. Though I admit that it is highly seductive to speculate on some scheme of mechanizing the note-taking - and I myself could easily supply you with many schemes for mechanized note-taking - I am utterly convinced that none of my schemes, or those of anybody else, will work in the sense of producing notes which are even remotely comparable in their quality with those taken by an intelligent reader. I would not like to sound too dogmatic, certainly not before this audience, but I must insist that the *onus probandi* lies entirely on those who would claim that note-taking is performable by machines presently existing or in the stage of development.

Before one reads a document *in toto* for the purpose of note-taking, one often prefers to read first an abstract or review in order to decide whether extensive reading of the whole document is really profitable. It is probable - though I know of no serious investigation along this line - that this approach is time-saving and should in general be encouraged on condition, of course, that an authoritative abstract or review is available. For some purposes, an author's abstract might suffice, for others an abstract or review by a recognized authority in the field is requested. Let me dismiss, from the beginning, the possibility of having a critical

review mechanically prepared, I shall give no reasons, first because the idea strikes me as too absurd to require serious consideration, secondly because I shall presently argue against the possibility of performing mechanically must less demanding operations.

The most one could in all seriousness think of as being performable at the moment by a machine at this stage of a literature search would be the preparation of an abstract of a certain severely restricted kind which I shall call, for the lack of a better term, an *auto-extract*, i.e. an automatically prepared partial sequence of sentences from the original document. The term is coined in analogy with the term 'auto-abstract' used by LUHN *(ref. 1)* against whose approach the present critical remarks are partly directed. It is again very easy to invent countless different methods of assigning "significance-values" to all the sentences of a given document and then to print out so many of the sentences with the highest significance values in their original order as an auto-extract, according to some criterion or other. One of these possible methods has actually been employed recently *(ref. 1)*; it is by no means the most attractive one.

Though the final proof of this pudding will again be in its eating - it has been claimed that the results obtained so far by the mentioned method have been encouraging *(ref. 2)*, though it has not been mentioned by what standards they were so - it is not difficult to point out its many defects which make it highly doubtful whether it could possibly attain its restricted aim. But the first thing to be clarified is whether this aim is at all a valuable one, to begin with. Now, would you be satisfied with a 100-word extract of a 2000-word paper, even if prepared by the greatest authority, in lieu of a 100-word abstract, where the abstractor is free to choose his own words? I personally doubt very much if an extract could, in general, be even remotely as efficient as an abstract of equal length, but I admit that the reasons I would bring forward in justification of this evaluation of mine will not be very compelling for someone who has strong feelings in the opposite direction. Nevertheless, I would insist that before one embarks on developing and improving mechanical extracting, the worthwhileness of extracting as such should first be tested. Notice that for a human being abstracting and extracting are processes which probably differ only slightly in the effort required for their performance, whereas for a machine the difference is enormous, indeed so much so that it is very hard to imagine what abstracting in the machine's own words could possibly mean.

The next stage - you recall that we are working backwards - would be the selection of abstracts or reviews (or of the documents themselves - in case that no authoritative abstracts or reviews are available) to be read by the investigator (or by some member of a team of investigators). It is here that one easily discovers great possibilities of effective mechanization, and it is at this stage that a certain amount of progress has been

definitely made. Assuming that each document of a given collection - for the present purposes we shall disregard the quantitative aspects of whether we have to deal with some private reprint collection, the document collection of some medium-sized industrial outfit, the Library of Congress, or some fictional World Center of Documentation containing copies of all the documents that have ever anywhere appeared - has been somehow assigned a set of *retrieval characteristics*, there exists no theoretical problem whatsoever of mechanically procuring a list of all and only those documents that fulfill any Boolean function of these characteristics. If each document of the given document collection is indexed by some subset of the total set of *indexes* (I shall use this term as short for 'retrieval characteristics' and beg you not to be misled by its other connotations), there are innumerably many devices, among them many working ones, that will present you, after so much time and with a cost of so many pounds, with a list of those documents from the collection which fulfill the condition that their index-sets contain, say, the indexes i_{a_1}, i_{a_2}, ..., i_{a_n}, do not contain the indexes i_{b_1}, i_{b_2}, ..., i_{b_m}, contain exactly one of the indexes i_{c_1} and i_{c_2}, contain i_{d_1} if and only if they contain also i_{d_2}, etc. It is a pity that so much ado has been created by some of the workers in the field of mechanization of information retrieval around this theoretical triviality.

I have not the slightest intention of belittling the ingenuity of those inventors who succeeded in constructing working machinery that performs the task just discussed in a time and money saving way. But it must be stressed that thereby only a small portion of the total literature search process has been mechanized, though I would not dare to give you a quantitative estimate of the importance of this portion.

Many different methods of assigning indexes to documents exist, and the discussion of their relative efficiency for manual retrieval is still going on. It is an interesting historical fact that though retrieval of documents indexed by any of these methods is mechanizable, new methods of indexing were developed of which their inventors claimed that they allowed for still more efficient mechanization of the retrieval process *(refs. 3, 4, 5 and 6)*. It even seems that some of these new methods may have their advantages for manual retrieval, which would then be a rather interesting, though not wholly unexpected by-product of the pursuit of mechanization in the field of literature search.

Some of the new methods of indexing are intended not only to improve the economical aspect, measured in time and/or money, of the retrieval process but also its quality. So far, mechanization meant just a speedier and perhaps cheaper procurement of a list of documents which could have been provided also by more orthodox means. The quality of these lists would have been the same. So many vital documents would still be missing, and so many documents would still be recommended for reading that would turn out to be

irrelevant for the investigator's purpose. Can mechanization be utilized for improving the quality of the reference list put out by a literature search system? I personally regard this problem as the most pressing one in the field. I am sorry to state that not only has very little of value been achieved so far in this respect but that incompetent and faulty theorizing has created a pretty thick fog around the issue, so that progress can be expected only after this fog has been dispersed. Since, however, these faulty theories have almost nothing to do with mechanization as such, their failure lying in an earlier level, I shall not elaborate here upon my dogmatic verdict, especially since I have been doing this on other occasions, *(ref.7)*.

Almost everybody in the field would agree that the success of any literature search system depends highly both on the quality of the indexing and on the coincidence of the various index sets with the formulation of the search question by the investigator (or its reformulation by the search expert). It is easy to see that these are not two independent aspects of one and the same problem; consequently, any attempt to improve the first aspect alone in disregard of the other is almost certainly doomed to fail in improving the whole system. This remark is meant as a criticism of all present and even future attempts at *auto-indexing* for retrieval purposes and, more generally, of any attempts to assign to documents an index-set composed entirely of expressions occurring in the document itself and to let it go at that.

Let me propose here a system of auto-indexing which, to my knowledge, has never been publicly proposed before in this form and which seems to me superior to any other system I have heard of. (A certain variant of this system would, incidentally, also provide for an auto-extracting system which would, as I see it, be better than the one indicated above.) Assume that, after some convention or other on what an English "word" is - the exact nature of this convention would be highly important, but I shall disregard this point for my present purpose - has been adopted, we are given a list of the average relative frequencies of all English "words" - I shall again not enter into the innumerable questions arising in this connection, especially in connection with the term 'average'. It would then be possible, for any given document, to rank-order all the "words" occurring in this document according to the value of the ratio of their relative frequency within the document over their average relative frequency. By some mechanically implementable standard or other, an initial segment of this list is selected as the index-set.

As I said before, I believe that this system of auto-indexing is superior to any other I know of - notice that only the first step would be manual, whether this would mean encoding the document into so-called "machine-language" or inserting the document into a mechanical document-reader, if and when such a device is put in production; it is still very easy to point out its many shortcomings. Even if there should exist a statistically strongly

significant correlation between the words with the largest ratio of their relative frequency of occurrence in a given document over their average relative frequency in the language and their membership in the index-sets prepared by the most authoritative indexers - and offhand, in the absence of any empirical tests, I think that some such correlation will indeed exist -, a failure of coming close to a satisfactory set of indexes in 20% or even 10% of all indexed documents would almost certainly disqualify this method for all serious purposes.

You will probably have noticed the close parallel between my present argument and that used above against auto-extracting. The analogy reaches still farther, though perhaps not with the same cogency. Above, I expressed my doubts about the effectiveness of any method of abstraction by extraction, i.e. by presentation of some sub-sequence of the total sequence of sentences of the original document. I would now like to express my equally strong doubts about the effectiveness of indexing, even if done by human authorities, a given document by exclusively using expressions occurring in the document, unless supplemented in some way or other by indications that should be effective in bringing this index set into coincidence with the formulation of the investigator or of the reference librarian who does the mediating transformations. Since this supplementation is clearly beyond the capabilities of presently existing or envisageable machinery, and since having it done by some human post-auto-indexer would probably involve an amount of effort of approximately the same degree of order as preparing an index-set *ab initio*, I regard auto-indexing by any method, including my own, as definitely unsatisfactory, even after the introduction of all kinds of refinements and even when the preliminary problems which I mentioned above but refrained from entering into have been satisfactorily solved.

It is perhaps worthwhile to dwell a little longer on just one specific shortcoming because of its implications, e.g. for machine translation. It is easily conceivable (and I shall therefore not bother to exhibit examples) that a document could contain many occurrences of a certain string of two words - of a certain *digram*, in the "lingo" of the trade - of very low average relative frequency, though each word as such is perhaps quite frequent. Our method would fail to select these two words as indexes, and the fact that their combination is highly indicative of the document would be entirely lost. One could think here - as in a similar situation arising in machine translation - of improving the method by taking into consideration not only the average relative frequencies of each single word but also those of all digrams. But the number of such digrams would probably be of the order 10^8 in English; and no practical method is in view how to arrive at their relative frequency list, in addition to the fact that the theoretical difficulties of preparing word frequency lists would here be multiplied many times.

Let us summarize: Only one stage out of the many stages, of which a literature search is composed, seems at the moment susceptible to mechanization. This step consists in the selection of all those documents (or, for most such systems, in the preparation of the list of all those documents) belonging to a given document collection whose index-sets fulfill any Boolean function over the subsets of the universal index-set of this collection. This step is of sufficient importance to warrant further development of mechanical devices by which it could be carried out with ever increasing efficiency. It is now almost beyond doubt that different devices will be optimal for different sizes of the document collection to be searched. There is no point going here into the details of the various types of machinery in operation or development.

But the importance of this step is strictly limited, and the complaints about the inefficiency of present methods of literature searching will not be allayed by its improvement to any decisive degree. Though I think that something can be done to increase the efficiency of other stages of the literature search process, their improvement will be brought about rather by better organization of the abstracting and indexing services and by the recognition that the adequate provision of such services requires more and better-trained people. Though here and there certain partial steps could again be mechanized, I regard it as chimaerical to expect that abstracting and indexing as such could ever - and by 'ever' I mean 'during the next two or three decades' - be mechanized in a satisfactory fashion.

Many people believe that there exist strong analogies between the machine literature search and the machine translation problem and seem to expect that a solution of one of these problems would greatly contribute to the solution of the other. I myself have dealt at times with both these problems, but in spite of this - or shall I say, just because of this - I consider this belief as almost entirely unsubstantiated, based upon misconceptions enhanced by certain semantical traps, and definitely misleading. Abstracting and indexing seem to me to be processes in which routine plays a considerably smaller part than in translation. In a translation, the sequences of sentences in the target-language will, in general, consist of sentence-by-sentence equivalents of the sentences of the source-language sequence (though occasionally sentences will be combined or split up, for stylistic reasons). An abstract is not equivalent to the abstracted document and does not carry the same information; for certain types of abstracts it is not even true that they carry less information than the original document. An index-set carries no information at all, in any serious, non-metaphorical sense, and the customary declarations to the contrary are, in my opinion, based on no more than carelessness. The assignment of an index-set to a document neither preserves its information content nor part of it; its only task is to provide clues by which this document will be brought to the attention of an investigator.

Since completely automatic high-quality translation seems to me a pipe-dream, completely automatic abstracting and indexing are even more so. I am quite ready to subscribe to the already mentioned slogan that "whatever a human being can do, an appropriate machine can do, too"; but I do this only because I regard the slogan as utterly trivial. At the moment, I am not talking about what machines could do *in principle* but only about what actually existing or blueprinted machines could do, and it is with regard to these that I utter my definite opinions. If someone wishes to write science-fiction about information-processing centres of the (undetermined) future, let him do so and I shall discuss it with him over a glass of beer and even offer some startling suggestions of my own. If he is interested in improving the literature search process today, I would strongly advise him to forget about mechanizing abstracting or indexing. May I add that it is with a good amount of sorrow that I have come to this conclusion which is quite counter-to my temperament and my convictions (never published) of a few years ago.

REFERENCES

1. LUHN, H. P.: "The automatic creation of literature abstracts", *IBM Journal of Research and Development*, 1958, **2**, 159.
2. Ibid, p.162.
3. PERRY, KENT and BERRY: "Machine Literature Searching", *Interscience Publishers, New York and London*, (1956).
4. TAUBE, M. et al: "Co-ordinate Indexing", *Documentation Inc., Washington* (1953/8. *Four volumes).*
5. MOOERS, C. N.: "Zatocoding and developments in information retrieval", *ASLIB Proceedings*, 1956, **8**, No. 1.
6. "The need for a faceted classification as the basis of all methods of information retrieval", *The Library Association Records*, 1955. **57**, 262.
7. "A logician's reaction to recent theorizing on information search systems", *American Documentation*, 1957, **8**, 103, and papers in preparation.

DISCUSSION ON THE PAPER BY PROF. Y. BAR-HILLEL

CHAIRMAN, THE EARL OF HALSBURY: I wonder how much the difficulties stem from the fact that the words of a finite vocabulary cannot have a precise meaning; otherwise we shall find ourselves with nothing to say on most occasions. If you do a computer translation from, say, English into Russian and back again, a proverb like "out of sight, out of mind" may come back as "invisible idiot". This is an extreme example of how the spread of meanings can cause confusion in the use of words.

DR. L. MEHL: It is said in the scripture:- "Perseverave diabolicum" - I persevere in my thinking. Prof. Bar-Hillel said, if I understood him correctly, that transformation of information is a bad concept or at least imprecise, but I only mean by *transformation* of information that the machine - and I think especially the machine for legal argument - can only give us what we have put into it. In other words the machine cannot create information, and I agree with Prof. Bar-Hillel when he says that it is impossible to mechanise completely logical operations. But it is possible I think, to build a machine which is able to answer questions. I must also point out there is a difference of great importance between the *general* problem of information retrieval and the problem of information retrieval for *legal questions*, because the legal provisions (laws, acts, regulations, bye-laws) and the decisions of jurisprudence do not represent an innumerable amount of documents. The concepts utilised for expressing these provisions are generally precise and the problem of abstracting and indexing is not impossible to solve.

When I studied the question I noticed that, contrary to common opinion, the system of law is generally logical and precise. When you read a legal text you have the feeling that it is very complicated; that there are a great number of basic concepts in it; but if we make the effort to analyse and rationalise the juridical problems, the matter becomes clearer and clearer and you will notice, and this is very important, that the fundamental concepts are not very numerous. It is the reason why I think that the questions of information retrieval in law are perhaps easier than in the other areas of knowledge, despite first appearances (if we except the exact sciences like physics and chemistry).

Prof. Bar-Hillel said also - and I agree with him - that abstracting and indexing resist satisfactory mechanization at the present time. Abstracting

and indexing pose difficult intellectual problems, they are human jobs, and there is a great difference in efficiency between manual retrieval and mechanical retrieval. Of course, as Prof. Bar-Hillel said, the essential condition is to have good abstracting and good indexing, but the advantage of a machine for information retrieval, if this condition is realised, is that it is then possible to take the characteristics - the basic concepts of the problem - in any order. On the contrary, when you use a manual retrieval process, for example an index or a table, it is necessary to follow a certain way, and if you make an error on the way it is possible that you will never find good information. I think it is a very important difference particularly when problems are complex. What is also important - and I explain this in my paper - is that if you put a question to the machine omitting part of the data of the problems, the machine answers that it lacks data. That is very important if you now consider the machine for legal argument and if you put a question to the machine in the form*:- "Is it true that .....?" it will answer yes by a series of 1, 1, 1, 1, in a binary system if you are right. If the machine answer no, the position of the zeros indicates why the implication is not verified. You see that the behaviour of such a machine is not quite passive, and that there is a possibility of a dialogue between the machine and its user.

MR. E. A. NEWMAN: In certain respects I am in agreement with Prof. Bar-Hillel's highly iconoclastic paper: in a number of respects I am not. I am in agreement with Dr. Mehl most of the way.

What you want to do to make an ideal information service is to scrap all the books you have, and instead to have records in separate 'pigeon holes', each record containing the organised meaning of that information from assorted books relevant to the solution to one of your problems. Whenever you ask a question you want all the information you need in a neat parcel, an abstract if you like. In certain circumstances the parcels will have material in common. This does not matter. Nor do you mind if you have to repeat parcels for different questions. The labelling system must be such that it gets you directly to the parcel you want. This is a perfect system provided you can completely anticipate every question that your questioners are going to ask. If a librarian knows this he can arrange to have all the answers set out in the right store locations and can organise these correctly. The difficulty in general is that we do not know at all what question will be asked. In a manually operated library, information stored in the librarian's brain is part of the store location system. If the information is incorrectly organised to suit the question, the questioner and the librarian can mutually reorganise it by talking together.

---

* In this form of the question, the search by the machine is easier, and the search time shorter.

Even after this, however, the information obtained usually contains much not relevant to the question. In a case like Dr. Mehl's, the questions that are going to be asked are very limited in kind, further the people who are going to ask them are a very definite type of people. Thus one can make a reasonable forecast of the questions they are going to ask.

In his paper Prof. Bar-Hillel discusses the difficulties that one has in an information retrieval system. One is not getting all the information you might possibly get from all the sources you have. I think most of us find that even all the relevant information is far too much, and few of us would worry much about not having quite all the information we could get. Our real trouble is we get a lot we do not need.

In paragraph 2 of page 792, he defines an ideal system. This seems to me far from being ideal; you push an appropriate pattern of knobs to get the information you want, but do not know what the appropriate pattern is, so you are in precisely the same difficulty you would be in without the assistance. You do not want to have a complicated transformation to get at the right knobs to push: Prof. Bar-Hillel says this 'ideal' system contains ambiguities, vagueness and obscurities. It seems to me that the definition of the 'ideal' system contained no ambiguities, no vaguenesses and no obscurities. It was all very clear indeed, it was merely no good as a system.

Prof. Bar-Hillel says that one difficulty is that different people want different information. This is very true. They might be solving different problems or have a different clue. Because of this they must in fact sort out what they want; but it is not true to say that the information is not necessarily either relevant or not relevant. To the problem they want to solve, some of the information is precisely relevant and some is not relevant. One surely cannot imagine a situation where it is partially relevant. It either fits the pattern or not. Of course, they may have some difficulty in deciding whether relevant or not but that is another matter.

On page 793 he does ask how a questioner knows which knobs to push. He then asks a quite different question which he says is the same, that is: how do you make these knobs get to the record you want? Could he explain how these questions are identical?

On page 798 of his paper Prof. Bar-Hillel implies that the major problem with library retrieval, that of allowing for context is also that with language translation. On page 799 he implies that library retrieval and language translation have little in common. What does he really think?

MR. P. E. TRIER: I would like to comment on Prof. Bar-Hillel's delightful proposed system of auto-indexing, based on the excess ratio of frequencies of key words above the statistical mean. I shall not shoot down the system, because Prof. Bar-Hillel has done this himself, but I can illustrate the

shooting-down process by an example: 50 years ago E. W. Hobson wrote a classic treatise on the Theory of Real Functions. Littlewood, more recently, was heard to remark, perhaps apocryphally, that in the whole work he had only found one single mention of real functions, in a footnote to the preface. It read as follows: a lecture is the place for the sort of provisional nonsense whose real function is to open the door to systematic study.

DR. J. PATRY: Prof. Bar-Hillel wrote in his paper that the documents are relevant or irrelevant to the problem you are working on. It is more useful, I think, and more important to say: Some documents are useful and others are not. They are relevant to a problem, but they are not always useful to one person. It is much more difficult to mechanise the search of papers from the second point of view, because it varies from one person to the other, and for a particular person it varies with the time. On the other hand, I agree with Prof. Bar-Hillel that auto-indexing is very difficult, because of that personal influence. The indexing must be done by a human being and not by a machine, because the importance of the different parts of the contents vary from one documentation centre to the other.

MR. H. W. GEARING: The coding problem for retrieval would seem to be essentially similar to the statistical problem of definition and classification. Where the definition does not automatically carry with it some measurable characteristic as, for example, when we classify children into age-groups, or postal or administrative districts according to some latitude or longitude definition, then we have to decide, before drawing up our code, what will be the most convenient code headings and their most convenient sequence, for retrieval, or for subsequent presentation in a summarised form. For example, in the comparatively trivial case of a sales analysis in a large company, we are presented with a large number of possible bases for classification of our data, in order that we may subsequently be able to retrieve it to answer the sort of questions that the managements in the different departments ask. The problem is similar again in the case of the filing of machine drawings, where the component described in the drawing may be used in a multitude of machines. It becomes more, perhaps much more, complicated in the case of filing patent specifications.

It would seem, in the field of literature and law, that if we could have a committee of librarians who could list the attributes under which any piece of literature or law could be classified, it would only remain to devise a logical sequence of codes, within each attribute. For example, the attributes might be under the headings of historical, geographical, industrial, legal, and scientific with subdivisions of scientific

classifications, and so on. The setting-up of such a system, after a survey of a sample of the literature to be classified, should not, I suggest, be difficult. The difficulty would arise subsequently in finding enough qualified human beings to read through the mass of literature in detail, particularly those books without index, so as to code all the paragraphs of possible future interest under the attributes involved. This problem in a smaller form has already been met by those who set up coding systems for economic and statistical information. It is essentially one of training and supervising the people concerned. As I see it, in literature and law, the principles have already been established in the past experience of librarians, of statisticians, and, in the particular case of the law, by the work of those who like your noble predecessor, Sir (Reference to the Chairman, The Earl of Halsbury) have prepared extensive summaries of the statute and case law, as it stood at a given date, without which the law students of today would make very poor progress. In commercial statistics, we have met a further problem! We find that we have to train those who ask questions of the tabulating room, to frame their questions in such a way that they can be answered from the coding system. Presumably in the case of literature and law also we should have to train the people, who came to the library, to ask their questions in an appropriate form, but where they did not know how to ask them in that form, we could probably find out the codes in which they were interested by a suitable process of interrogation.

MR. W. S. ELLIOTT: Why are we talking about dealing with the literature, the scientific papers, that have been produced in the past? Let us forget them. Could we get the professional institutions together - the institutions which are going to publish serious papers, worthwhile papers, in the future - and get them to do something from now on about a universal code which can be used some time in the future when perhaps we will have machines of sufficient power to use it.

PROF. BAR-HILLEL (in reply): There are about a dozen specific comments I should have to react to in the five minutes allotted for this purpose. Therefore, I hope to be forgiven for not reacting to all of them and not doing justice to some of the others.

(1) With all due regard to Dr. Mehl's authority, I see no attempt on his side to justify his claim as to the basic simplicity of juridical language and theory. Until I see such an attempt carried out to some serious degree, I intend to remain skeptical and to continue regarding legal problems to have at least the same degree of complexity as the average problems for whose solution information retrieval systems are set up.

(2) Whether too much suggested reading material is the major drawback of existing information retrieval systems, as Mr. Newman thinks, or too little,

as most workers in the field tend to believe, is not a question that can be uniquely and simply answered. I am sure that there are situations where you will be critical of a system that does not supply one with some vital reference because somewhere the indexing system has misfired, and that you would, in such situations, prefer a system that provides you with this reference together with a certain amount of useless material.

(3) It seems that Mr. Newman did not take seriously enough the quotes around "ideal". I expounded this "ideal" system not only to knock it down -- and I am grateful to Mr. Newman for helping me in this task -- but also because I thought it would be of some pedagogical help in explaining certain fallacies in the thought of those who believe in the possibility of far-reaching over-all mechanization of information processing and retrieval. I am still not sure whether Mr. Newman thinks that I did not succeed in exploiting this artifice well-enough or whether I overdid it.

(4) I find it very difficult to understand why Mr. Newman is so sure that documents must be classified as either relevant or irrelevant to a certain problem, as well as his problems in imagining a situation where you would want to say that a document is partially relevant. Assume you are interested in diseases of cats. Is a document discussing diseases of dogs relevant or irrelevant to your problem? Is it not more helpful to regard it as partially relevant in the sense that for X's interest in the problem it is relevant and for Y's interest irrelevant (as well as in other senses)?

(5) On page 798 of my paper, I say indeed that library retrieval and language translation have in common that certain initially (because of semantic traps) attractive methods fail to work satisfactorily in either. Is one really entitled to say that they have much in common, thereby contradicting what I say on page 799?

(6) I am grateful to Mr. Trier for his help in shooting down the auto-indexing business, but I should like to add, in all fairness, that Mr. Luhn with whom this whole idea originates is aware of the occurrence of such cases as mentioned by Mr. Trier and proposes to deal with them by assigning special weight to terms occurring in titles. However, adding this refinement and the innumerable other ones that will be needed to meet other sources of failure will in all probability make the resulting system too complex and costly to be of practical use. The original system, on the other hand, is elegant and quite cheap -- but not good enough.

(7) I have no objection to Dr. Patry's proposal to use 'relevant for' as a semantical term, denoting a binary relation between a document and a problem, and to use 'useful to...for- - -' as a pragmatical term, denoting a ternary relation between a document, a person and a problem. In these terms my point was indeed that auto-indexing is so unpromising because it must be extremely difficult to weight the index terms in accordance with the usefulness of the indexed document of some group of prospective users

for their prospective problems, much more than weighting the index terms in accordance with some user-abstracted criterion of relevance.

(8) Mr. Elliott's proposal -- not a very original one, as he is doubtless aware -- has to be put before UNESCO or perhaps the General Assembly of the United Nations where it will probably suffer the same fate as the related proposals of using an International Auxiliary Language for scientific publications. I can only pray that both proposals will eventually be accepted, but in the meantime we had better go on and investigate carefully to what degree translation and information processing and searching can be mechanized, trying to correct man's irrationality by the use of machinery, a procedure which might not seem very attractive to some but is still practically more or less necessary.

SESSION 4B

PAPER 3

TO WHAT EXTENT CAN ADMINISTRATION BE MECHANIZED?

by

J. H. H. MERRIMAN and D. W. G. Wass

BIOGRAPHICAL NOTES

Mr. J. H. H. Merriman was educated at King's College School, Wimbledon, and King's College, University of London. He obtained his B.Sc. (Hons.) in 1935 and did Postgraduate Research at King's College London obtaining his M.Sc. in 1936.
He entered G.P.O. Engineering Department, Radio Research Branch, Dollis Hill, in 1936 and was associated with development of long distance radio communication systems. He was Officer-in-charge Castleton radio research station 1940-8, and from 1948-53 in the Office of Engineer-in-Chief G.P.O. and responsible for microwave system development and planning. In 1954 he went to Imperial Defence College; in 1955 he became Head of G.P.O. Engineering Department O & M unit. In 1956 he joined H.M. Treasury, and is now Deputy Director, Organisation and Methods Division.

Mr. D. W. G. Wass was educated at Nottingham High School and St. John's College, Cambridge, where he obtained his M.A. in 1947. He served in Admiralty as Scientific Officer 1943-45, and entered H.M. Treasury as Assistant Principal 1946. He was Private Secretary to Sir Wilfrid Eady 1948-50, and Head of Civil Service Establishments Manpower Statistics Unit 1950-53. He was in the Overseas Finance Division of Treasury 1953-57, and is at present on a Commonwealth Fund Fellowship at Princeton University studying "U.S. Monetary Policy".

TO WHAT EXTENT CAN ADMINISTRATION BE MECHANIZED?

by

J. H. H. MERRIMAN and D. W. G. WASS

SUMMARY

MOST current examples of automatic data processing (A.D.P.) may be regarded as slightly advanced mechanical models of tasks performed by clerks. The paper examines the extent to which a less mechanistic approach may be possible and suggests limitations that may be imposed not only by human limitations but by difficulties of correspondence and significance between machine and manual situations.

---

1. INTRODUCTION

LET us assume that automatic data processing (A.D.P.) can do the things that we are planning for it to do at present, such as payroll, stores accounting, and statistical analyses. There will, of course, be many problems to be solved before these tasks can be regarded as satisfactorily completed, and before we can speak with confidence out of experience. But these problems do not appear to have any fundamentally insuperable content. The difficulties are manmade rather than intrinsic. They originate in part from the difficulty of adjusting the organisms of office life to new rhythms, new environments, new relationships, in part from imperfect understanding and appreciation of the power and range of new techniques, and in part from a lack of perception of the limitations and deficiencies of these systems. We may reasonably suppose that, during the course of the next five years, these difficulties will be overcome and that, throughout Government Departments and Industry, there will be a growing number of installations at work on these jobs. With this perhaps over-simplified premise, it is not too early to start thinking about a possible future form of A.D.P. in Government Departments in, say, ten or fifteen years' time.

2. THE IMMEDIATE FUTURE

2.1 *The move towards integration*

Although computers were developed as an aid to scientific work, and as such were general purpose, maid-of-all-work mathematical slaves to their scientific masters, most commercial applications have up to now been planned as single purpose installations. In Government Departments we are, by and large, installing equipment which will do only one or, at the most, a few jobs. We have installations for pay, and to these we are adding statistical jobs. We have installations being planned for stores accounting, and to these we are adding provision and purchasing functions. These additions, however, are being regarded as secondary objectives and in most cases the installation is regarded as having a single main purpose.

We are, therefore, likely to see in the immediate future a movement away from the concept of single purpose automatic data processing installations to installations or systems of installations which, in the first place, will be multi-purpose and, in due course, integrated. At present though, we cannot achieve integration without human assistance and the written word.

2.2 *Prerequisites for integration*

In an automatic system, integration is achievable only if data are able to be exchanged readily and freely between the component parts of the system. One prerequisite of integration, therefore, is the existence or availability of a common language. In its simplest form, this language could be commonly agreed code marks on pieces of paper; in a slightly more advanced form, it could be agreed standard forms of punched cards and codes. In a still more advanced form, the common language could mean completely standardised programmes, order-codes and magnetic tape codes. Although the prerequisite of a common language is a desirable one, it is not essential. Civilisations have managed to exist and to communicate without such a common language, even though there have been excursions from time to time into the possibilities of one. If a common language is not practicable, then automatic translating devices (or dictionaries) must be considered. These seem to be inescapable premises on which to build integration.

The extent to which it is possible to achieve integration is, however, controlled by two other factors. The first of these is the ability to transmit data from processing centres in one part of the country to another without fear of error, and the second is the extent to which it is economical to build bigger and faster central systems. Up to a certain stage, the bigger the computer the cheaper. Beyond this point, however, there is a line of diminishing return, both in terms of price to be paid for the apparatus and the extent to which it becomes humanly possible to build in the necessary intelligence to manipulate very large installations and to

programme them. We may suppose that the price factor is not a substantive one. Changes in technology may reduce its importance.

The ability of most human minds to keep pace with the increasing complexity of integrated data processing may, however, in the end, set limits to the size and complexity of the installations themselves. If we are, therefore, to imagine large complex multi-purpose integrated data processing systems, we must imagine them to be serviced, to an increasing extent, by separate installations which will analyse the operations of the integrated system, determine the most appropriate operating conditions and which will, to some extent, relieve the burden of programming by automatic access to inbuilt programming routines. It is from this concept that we would expect major developments in overall economy and efficiency to stem as these "monitoring" computers assess the efficiency and failures of the processes being undertaken in the main system, and thereby "learning" from these experiences. It is from studies of the subjective probability of partial success or partial failure in various programming situations and in the way in which the machine complex may react to certain circumstances that the worth-while application of A.D.P. to tasks of management may evolve.

2.3 *The Technical Problems*

These developments for the immediate future depend upon techniques which, generally speaking, are available at present but need further development to bring them to points of commercial and economic reality. These techniques are:-

(i) The creation of data as an automatic by-product of other human activity. Techniques are required for the creation of data in a form which can be readily and directly assimilated by machines without intermediate translation processes. Techniques such as automatic character recognition are already possible, but clearly more work has to be done to convert them into thoroughly reliable commercial tools. In many situations, however, an organisation may have virtually complete control over the data at their point of origination and their final point of reception. Under such circumstances, therefore, it is necessary to consider whether automatic data creation by the recognition of written characters is necessary, or whether greater overall economy could not be achieved by the creation of a parallel chain of data in machine language automatically at the same time as the data is being created in a visible record form. There seems to be no reason why machine development should be constrained by the complexity of arabic or latin script when simpler coded forms capable of being printed at the same time as the arabic or latin characters would enable simpler machine processes to be adopted.

(ii) Techniques for the abolition of the written word as a means of data transfer from point A to point B. Analysis of the internal organisation of offices shows that the cost of letter preparation, enveloping, sorting,

franking, distribution in the office, filing, outweigh grossly the cost of transmission in written form through the mail or by private carrier. Great potential economy, therefore, exists in the abolition of written records at all points except those where access to information in written form is essential for the conduct of the business. Before these economies can be secured, however, the technicalities of data conversion, data transmission and data recreation require to be progressed further into commercial reality and economy.

(iii) Analysis of the A.D.P. systems being planned for Government Departments shows that over 30% of the cost of these installations is accounted for by programming. The essential technical problems of automatic programming may be said to be overcome, but much remains to be done before automatic programming and coding is an acceptable commercial device.

(iv) There is a growing demand for systems for very large information storage and associated with these systems versatile information retrieval. Before such systems may be realised in practice, considerable development is required on techniques of storage and the philosophy of its retrieval. Many techniques can be seen to be possible in this field. As yet, no problem has been sufficiently clearly formulated to enable the worth-whileness of the various techniques to be estimated.

3. LONGER TERM PROBLEMS

3.1 *General*

We must now assume that the technical problems posed in the foregoing paragraphs have been solved, and look to the impact of these developments upon even longer term planning, so that the possible form of A.D.P. systems of 10/15 years hence may be assessed and courses of research and development postulated. By such examination, we may be able to lay down, if not the philosophy of systems design of the future, then at least the constraints within which those system designs may operate.

3.2 *Problems of Symbolism*

It seems possible that the growing complexity in A.D.P. systems will lead to a situation in which the relationship between symbols as they exist in the machine and the symbolism of a real world will be progressively more difficult to trace. This is no new problem. Business is, after all, now generally conducted by the making of marks upon a piece of paper. In most cases, these marks bear a definite relationship (usually a unique one) to objects or symbols in the real world. For example, descriptions upon a ledger relate to physical identifiable items in a stores bin. The problem is, however, made more complex in the A.D.P. system of the future, since

the symbol relating to a particular external object may no longer be humanly identifiable without the aid of a machine interpretation process. The relationship between the two is therefore dependent upon the behaviour of the machine, which in turn is dependent upon the programmer and the designer. Knowledge, therefore, of the characteristics of objects both in the real world and in the machine will tend to become restricted to a relatively few people. This will create difficult problems of access to that information under certain conditions and may therefore impose, of itself, constraints upon the universality of application of these machines.

3.3 *Problems of Interrogation*

The greater the volume of information held in mechanical or electronic form, the greater will be the task of making that information available in cases other than those where a direct "Yes"/"No" answer is sought. Before information can be held in machine code form, it must be translated into that language. The translation process depends critically upon the extent to which there can be a unique one-one relationship between the external meaning of the information and the machine meaning. In many cases, information to be encoded in this way is not capable of being defined uniquely on a one-one correspondence basis and there, therefore, has to be a subjective assessment of the information content of the data subjected to filing. Clearly then, these data cannot be satisfactorily interrogated unless the same subjective translation process is followed. If it is not followed, then the interrogation can, at best, only be partially successful. Partial success, however, may not be recognisable as such. It may, therefore, be increasingly difficult to determine the degree of precision with which interrogation is being answered unless the form of the questions is rigidly related to the form of the filed data.

3.4 *Problems of Rigidity, Resilience and Feed-back*

Clearly, the greater the absorption by A.D.P. of an organisation, the greater will be the dependency of that organisation upon any constraints imposed by the A.D.P. system. In practice, most systems will be designed and programmed to meet the specific requirements of the organisation. If these requirements change, there is no guarantee that the A.D.P. system will be able to accommodate them. Unless the A.D.P. system, therefore, is a resilient one, it may be found that it is imposing constraint upon organisational changes that are admissible. Either, therefore, A.D.P. systems will have to be made sufficiently flexible to accommodate the estimated degree of change, or else a knowledge of the constraints of the system will have to be fed back into administration and brought into the consideration of changes of policy.

3.5 *Human Problems*

The development of A.D.P. systems will lead to a gradual absorption of the routine, easier and more mechanical tasks. These tasks have, in the past, been one of the training grounds for the less routine, more difficult, more advanced operations. Training techniques will, therefore, have to be modified to take account of this changing situation. This will exist, not merely in the mechanistic field of A.D.P. system operators, but will affect those coming into contact with A.D.P. systems.

There is also a potential human problem in determining how best to organise an A.D.P. system to meet the unexpected demand. Whereas, in the past, it has been possible, for example, to aggregate a considerable volume of human experience in clerical and executive tasks and to evoke from that experience considerable skill in dealing with the occasionally unusual, in the future experience in the routine operation will be contained within the machine system in machine language. The difficult problem will, therefore, have to be solved (if it cannot be solved by the machine) de novo. This problem may become acute for example in very large information storage retfieval systems at times when it is not known in precise terms what question should be asked of the system and, indeed, the degree of precision necessary cannot be reached until there has been considerable browsing. The provision of the machine equivalent of browsing (i.e. the tentative, testing, probing search that is sometimes divergent and only in moments of rare success, convergent) seems difficult to visualise.

Associated with all these problems is the problem of maintaining interest in a fully-mechanised relatively highly reliable system, so that the necessary degree of skill may be exercised at times of crisis and unreliability.

Underlying these 'human' problems of A.D.P. is, however, the greater problem implicit in all mechanisation, that of the growing dependence of the human race upon the complexities of technology, This dependence is defensible and, indeed, acceptable insofar as technology provides a greater degree of perfection or satisfaction to human ambitions. If situations develop where this is not so, then not only is the position indefensible but it may also produce worse result than had mechanisation not been attempted. In the field of A.D.P. it is possible to visualise a situation where dependence upon devices in a particular sector of human activity becomes complete and our premises of providing a greater degree of perfection or satisfaction of ambition were satisfied. A change of external environment, caused let us say politically, and not foreseen by the designer of the device, may then be supposed. The device is useless. But worse; the experience of the immediate past cannot be tapped. The situation may be retrieved only by a painful manual recollection of events leading up to the externally imposed change. The more perfect the machine, and the more widespread its acceptance, the

more explosively imperfect is the overall situation in the consequences of failure under conditions not foreseen by designers and administrators.

4. THE "UPPER" LIMITS TO A.D.P.

An attempt at this point in time to predict the extent to which A.D.P. is likely in the long term to supplant the human agency in 'Administration' would be most hazardous. It is perhaps possible, however, to make some assessment of the area within which, for one reason or another, a purely mechanistic agency would be unsuited. Broadly speaking, in any organisation, the higher one goes in the hierarchy the less routine becomes the type of work. By the expression "less routine" is implied "the less are decisions taken according to some formulated rule". In the highest reaches there may be a complete absence of formulated rules and many of the decisions are made "on merits", that is to say all the relevant considerations are taken into account and a conclusion reached to which, in the opinion of the person forming it, the considerations point. The decision is thus a subjective one and depends on the weight attached by the person concerned to the various relevant factors. Once the relevant considerations have been elicited and the weights to be attached to them assigned, the process of forming a conclusion is relatively simple.

The decisions of the higher administrator, therefore, tend to differ in two respects from those of the clerk:-

(i) there is far less scope for referring to an authority e.g. a code of rules; indeed, even where reference to an authority is made then this is usually oblique and indirect. (Were the references direct and unequivocal, it seems doubtful if the case merits the attention of administration), and (ii) there is less repetitive work.

Thus the two main features of clerical work which make it amenable to A.D.P. tend to be absent from higher administrative work. The actual process of administration may not be entirely unsuited to A.D.P., for, since it must have a logical basis, it may be executed by a computer. But the extent to which it may in fact ever be performed mechanically as an integrated process would seem to be limited by two factors:-

(a) there is an element of subjective judgment in every administrative decision as to what is a relevant consideration; and

(b) there is a further element of subjective judgment as to what weight should be attached to each relevant consideration.

It is difficult to conceive of a computer - of finite size - having the capacity to weigh what are normally called "imponderables" and balance them against other imponderables, unless a similar exercise with the same imponderables in the equation had previously been carried out. Yet this is what in the main higher administration amounts to.

But against this the possibility must be admitted that A.D.P. may provide administration with some most powerful aids to judgment. A good deal of the subjective judgment to which an administrator is driven by defects in his present equipment could be made objective if the means to processing were available. For example, decisions based on purely economic considerations are likely to be made more soundly and with greater realisation of their consequences as the theories of econometrics are developed and the formulation of accurate numerical equations in economics and their solution by A.D.P. becomes possible.

But in some ways this is only another way of saying that wherever possible objective judgments should replace subjective judgments. This is truth which would command general assent. It remains equally true, however, that in a world peopled by human beings the judgments which bear upon them inevitably have a subjective element, and that element can never safely be devolved upon a mechanistic agency, for to so devolve would deny in the last analysis the spiritual nature of man.

DISCUSSION ON THE PAPER BY MR. J. H. H. MERRIMAN AND MR. D. W. G. WASS

MR. A. J. BROCKBANK: My interests in the computer field are connected with the practical problem of applying the use of a computer installation to a commercial organisation, and in that context I have an appreicable interest in what Mr. Merriman has just had to say. I was interested to have clarification of his definition of the "administrative level". I would respectfully suggest that so far as industry is concerned his definition is not wide enough. There are significant groups of people employed in industrial organisations who are classified as occupying administrative positions but are not concerned with these higher levels to which Mr. Merriman has made reference. I accept Mr. Merriman's contention that at the policy level the possibility of computers replacing intellect is unlikely to come into operation for some considerable time. In regard to the middle level group, I am, however, being driven more and more to the conclusion that in the relatively near future the utilisation of computing techniques is going to affect such groups very significantly. The ones concerned are those whose job is to bridge, from their experience, the lack of available information owing to inadequacies in existing data processing systems. In my view with the ability to have all the facts relating to a given situation before us, then the need for this human link becomes unnecessary.

The other aspect of this matter which I think is important is that in certain spheres of industrial and commercial activity, where even with all the facts available human judgment is exercised, a computer can be programmed to adequately take a number of decisions in this field.

On the question of rigidity, one aspect of computers which at least has impressed me, and I hope I have not here drawn a false conclusion, is that they appear to enable one to change organisational approaches and systems to meet changed requirements far better than is possible with existing mechanised type of equipment.

MAJOR E. J. GUTTRIDGE: When reading Mr. Merriman's paper I hoped to find a complete specification on Automatic Data Processing Equipment for industrial and commercial use in the next ten years.

Obviously this would have been a very bold action on his part and I would therefore like to congratulate him on the many sign posts which he erected.

The paper does create an impression that during the next ten years there will be a progressive move towards centralisation of Data Processing. No doubt there is unintentional. There is sufficient evidence to indicate that a movement away from over centralised administration is beginning and it is essential that designers of data processing systems take note of this in order that the smaller organisation is not neglected. The remote communication of input and output will only provide a limited solution.

It is my opinion that the major development during the next phase will be devoted to the requirements of small decentralised organisations. The first steps to be taken are of a systematic nature as distinct from sheer hardware development.

Most of the more novel technical advances are still on a research basis and the near future should be used for the exploitation of established techniques coupled with a truly integrated approach to the systematic requirements of business and industry.

MR. C. STRACHEY: I would like to take up a point Mr. Merriman made at the beginning of his paper, though I do not think he mentioned it again in his introduction. He said that he was going to assume that all the data processing problems we were giving computers at the moment were solved, and that they did not present any insuperable difficulties; he said, I think, that all the difficulties were man made. I am not sure that I agree with this, I think there are some data processing problems, particularly those concerned with timetabling, or shop loading, or works control, for which no adequate method of solution yet exists. These problems are essentially combinatorial and their complexity increases very rapidly with their size. If we consider the problems of organising any sort of time-tabling procedure - say loading machine tools - so long as the unit is small a human brain can make quite a good job of it, though the methods by which it works are not at all clear. As the organisation grows there comes a moment when the time-tabling gets beyond the capacity of the man who is doing it. It is then necessary to increase the proportion of people operating quite disproportionately because no one person can comprehend the whole field of the problem. The trouble really is that the complication goes up exponentially and for large problems becomes quite out of hand. I do not think we yet have any satisfactory way of tackling this sort of problem. Almost the only method suggested so far is linear programming, but for this type of problem it involves matrices which are far too large for any reasonable sort of machine to deal with. The whole of the area seems to me to be one of considerable intellectual difficulty and I am not at all convinced that we know how to tackle these problems yet.

DR. S. GILL: I think I would agree with Mr. Brockbank that we are more likely with electronic computers to be able to adjust the system to changing circumstances than with older forms of equipment. But I think that the problem will arise from the extension of the mechanised system to cover more and more of the enterprise. This of course will lead to the problem of rigidity which Mr. Merriman has pointed out. And I think it is important that when we do mechanise a large system we should arrange the mechanisation so that suitable data is printed out which can be readily comprehended by humans, so that some one, or two or three people preferably, can keep themselves familiar with the progress of the enterprise and are in a position to take over control if the mechanised arrangements look like failing under a certain change of circumstances.

With regard to the problem of centralisation which Major Guttridge raised, it does seem to me that the problem of communication between a number of different places is likely to set a limit to the practical size of a computing installation, and this factor, together with the factor of reliability, seem to be the two things that will limit the useful size of computers that we are likely to be able to build. The two factors mentioned in the paper, namely, the price factor and the difficulty of programming a large job on a computer, seemed to me to be somewhat irrelevant. In fact, if anything, I would say the price factor operates the other way, and tends to favour a larger installation.

MR. B. R. ASTON: I would like to start with the first point Mr. Merriman made, as to what is an administrator. The way we are looking at this is that an administrator is someone who makes a decision, and he is going to make that decision on the basis of information that is handed up to him. When he makes the decision he will then hand an instruction down, and very rapidly we feel we are getting a situation very similar to the closed loop of a control mechanism - certainly a loop effect, and I would like to ask Mr. Merriman what he thinks about this idea.

MR. R. BENJAMIN: I have a small comment on Mr. Strachey's remark, concerning the point where the loading of the machine "gets out of hand". It gets out of hand quite as much for the man as for the computer. In each case the number of possible solutions in the matrix becomes excessive, and some sort of short cut is required. However, it is fairly easy to think of reasonable short cuts: as a first approximation, one could classify manufacturing tasks in order of importance or difficulty, and classify the machine-tools in order of desirability of use - that is roughly the inverse of versatility -, so that if the most urgent problem is solved with the least versatile combination of machine-tools that are otherwise sensible for doing the job; you will then have preserved the maximum flexibility in deploying your remaining machine-tools for the next urgent job, and so on. By doing this sort of thing

you can keep the growth of complexity with problem size roughly linear- instead of a square law, and you can still get a very good approximation to the best loading; although I admit you cannot guarantee to get the best possible loading of your workshop. It is quite feasible to refine this general approach to take other factors into account, without getting excessive complication in your computer programmes.

MR. M. A. WRIGHT: Mr. Benjamin says that short cuts are needed to solve machine loading problems: I agree. He says, that one method of doing this involves classifying work-pieces and machines. It is possible to imagine that work-pieces could all have different prescribed priorities and if so Mr. Benjamin's suggestion would lead to a useful solution. If some artificfal classification system were introduced, the number of possible solutions of the problem would be reduced; but the restrictions imposed by using the classification system would exclude the possibility of finding a large number of solutions. The remaining solutions could all be very inferior to the true optimum solution. Thus, the method of classification may be very important.

It is interesting to note that methods of classification are important in many other applications. They offer prospect of a simple, but often a poor, solution unless the appropriate classification method is chosen.

In their paper Mr. Merriman and Mr. Wass say (see p.817) that there is an element of subjective judgement in every administrative decision as to what is a relevant consideration. I am not very clear on what this means so to begin with I have assumed the following definitions:- a relevant consideration is one which affects the decision and a "subjective judgement" is a judgement on what will affect the decision, which is made without knowledge of the decision. I would guess that decisions based on this kind of subjective judgement would be likely to be wrong. However, our decisions are often right and I suspect we are able to make them because we have some experience of similar problems. We may use a classification technique and may not be conscious of the processing. For example, we may make a tentative guess at the answer so that we can check whether we have included all the "relevant considerations". If we have not, then we change "considerations" and repeat the process. Dr. Selfridge has explained a machine which works on similar principles. His machine does not require an "infinite" store so it is possible that the selection of relevant considerations may not limit the application of ADP to administrative decisions as is suggested in the paper. The other limiting factor, mentioned in the paper, is the "attachment of weights" to relevant considerations. It is possible to measure some considerations either absolutely or relatively; but where human reactions are involved, this is difficult. But even these difficulties may not be insoluble even with relatively small computers. If this were, perhaps the main difficulties in using ADP for decision making would be the time taken by a machine to gain experience.

MR. J. A. GOSDEN: I would like to ask Mr. Merriman about his statement that 30% of the cost of an installation is accounted for by programming because he claims that the techniques of automatic programming will make large savings in the costs of programming. Before commenting on this statement it is necessary to be sure that programming is properly defined, putting a task onto a computer takes three phases

(a) Defining the Task.

(b) Planning how the Task is to be effected.

(c) Translating the Plan into Computer Language.

It is possible that Mr. Merriman includes all three of these in his estimate of costs. I would like to make a distinction between (a) and (b) which are Programming, and (c) which is Coding. I would say that the automatic programming that exists and has been discussed here this week, is only automatic coding. The problem of automatic programming is still largely unsolved. Systems such as FORTRAN have introduced techniques to optimise the organisation of a single programme within itself, but nothing has yet been done about deciding how to organise a suite of programmes to tackle one task. The high costs are associated more with phases (a) and (b) rather than (c), and I do not think that a significant reduction of their costs is near at hand.

Finally, I should like to reinforce the remarks made by Mr. Strachey about the complexity of combinatorial problems. There has been a large amount of work done on one particular problem, playing chess, and it has not yet been possible to attain the ability of an average human chess-player, and the most recent programme *(ref. 1)* expects to take 8 - 10 hours making a move in the middle game. Machine shop loading is a similar type of problem. Humans can get a solution that works but is probably a long way from the optimum.

MR. J. H. H. MERRIMAN (in reply): Three points of substance have been raised in the discussion:-

(1) Rigidity and flexibility in A.D.P. systems. To say that "computers appear to enable one to change organisational approaches to meet changed requirements" is to overlook the growing number of systems that are being built around, not general purpose, versatile computing centres, but special purpose, custom-built units, the logical content of which is closely matched to the tasks they have to perform (for example, on-line control of continuous flow chemical processes, air-line seat reservation systems, air traffic control assistance, banking). In such cases, unless the direction and rate of possible future change can be guessed at

REFERENCE:

1. NEWELL, A., SHAW, J. C. and SIMON, H. A. Chess-Playing Programmes and the Problem of Complexity. *IBM Journal of Research and Development*, 1958, 2, No. 4.

reasonably accurately, and provision made now to accommodate this change when it occurs, the A.D.P. system may in fact be as difficult to convert as were the earlier broad gauge railway systems to narrow gauge. General purpose A.D.P. systems, however, with their usual large volumes of logical redundancy present somewhat less difficulty provided programming changes can be accommodated.

(ii) The pressure of A.D.P. to centralise processing. Moves that there may have been away from centralisation during the last two decades have now mostly halted or reversed as a result of the growing economic pressure in favour of centralisation - at least of the processing element. To decentralise, say the 'logistical' control of a large organism is to predicate such a large increase in two-way, near simultaneous, closely dependent data flows between the component parts for any, say optimising, operations as to make the operation impracticable by presently foreseeable techniques at least. Centralising, in contrast, calls for, in the main, simple, (though bulky) unidirectional independent data flows.

(iii) Administrative decisions and their nature. To attempt to assign to such decisions a value to indicate "rightness" and "wrongness" must itself be a subjective process and incapable of expression in other than transient terms. A more fundamental difficulty would seem to be that because the value scales tend to be transitory (in that they depend in turn upon human reactions) the labour of encoding, both logically and in machine terms, isn't worth it.

SESSION 4B

PAPER 4

POSSIBILITIES FOR THE PRACTICAL UTILISATION OF LEARNING PROCESSES

by

DR, S. GILL

BIOGRAPHICAL NOTE

Dr. Gill was employed at the National Physical Laboratory during 1946 to 1948 on punched card computing and the design of the Pilot ACE. He later studied programming with Drs. Wilkes and Wheeler at Cambridge where he become a Fellow of St. John's College. During 1953 and 1954 he visited the Massachusetts Institute of Technology and the University of Illinois, and he is now Head of the Computing Research & Service Group of Ferranti Ltd.

POSSIBILITIES FOR THE PRACTICAL UTILISATION OF LEARNING PROCESSES

by

DR. S. GILL

INTRODUCTION

IN the beginning, practical applications of high-speed computers were limited by a number of factors including the following:

1. available input and output equipment.
2. capacity of storage devices.
3. computing speed.
4. the practicability of preparing programmes.

A considerable amount of research has now been carried out on all of these factors, and technical advances have greatly lessened the limitations imposed by the first three. We are moreover within sight of even greater achievements in these directions. However, although a great deal of work has been done on the subject of programming, it seems likely that before long there will be many extremely promising potential applications of computers that we shall be prevented from realising solely because of the difficulty of preparing the programmes.

Language translation is making great strides at present. The possibility of applying computers to this task is as yet somewhat dependent upon the available storage devices and on the means for feeding text into the system. However, with the development of character reading machines and of larger stores, language translation will be limited only by our ability to write programmes defining the translation process. Crude programmes already exist, but there will be tremendous scope for their refinement.

Similarly traffic control (particularly the control of air traffic) is potentially a very important application of computers whose realisation is being delayed principally on account of the complexity of the programming. Suitable computers with the necessary external data links could soon be made available if it could be shown that effective programmes could be written.

In the field of business and industrial management, it is conceivable that a computer could play a large part in making policy decisions, taking

into account a great many relevant facts and combining them with a much greater degree of precision than unaided people can. In this case it is quite possible that existing methods of storage would be adequate and that the problem would not take many hours on existing computers. The stumbling block is the production of a programme embodying the rules for making the appropriate deductions from a great variety of facts.

Looking a little further into the future, it is perhaps not entirely fanciful to picture a computer in control of a surgical operation. No new techniques of instrumentation would be needed in order to feed information to the computer about the patient's respiration, blood circulation, body temperature, etc. The computer could take into account many more such measurements than can at present be made by the surgical staff and could wield (by means of servo-mechanisms) a large number of surgical instruments. Provided that the necessary programme could be prepared, the computer could arrive at an accurate decision on its course of action far more rapidly than can a human surgeon, and thus may be able to perform operations which are at present made impossible by the time factor. It is true that a very large store would be required to hold the coded equivalent of the surgeon's medical knowledge, and the computer would have to be provided with an input in the form of a television camera to scan the site of the operation, but neither of these requirements seems beyond the range of foreseeable developments. Preparing the programme would however be a colossal task.

It is worth noting that, whereas each human surgeon requires long years of training to prepare him for a professional career which lasts only for a limited number of years, a computer programme, once it is prepared, could be immediately duplicated and made available to any number of similar installations, and would be permanently useful.

TYPES OF LEARNING PROCESSES

Present efforts at assisting the programmer in arriving at a working programme are concentrated in the field of automatic coding, which enables programmes to be written in a language that is as close as possible to the language in which the problem originates. However, it does not relieve the programmer of the task of extracting from the mass of more or less relevant information associated with the problem a concise statement of the task to be performed by the machine.

The object of a learning process is to carry out the extraction of a comparatively few relevant facts from a mass of data. If the process is suitably designed, these facts might comprise a programme, or part of a programme, for a computer.

There are a number of ways in which various examples of learning processes may be classified, in particular the following:

1. The process may be limited in scope merely to determining the appropriate values of some parameters whose meanings have been laid down in advance, or it may be sufficiently sophisticated to be capable of building up logical structures of a very broad general type.
2. The process may be entirely automatic, or it may be effected by a combination of human and mechanical activity.
3. The results of the process may be put to use in the same computer as that which carries out the process itself, or they may be incorporated in a programme or in the logical design for another computer. (In the former case the process may be carried out continuously and may therefore be able to take into account long term variations in the prevailing conditions.)
4. The process may concern itself purely with an examination of the data which is presented to it, or it may be able to make a selection of the data which it is to receive. In the latter case it would obviously be able to save some of the time which would otherwise be wasted in scanning through irrelevant data.
5. The data which the process scans may be obtained from the outside world through a suitable input channel, or they may be generated within the computer by some computing process such as a simulation of a physical system. (In the latter case, the process would certainly be able to exert the power of selecting those items which it wished to examine).

Although it is not yet possible to see clearly the way in which learning processes are likely to be developed in the future, it seems probable that examples of many different types of process, according to the above classifications, will be used.

Applications in the Training Field

A learning process is already being used in the Solartron machine known as "SAKI" which is used in the training of card punch operators. This machine adjusts itself to the operator's rate of learning and presents her with tasks suited to her current ability. In fact it extracts, from the record of her reactions, the relevant facts to enable it to provide the appropriate exercises.

There seems no reason why similar principles should not be extended to the teaching of other skills, provided that the necessary data links can be constructed. Typing is an obvious subject; another possible application is flight training.

In these learning processes the learning is carried out concurrently with the teaching process which makes use of the experience gained by the machine; the machine is therefore able to adjust its behaviour as the pupil increases his skill.

Music

One of the first applications of a comparatively elaborate form of automatic learning is likely to be the composition of music. It is true that this will call for a rather sophisticated type of learning process, since the structural characteristics of the type of music to be composed may be somewhat elaborate. However, music lends itself conveniently to digital coding, and various attempts have been made in the past by experienced composers to lay down detailed rules of composition, so that a basis for automatic composition already exists. Some simple tunes have in fact been composed by computers (Brooks, Hopkins, Neumann and Wright, 1957, *ref. 1*). There is still however considerable scope for extending the rules so as to cover more ambitious melodies, and also to orchestrate them, and this could probably be done by applying a learning process to a large number of existing compositions.

By varying the list of compositions presented to the computer, the style of music which resulted could be altered. It is possible to envisage a two-stage process, in which the computer began by making a passive analysis of some given works, and then proceeded to compose a series of works itself which could be criticised by a human musician, thus adding to the computer's experience.

Information-Handling

There are many problems involving the recognition of various representations of information, e.g. printed, written, or spoken numbers or words. In all of these problems the chief difficulty lies in maintaining a useful degree of discrimination in spite of the presence of noise and distortion of the original information. A satisfactory recognition process must usually take into account the relative probabilities of various types of distortion, and it is quite possible that a learning process could be used in order to study the actual distortion experienced. In the case of written and spoken information, these distortions can be quite complicated, so that a powerful learning process might be required.

A similar problem is the interpretation of mis-spelled words; this has already been mentioned as a possible application of learning processes (Tizard, 1957, *ref. 3*). In all these recognition problems, the learning process would need to examine a large number of specimen items in order to build up its experience of the noise and distortion, and it would be necessary for a human to supply the correct interpretation of each specimen.

A process which has been suggested (Brown, 1958, *ref. 2*) as a means of arriving at a successful programme for language translation is in fact a type of learning process. Its originator suggests that some very elementary rules should be applied to a sample of text, and, whenever the rules break down, they should be amended in order to cover the cases so far encountered.

His suggestion is that a human should not only detect the cases where the rules break down, but should also then make the necessary amendments to the rules; the computer would be used merely in order to apply the rules to new specimen sentences. It will perhaps be possible one day to use a learning process which merely examines a lengthy sample of text and a translation of it, and devises its own rules for the translation process; this would clearly call for an extremely powerful learning process and is not likely to be practicable for several years. However, in the meantime Brownʃs suggestion has a great deal to commend it.

The Control of Physical Systems

There is a class of problems arising in a variety of different fields, involving a system of continuous variables which can perhaps be treated for a first approximation as a linear system, but in which the parameters of the system are not capable of direct observation and may perhaps change slowly with time. In such cases quite a simple learning process might be able to determine the parameters by investigating the behaviour of the system over an initial period, and thence to make predictions of the future behaviour of the system. Such problems might include the studies of economic systems and of the water flow in large systems of rivers. In so far as such systems could be considered as linear, the learning process would be comparatively straightforward; if however one attempts to take into account departures from linearity there would be considerable scope for ingenuity in the learning process.

It is perhaps possible that this approach might also produce useful results in the sphere of weather forecasting, in which the rigid application of the laws of physics requires a somewhat lengthy and cumbersome computation. Some work has indeed been done in the U.S.A. with a view to making forecasts by examining past records of similar circumstances, and this is a step towards the application of a learning process.

Many problems are arising nowadays in which the variables are not continuous and in which tasks such as optimisation are therefore difficult. There are several such problems in the field of production scheduling which require the optimisation of a function of a permutation (the permutation being the arrangement of processes to be performed by various tools in fulfilling given production requirements). In many such problems a complete optimisation is impossible, and a near-optimisation is sufficient. Various ad hoc procedures can often be devised which will usually produce a fairly close approximation to the optimum solution. However, the best procedure may not always be easy to find, and may depend on the incidental features of the problem. It is therefore possible that learning processes might be able to evolve successful optimisation methods adapted to various types of problem.

Such a learning process may turn out to be similar in structure to the one used by Newell and Simon in the United States to develop successful methods of manipulating logical formulae in order to prove mathematical theorems. Research in this line would seem to be promising. It is likely to find direct applications in the design of switching circuits, and also the learning principles involved may well be valuable in a number of problems in which classical mathematical logic is not directly applicable.

THE PROBLEM OF OBTAINING EXPERIENCE

In the field of traffic control it would not be appropriate to apply a learning programme directly to a real traffic system, because the whole system would probably be wrecked before the learning process had accumulated enough experience to be capable of handling the situation. However, in this case the physical part of the system could be simulated sufficiently well to enable trial runs to be made by a computer alone, so that the learning process could operate on a hypothetical series of situations in order to gain its experience.

In the case of surgery however the problem is much more difficult. One would not be able to afford to let the computer learn by a long series of mistakes, and it would therefore be necessary to let it draw on the experience of generations of human surgeons. In other words, the programme must be presented to the machine in a pre-digested form. However, this does not mean that it must be presented in any of the types of coding that are normally used for computer programmes. The ideal solution would be to prepare a "compiler programme" that could accept information in the languages in which existing text books are written. Once such a compiler has been produced, the whole of the existing store of medical knowledge could be incorporated into the computer programme. This would of course need to be supplemented by sections of the programme to deal with the interpretation of the input signals and the control of the surgical instruments.

Perhaps "compiler" is not an appropriate term to describe the programme that would be required for such a task, since it would need to be tremendously more powerful than any existing automatic coding system. It would need to be capable of applying the rules of grammer to the interpretation of the sentences presented to it, and of coding and tabulating in some systematic way the information which they contained. The production of such a programme would in itself be an enormous task, but it would be of value in many fields other than surgery. It would in fact give the computer access to the whole accumulated fund of human literature. As yet one cannot foresee the lines along which such a programme might be constructed, but again it seems likely that some kind of automatic learning procedure may assist in its development.

CONCLUSION

In comparison with the foregoing suggestions, the present stage of computer usage seems extremely rudimentary. The subject differs from most technologies in that its development will depend comparatively little on the discovery or exploitation of physical laws, but primarily on the sheer application of intelligence. This however will be needed in great abundance. Even the comparatively crude programmes that we are using today call for a great deal of brain power for their construction, and it is very difficult to say how much more effort will be required in order to realise the more exciting possibilities discussed above. The author feels however that this is purely a question of time, and that in due course we shall see computers performing tasks which are now regarded as essentially calling for intelligence.

ACKNOWLEDGMENT

The author wishes to thank Messrs. Ferranti Ltd., for permission to publish this paper.

REFERENCES

1. Brooks, F. P. Jr., Hopkins, A. L. Jr., Neumann, P.G., and Wright, W.V. An Experiment in Musical Composition. *I.R.E. Transactions on Electronic Computers*, 1957, EC-6, 175.
2. Brown, A. F. R. Language Translation. *J. Assoc. Computing Machinery*, 1958, 5, 1.
3. Tizard, R. H. Electronic Computers and National Insurance, *The New Scientist*, 9 May 1957, p13.

DISCUSSION ON THE PAPER BY DR. S. GILL

MR. J. A. GOSDEN: First, I would like to point out one of the problems which we have to face in giving effect to some of the suggestions made by Dr. Gill; that is the problem of producing good systems that "fail safe" and organising methods for trying them out. If we take the case of teaching a machine to compose music, any errors may make us uncomfortable, but errors in learning to wield a surgeons' knife are much more serious. The difficulty is to try and predict the internal and external conditions that have to be guarded against. Errors exist in all sorts of systems and we have learnt to live with them. Some of these systems are going to require quite a lot of re-orientation before we can live with them.

I would also like to ask a question that is not particularly directed to Dr. Gill, and it concerns the way administrators work. An administrator lays down a procedure for his staff and only situations outside the scope of the procedures are referred to him for decisions. After several similar situations have turned up, he saves himself further labour by generalising his decisions into a new procedure. Now, can we teach machines to generalise, and if so how can we discriminate between a good and a bad generalisation?

CHAIRMAN, THE EARL OF HALSBURY: It seems to me that we need more insight into the hierarchy of difficulties that are involved in some of the things we have been talking about. The type of difficulty Mr. Strachey referred to is one type. Another type of difficulty can be illustrated by the sort of experience I had once in attending a derating appeal where the business before us was to construe the words "scientific, charitable or social activities" for the purpose of the Act. What was a "social" activity? Learned Counsel on each side proceeded to argue this case for two hours. They took instances of the use of the word social, as in the context of social worker, social life, social science and so on. In each case the word "social" has a different connotation. These differences represent one kind of difficulty. The judge was asked on this particular occasion to give a considered judgment and state his reasons, which he did. What he did not state were the reasons that lay behind the reasons. Those of course were embodied in a degree course in law, 20 years experience at the bar, another 15 years on the Bench and so on. Computerising these would involve difficulties of a much higher order.

Speaking as a computer with some 10^{10} multiple threshold relays in it, a computer which spent three years taking a degree course, I ask: could one

design an appreciably simpler organism to do the same sort of job? I doubt it. We must be clear as to the hierarchy of difficulties involved in thinking and learning processes, and which of them could be solved by the type of computer we could hope to build. I agree with Dr. Gill that the old fashioned computer of 10 years ago has still a very long and useful life to run.

MR. R. H. TIZARD: When one considers the application of a learning type of machine to practical problems, I think there is one factor that is overlooked. There is an obsession on the part of people concerned with developing learning machines to start off with a machine which is entirely empty of previous experience and learning. Not only is the sort of machine we can make now extremely small in its capacity compared with the capacity of the human brain, but one must also bear in mind the length of time spent in putting experience into the human brain before it can be considered to be of any use whatever. For instance in political affairs this time is considered to be 21 years. Therefore an important requirement for any application of this nature is to build in in the first place as much experience as one can. This can be done in two ways: first of all in the design of the equipment itself; secondly in a method of supplying experience to an empty machine in a very rapid way before it is put on to carrying out its work. After this is done there is a third way in which one can improve its rate of experience, that is by teaching it - allowing humans to teach it - as it does its work. Here I think there is one very important practical problem which is overlooked, and that is that the methods of communication between the human and the machine at the moment are so elementary that they will form one very serious obstacle to the practical use of learning machines.

MR. P. REDFERN: It seems to me that many of the problems which we shall have to tackle in the next decade or so will be of such complex mathematics that we will have to rely to a great extent on automatic programming procedures to help us if we are to reach our goal. That leads me to suggest that in designing machines we must pay more attention to the order codes that are used. This is because in preparing an automatic programming procedure for the machine we must in fact tell the computer what its own rules of procedure are.

If these rules of procedure embodied in the order code are powerful but nonetheless essentially simple, then the task of preparation of an automatic translation programme will be eased. But if the rules of procedure are complicated by many restrictions and exceptions (e.g. if there are complications of optimum timing), the task is made more difficult and the possibilities of producing a powerful and efficient automatic programming procedure is reduced. If the order code of machine A is C times as complex

as that of machine B but of equal power, then I would venture to suggest that an automatic translation programme for machine A might be of the order of C^2 times as complex as that of machine B.

MR. C. STRACHEY: I have three rather disconnected comments. The first is about the problem of a machine making mistakes. I should not mind being anaesthetised by an anaesthetising machine provided it made fewer mistakes than a human anaesthetist; after all, human beings occasionally make mistakes. In general when replacing human beings by machines the important thing is not that the machines should make no mistakes, but that they should make fewer than the human being they replace.

Secondly, I should support what Tizard said about teaching learning machines. I was a schoolmaster for some period and spent a long time teaching some rather elementary learning machines, and I can assure you that the problem of communication both ways is exceedingly important. I think that if a machine which is trying to learn, cannot actually ask questions it must do something which is more or less equivalent. It could, for example, make experiments or suggest tentative solutions. It is only by something like this that one might hope to teach a machine to do something without making a great many disastrous mistakes.

The third point is about the desirability of designing machines to make automatic programming easy. I think it is important that the design of computers should be such that it is possible to write an automatic programming system, but I do not think that it is particularly important to make it easy to do so; after all a system like a FORTRAN has only to be written once. It is, however, of no use trying to write an automatic programming system for a machine which is intrinsically too complicated to make the resulting programme economically worth while. I think that machines having multiple-level stores probably come into this category.

MR. G. M. E. WILLIAMS: Pursuing Mr. Tizard's point of the problem of human beings communicating with machines, I am going to assume that Dr. Gill's paper and all the discussion so far is in terms of a purely digital machine. It always strikes me as odd in the case of punched card equipment applications for example, that so frequently lengthy tabulations are produced because the meaning of these long columns of figures is so difficult to comprehend by anybody looking at them. A decision on general policy or major stragety can surely be made more easily on a mass of data presented in graphical form. A little work has been and is being done in this country on output of digital machines in graphical form, for instance in examining the results of observations obtained by telemetry from guided weapons and also in elementary attempts to control processes by computing machines.

The input equipment problems of digital machines have advanced less than those of output in alpha numeric form partly because of the technical difficulty of eliminating the human element and partly because of the analogue to digital conversion usually required causing the speed of input equipment to fall below that which a digital machine could accept.

I think that a larger proportion of effort and a wider outlook than high speed printing is needed for output devices. The proper amalgam between analogue and digital devices should be the aim and not just analogue results alone expressed in two-dimensional form. A digital surface generator might provide another application for some of the work which has been expended on computer control of machine tools and such a generator would be a powerful means of communicating computed results to human beings.

MR. D. W. WILLIS: I would like to ask Dr. Gill whether he knows of a learning programme in which the performance of the machine has given us some information. I myself have seen learning programmes demonstrated when the programmer appeared surprised at the result the machine gave. It seems to me that, having worked out a learning programme, there is no longer any point in making the machine do it - in fact there is no point in programming it as long as you know it can be programmed. I would ask Dr. Gill why in fact he wrote the learning programme he mentioned; was it to measure the speed of his own reactions?

CAPT. S. JEFFERY: During this meeting we have discussed either automatic programs or mechanized thought processes with the aim of satisfying in some manner the problems which we cannot satisfactorily program today. Dr. Gill has briefly noted in his paper one extremely significant area. "The combination of human and mechanical activity". This man machine relationship is becoming increasingly important particularly in light of the present studies in computer technology. The next five years will make available machines operating 10, 20 or even 50 times faster than the present 1 microsecond devices. In view of this tremendous data handling capability one must become more concerned with the problem of how the human can effectively apply decision elements throughout a data processing event. The complex programming problems we are faced with today might conceivably be reduced to practice as a result of an integrated man-machine, where the man can apply tests, human decisions and effectively control all facets of the data manipulation or process.

DR. S. GILL (in reply): I agree that the problem of coping with the great mass of mistakes that are sure to be made as we learn how to make learning processes is one of the greatest difficulties which we have to face. It is one of the reasons why I feel that this second revolution in computers is likely to be at its peak in 1980 rather than in 1960.

I also agree that if we are going to make much practical use of a learning process it will probably not be very economical to start with practically no information in the machine and expect the process to accumulate all the information it requires. We should try to put in as much as we can at the beginning in order to give it a good start. On the other hand while we are in the experimental stage it is perhaps worth while experimenting with very elementary processes in a machine with practically no preconceptions, simply in order to get at the fundamental principles of learning.

I agree with Mr. Strachey that the most efficient use of a multi-level store is often something which cannot easily be achieved by an automatic process. On the other hand, the order code itself seems to me something that does not fundamentally effect the suitability of a machine for automatic coding. One of the aims of automatic coding is to insulate the ordinary user from the details of the order code, which have to be considered only by the man who writes the translating programme. To some extent therefore complications in the order code can be more readily tolerated when using automatic coding. It is only when these complications reduce the effective speed of the machine that they become serious.

With reference to Mr. Strachey's comment about the desirability of a machine being able to ask questions, I think that I do make the distinction in the paper between processes which do, and those which do not conduct experiments. Obviously the former will be more powerful, but there might well be situations in which experimenting is impossible.

In reply to Mr. Willis' question, the actual reason why I myself constructed a simple learning programme in 1951 was partly because it was fun, and partly because it added some impressive material to my Ph.D. thesis. The only sense in which I was "surprised" at the machine's behaviour was that the machine succeeded in beating me in a guessing game. Apart from this I learned nothing from the programme itself.

The question of graphical and other analogue forms of input and output is part of the whole question of the relationship between the man and the machine, in which there is obviously great scope for developments. If satisfactory techniques can be developed, it does seem quite likely that we shall establish systems in the future which operate by an intimate combination of human and machine activity. This problem even arises to some extent in running a conventional computing installation if the speed of the computer is extremely high, because satisfactory procedures must be established for scheduling and recording the problems being solved whilst keeping idle time to a minimum.

SESSION 4B

PAPER 5

AUTOMATIC CONTROL BY VISUAL SIGNALS

by

DR. W. K. TAYLOR

BIOGRAPHICAL NOTE

Dr. W. K. Taylor graduated from the University of Manchester with first class honours in electrical engineering and then spent a further two years on control system research before proceeding to the Massachusetts Institute of Technology. At M.I.T. he combined research on control systems and computers with post-graduate study in the field of communication and information theory. On returning to England in 1953 he spent a year on communication research at Imperial College London. Since 1954 he has been in charge of the Electrical Engineering Section of the Nuffield Foundation Cerebral Mechanisms Research Unit at University College London.

AUTOMATIC CONTROL BY VISUAL SIGNALS

by

DR. W. K. TAYLOR

SUMMARY

VISUAL signals, derived from spatial patterns or light intensity distributions by means of a transducer system, are in general analogue signals. A method of combining them to form error signals for controlling visually operated tracking servomechanisms is described. The more complex operations of sorting and reading can also be controlled by the signals if they are subjected to appropriate transformations but because the outputs of a sorting or reading machine are effectively digital signals there must be an analogue-digital conversion stage before the final output. The new methods of machine synthesis described in the paper preserve the analogue character of signals up to the input terminals of a final conversion stage before which only the simple operations of adding and subtracting appropriate proportions of selected signals are required.

Electrical networks for performing the signal transformations and analogue-digital conversion are described and analysed. The characteristics of sorting or reading machines synthesized by the new method are compared with those exhibited by the human operator.

It is concluded that the synthesis procedures described in the paper lead to the construction of sorting and reading machines that have many characteristics of the human operator, of which the ability to guess the operation to be performed on the evidence of visual signals that are distorted in intensity and/or spatial position is probably the most important.

INTRODUCTION

A MACHINE that is capable of learning to recognize simple visual patterns is described in earlier papers *(refs. 1 & 2)*. In this machine an input classification signal is supplied at the same time as visual signals representing the distribution of light intensity over a sample pattern.

The coincidence of the signals is recorded in a memory unit that increases the transmission of paths from the visual signal input terminals to an output terminal that corresponds to the classification, until the signal appearing at the latter terminal is sufficient to operate an indicating relay. This setting-up procedure is repeated for all patterns and classifications that are of interest in any particular application.

Information stored in the memory units of this machine after a period of learning can be extracted in the form of path transmissions and incorporated permanently in a second machine that is incapable of learning new discriminations but which otherwise has all the characteristics possessed by the first machine at the time of the information transfer. Machines of this second type can be considerably simpler than pattern learning machines containing the same information since they have static instead of dynamic memory units. They can be equally efficient, however, in controlling operations in response to visual signals if the response characteristics do not have to be changed by the signals. In addition to the simplification of a static as compared with a dynamic memory a fixed purpose machine can have a much smaller storage capacity than a learning machine that has to be potentially capable of storing information about the wide range of possible patterns that may occur during the training process. The learning machine must have sufficient memory capacity to learn any discrimination within its range of accuracy and the non-learning machine can be thought of as a frozen version of the former in which the memory is made static and the unused storage space discarded. The two types are thus functionally equivalent over a short interval of time.

A static type of recognition machine may of course be constructed without reference to the memory of a learning type if some other means of obtaining the necessary information about the visual patterns of interest is available. If we only wish to construct a machine that will recognize a small number of sample patterns together with variations of these within certain limits an approximate analysis can be carried out manually and an attempt made to synthesize the simplest machine that will give reliable results. This analysis and synthesis procedure forms the principal subject of the present paper.

INPUT AND OUTPUT SIGNALS

In the machine referred to above *(refs.1,2)* electrical signals corresponding to the light intensity at different parts of the visual field are obtained from a matrix of photomultipliers which can be said to simulate receptors in the retina. The disadvantages of this transducer system, namely the high cost, weight and volume of a large matrix, have led to its replacement by a television transducer system in which the

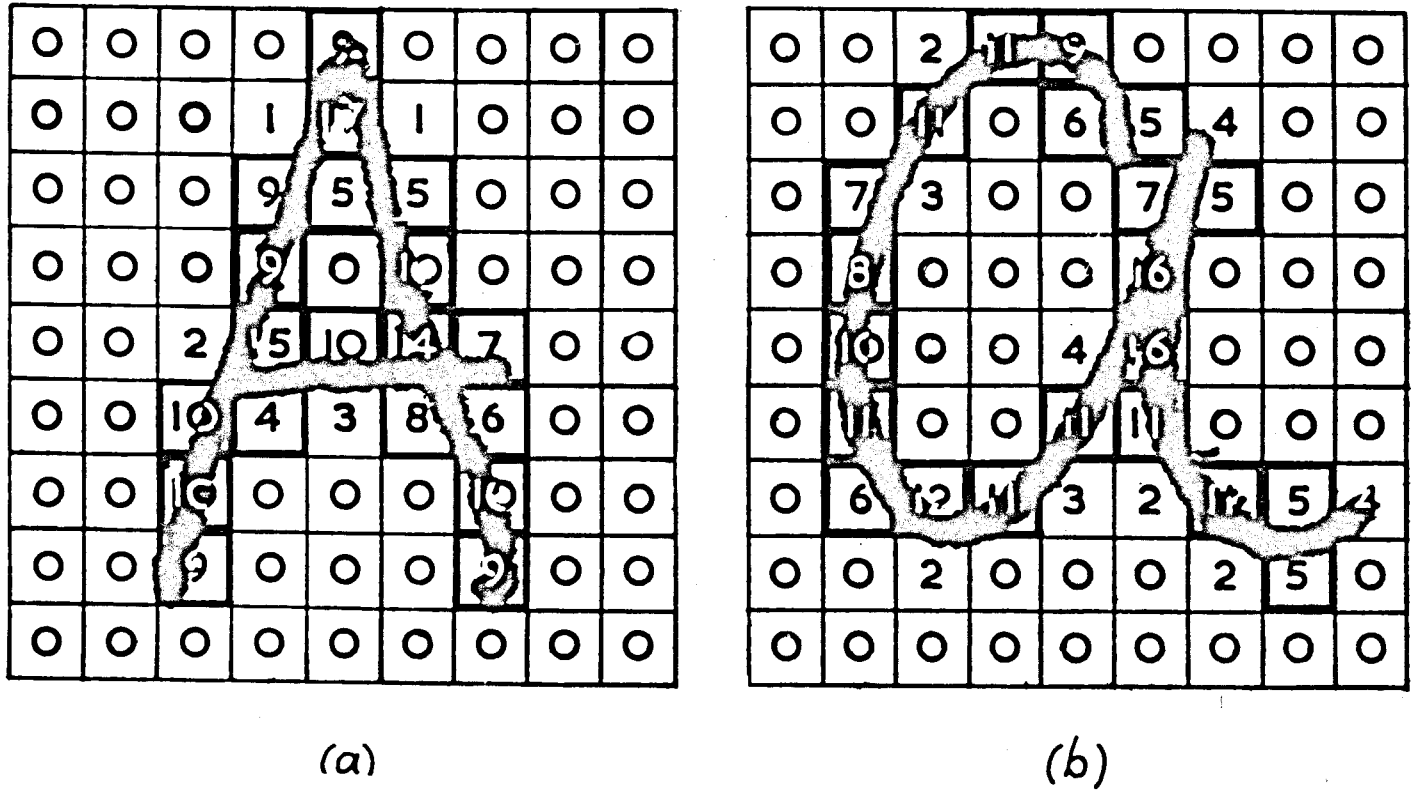


Fig.1. Visual Patterns and Input Signals.

video output waveform of a television camera is synchronously gated through suitable smoothing circuits to a matrix of output terminals. The smoothed output voltage at any terminal in the matrix is approximately proportional to the integrated light intensity over a corresponding square element of the camera tube surface and hence to the light received through the camera lens from a corresponding element of the visual field.

It has been shown *(refs.1,2)* that the matrix of signals produced by a solid object or shape can be transformed by a device that has been called a "detail filter" into a simpler matrix of signals in which only the outlines of the object or shape are preserved. It will be assumed that this detail filtering has taken place or that the original patterns of interest are in the line form, as handwriting is, so that a detail filter is not required. The enlargements of handprinted and handwritten letters shown in *figs.1a and b* serve as examples of line patterns but the following discussion applies to any conceivable distribution of intensity.

Analogue input signals $x_1 \ldots x_s \ldots x_S$, proportional to the integrated vidoe signal over corresponding squares of the visual field matrix, are formed by networks that switch the video signal to a matrix of S integrating capacitors. In the examples of *fig.1*, the resulting input signal amplitudes have the values shown if a constant black level over the letters generates positive video signals and if the background gives zero signals.

The object of the synthesis procedures to be described is the design of a machine that will generate binary output variables $z_1 \ldots z_c \ldots z_C$ in response to the S analogue input signals $x_1 \ldots x_s \ldots x_S$ according to rules specified by the designer. The machine can thus be represented by a box between the inputs and outputs as shown in *fig.2*.

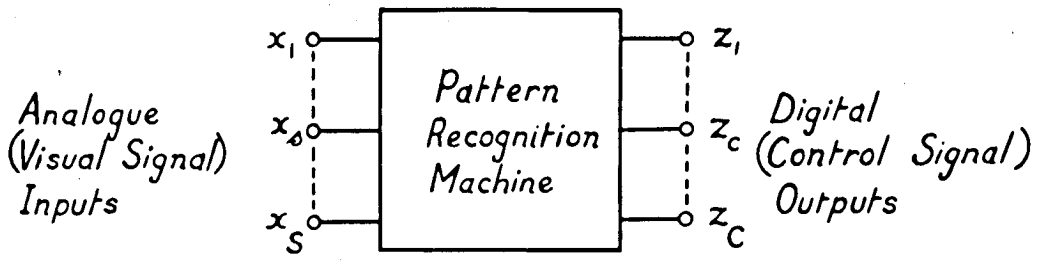


Fig.2

The binary outputs may be used to control a wide variety of operations. Letters, cheques, etc., may be sorted automatically by arranging for the outputs to open mechanical gates. In the case of letters a large number of variations in the writing and printing of an address is required to produce the same sorting operation but with cheques the acceptable range of a signature is limited to variations that fall within much closer tolerances. Information derived from written or printed instructions may also be fed through the recognition machine into an electronic digital computer thereby eliminating the intermediate stage of punched card or tape preparation.

Tedious statistical work such as the estimation of conditional probabilities of letters and words in a written language could be mechanized by supplying the binary output signals to the input terminals of a conditional probability computer *(ref.3)* as the pattern recognition machine scans the pages of books.

One final example of the many applications that could be mentioned involves the selection of recordings of spoken letters or words by the control signals as a page is scanned by the input lens, thus providing blind persons with a possible means of "reading" newspapers and printed books.

ANALYSIS OF THE INPUT SIGNALS

To minimize the cost of the machine it is necessary to employ the smallest number of input signals that is consistent with reliable operation. In *fig.1* there are 81 signals and it is clear that the minimum number will depend on the complexity of the patterns that have to be discriminated. Each member of the infinite set of different configurations that a fixed amount of light flux can take up within a small square produces the same output signal and the members of the set cannot be distinguished. In general it can be said that patterns must cross a different set of squares if they are to be distinguished but the sets need only differ by one or more squares.

The analysis of the input signals follows the procedure adopted in earlier work and involves the formation of the sums of sets, each set containing any number of signals from one to the maximum of S. In practice the average number of signals in the sets is usually far less than S. The following semi-manual method of selecting the sets is based on the automatic process that takes place during the training of the learning type of machine.

A pattern that is required to produce a binary output at terminal z_c is placed before the television camera and a lens that causes the pattern to occupy approximately 3/4 of a monitor screen is chosen. A meter is then

connected in turn to the x terminals and the M non-zero signals, denoted by the non-zero numbers in the examples of *fig.1*, are recorded.

There are $2^M - 1$ possible ways of forming sets containing from 1 to M and in general m non-zero signals. The sum of the M non-zero signals divided by $2m$ is subtracted from the average value of the x signals in the sets of m to form y signals given by

$$y = \frac{1}{m}\sum_{m} x_s - \frac{1}{2m}\sum_{M} x_s \qquad (1)$$

Thus for any given sample pattern that produces M non-zero input signals there are in general $2^M - 1$ values of y of which only the largest is significant for the synthesis procedure. In the example of *fig.1a* the maximum y signal is 4.47 units and occurs when m is 19, the particular set of 19 signals being those surrounded by thicker lines. For the *fig.1b* example the maximum signal is 4.00 units and the corresponding value of m is 21.

The actual synthesis of particular y signals is achieved by means of the network shown in *fig.3*. It will be noted that the signal $\frac{1}{S}\sum_{S} x_s$ is independent of m and can be used for supplying any number of networks of the form shown on the right of *fig.3*, providing the loading effect is negligible.

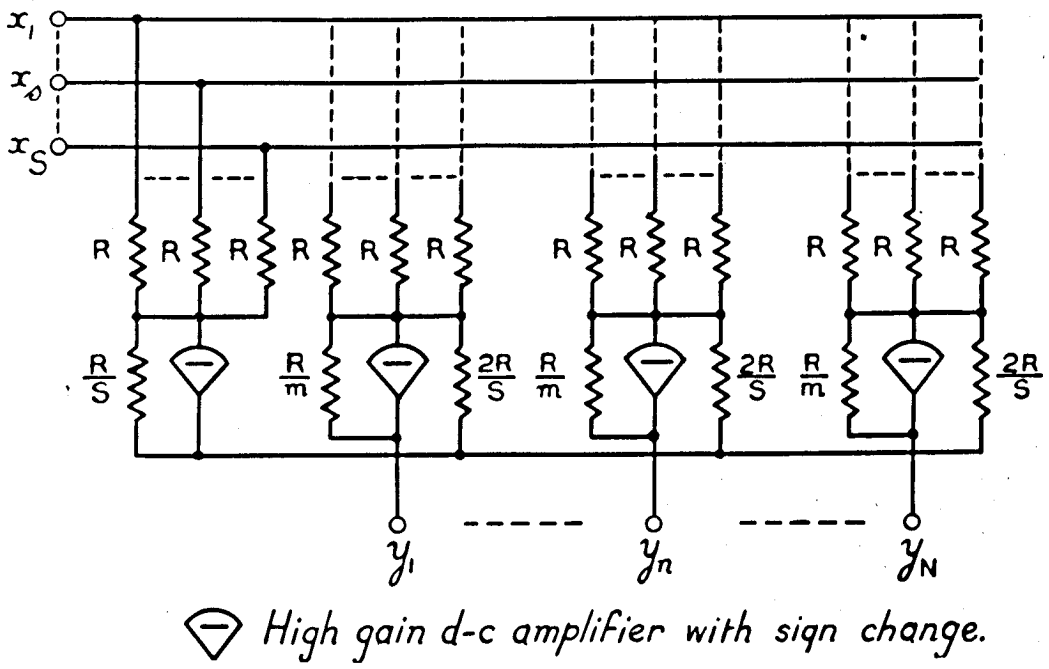


Fig.3. y-Signal Networks.

CHARACTERISTICS OF THE y SIGNALS

The relationship between the x and y signals given by equation (1) satisfies two conditions that are found to give desirable control characteristics in the case of visual input signals; firstly that all classified input patterns of equal intensity should give maximum y signals of equal amplitude and secondly that for any set of m equal non-zero input signals (the remainder $S-m$ being zero) there should be a y signal that is greater than any other y signal when the set occurs.

One consequence of the first condition is that a machine with an input matrix sufficiently large to accommodate words of any length would give approximately equal amplitude maximum y signals for all words, irrespective of length. This approximate equality can be illustrated by assuming that the m input signals in the set corresponding to the maximum value of y are equal to the average value x_{av} and that all other x signals can be neglected. The maximum value of y is then $y_{max} \approx \frac{1}{2} x_{av}$ which is independent of m and therefore of the word length. If the $M - m$ non-zero input signals not included in the m are taken into account by assuming them to be some fraction $c\ x_{av}$ of x_{av} instead of zero the maximum value of y becomes

$$y_{max.} = \frac{1}{2} x_{av} \left[1 - c \frac{(M-m)}{m}\right]$$

The value of c at which the outputs of the two networks designed to give a maximum output when c is zero and one become equal can be found from the equation

$$\frac{x_{av}}{2}\left[1 - c\frac{(M-m)}{m}\right] = \frac{x_{av}}{2}\frac{m}{M}\left[1 + c\frac{(M-m)}{m}\right]$$

If $m = M/2$ then $c = 1/3$ which implies roughly that a long word made up of two short words of equal length, one of which is only c times the intensity of the other, will be recognized as the long word when c is greater than 1/3 but as the more intense short word when c is less than 1/3.

According to the second condition the maximum y signal produced by a set of m non-zero and equal signals x_a should be greater than any other y signal, irrespective of m, when the set occurs. In terms of eqn. (1) this inequality may be written

$$Max.\left[\frac{\sum_{m+k} x_s - \frac{1}{2}\sum_{S} x_s}{m+k}\right] < Max.\left[\frac{\sum_{m} x_s - \frac{1}{2}\sum_{S} x_s}{m}\right] > Max.\left[\frac{\sum_{m-j} x_s - \frac{1}{2}\sum_{S} x_s}{m-j}\right]$$

$$1 \leq k \leq S-m,\quad 2 \leq m \leq S,\quad 1 \leq j \leq m-1 \qquad (2)$$

where it is understood that the set of m signals that the network with m

inputs was synthesized for is present. The integers $m+k$ and $m-j$ are the number of inputs of y-signal networks with more or less than m inputs respectively. When the values of the x_s are inserted the inequality simplifies to

$$\frac{m}{m+k} < 1 > 2 - \frac{m}{m-j}$$

which is independent of the actual value of x_a. The inequality will be seen to hold for any possible value of m and the corresponding ranges of k and j.

REJECTION OF INSIGNIFICANT PATTERNS

It has been shown that networks can be synthesized that will give a maximum output signal in response to any given sample pattern. The variation of this signal as the pattern is changed will now be considered and it will be seen that the network is insensitive to small changes in a complex pattern. If in particular a pattern that produces m equal x signals x_a is changed by having any k signals removed or by having k signals added the output of the network that was giving the maximum value of y decreases by an amount $\frac{k}{m}\frac{x_a}{2}$. The rate of decrease of output with k from the maximum value of $x_a/2$ is thus $x_a/2m$ on both sides and y is zero when $k = m$. It will be noted that complex patterns giving rise to a large m value can suffer more distortions than simple patterns before the output y is attenuated to zero but that the percentage distortion required to reduce the output to zero is independent of complexity, being 100% in all cases for this form of distortion.

If the reduced y signals produced by distortions are above a threshold level they can be made to select the same output as the maximum signal and the machine can then be said to generalize from a particular sample pattern to distorted versions of the sample. One form of distortion that is of particular interest is produced by moving the pattern relative to the input matrix without changing its shape. When the movement is only a fraction of a matrix square the value of y will in general only decrease slightly but a movement of one or more squares may reduce it to zero. Thus if a given shape is to be recognized at all positions in the field a y network must be synthesized for a sufficient number of sample positions throughout the field. This also applies to tilted or upsidedown shapes. Whether or not all the possible variations in the position of a shape are required to produce the same z output will depend on the application. The essential point to note is that the machine, in common with the human observer, can control different operations according to differences in the position or orientation of a given shape providing the differences are sufficiently great but that it is also a simple matter to arrange for the machine to give the same output for all positions and sizes of the same shape. These observations also

apply to projected images of three dimensional objects, so that the image of a cube when projected from various directions may give the same z output or different outputs as decided by the designer.

A second form of distortion that occurs in practice results from poor printing or writing in which parts of a character may be missing. The remaining parts make a contribution to the y signal that is approximately proportional to the ratio of the part to the whole. In other cases there may be contamination of the character by additional signals and again the y signal will be reduced, approaching zero as the complexity of the additional parts tends towards that of the character. If the diminished y signal will still select the appropriate output the machine can be said to ignore the distortions but a small difference can if necessary be detected since it is always possible to synthesize a separate network for the new pattern if the difference is such that a different maximum y signal can exist.

In summarizing these properties it can be said that a machine of this type tends to generalize from a sample pattern to patterns that resemble it unless the latter are analysed in detail and provided with separate y networks where the analysis indicates that this is possible.

ANALOGUE CONTROL OF THE BINARY OUTPUTS

If y signals can be derived from a set of differing sample patterns so that each signal is a maximum when its associated pattern occurs and if the maximum y signal causes a z signal to change from a resting level to an active level the change serves thereafter as an indication that a pattern identical with or resembling the sample pattern for which the y signal network was synthesized has been presented. Any new pattern that gives the same maximum y signal as a member of the set will not be distinguishable from that member but if the new pattern is not identical with the member it can always be resolved by increasing S until a different maximum y signal can be found.

A simple form of binary output that is often convenient consists of a set of relays, one for each z signal. The active level of a z signal can then be represented by one state of the relay and the resting level by the alternative state. To be specific the energized state, representing the active level, will be symbolized by the binary digit 1. Since relays operate at a threshold level of input it is important to have a signal that is nearly always well above or below a threshold if indeterminate operation is to be avoided. Furthermore it is essential that only one or no relay should be energized at any instant if each type of pattern is only to change one or no z signal from the resting level 0 to the active level 1. The possibility of two or more nearly equal y signals producing more than one indication can be reduced by means of a "maximum amplitude filter" *(ref.2)* that amplifies the

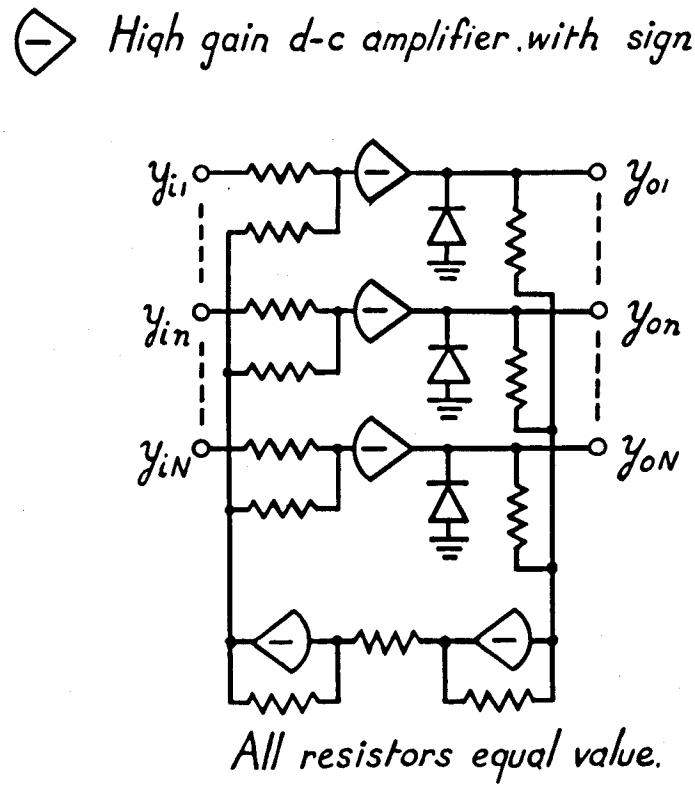


Fig.4. Maximum-Amplitude Filter.

percentage difference between nearly equal signals, attenuates equal signals and leaves distinct maxima unchanged except for a constant multiplier which can be made ± 1. One form of the filter is shown in *fig.4*. The inputs y_i are the y signals defined in previous sections which are assumed to be negative going or zero and the outputs y_o are the filtered inputs with a change of sign. The operation can be illustrated by assuming that y_{in} is the maximum negative input signal and that all the outputs except y_{on} are zero. Under these conditions y_{on} is approximately equal to $-y_{in}$ and since the voltage fed back to the units is greater than the remaining inputs the rest of the outputs must be held at zero by the diodes as assumed initially. In the unlikely event of r identical inputs of amplitude y_{ir} being larger than the remaining $(N-r)$ inputs the corresponding r outputs y_{or} are also equal and are given by $y_{or} \approx -y_{ir}/r$. The y_o signals can thus be used to operate relays and there is a low probability that more than one relay will operate at any instant.

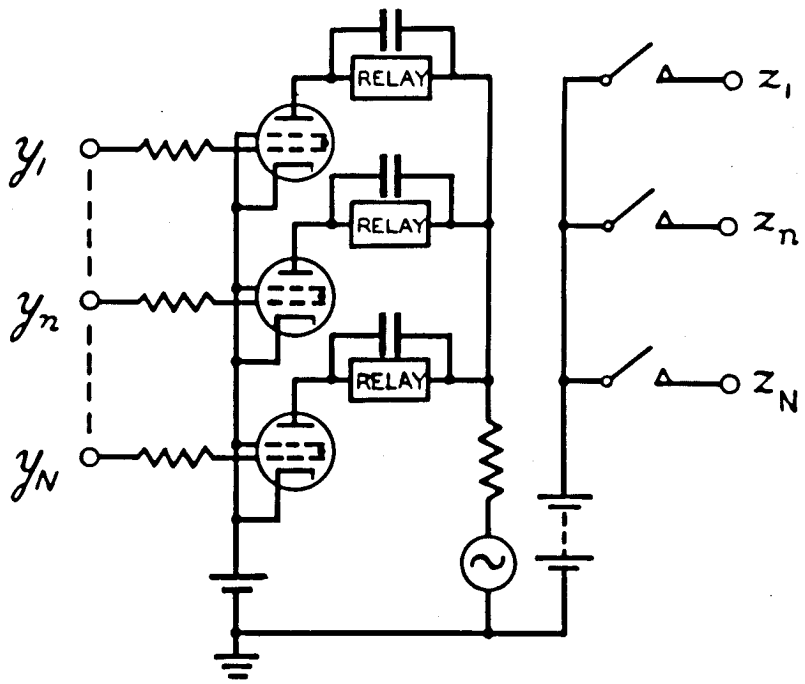


Fig.5. Output Stage.

An alternative form of output stage that may be driven by the original y signals after a change of sign or by the outputs y_0 of the maximum amplitude filter consists of a set of thyratrons supplied by an alternating voltage through a common resistor and the output relays as shown in *fig.5*. During the negative half cycles all the thyratrons are non-conducting and the first tube to conduct as the voltage rises for the positive half cycle is the one with the highest grid supply voltage. When this tube conducts the anode voltage of all the tubes falls to a level at which ionization cannot be initiated and the only relay to operate corresponds to the maximum y signal. In practice it is often necessary to arrange for the same z output to be operated by more than one y signal. Output z_1 for example may be required to indicate the occurrence of both styles of letter A in *fig.1*, as well as printed letter A's of various shapes and sizes. This result is achieved by connecting the z_1 relay to the anodes of all the thyratrons supplied by signals from y networks that correspond to the set of A-type patterns and in general this will reduce the number of final outputs to C, which may be smaller than N.

AUTOMATIC TRACKING OF VISUAL PATTERNS

When a pattern is magnified until it occupies a large proportion of the visual field the number of y networks corresponding to different positions of the pattern is reasonably small if the matrix only contains a small number of squares. For the A in *fig.1a* for example, the number is approximately 10 if only horizontal and vertical movements are considered. A small pattern such as a single dot may be of interest however, in applications that require fixed magnification and it is clear that the dot can in general occupy any of the S matrix squares. A considerable saving in the number of y networks can then be achieved by introducing an automatic control system that stabilizes the positions of patterns relative to the matrix and also serves as an automatic tracking device for moving patterns. A pattern can be kept at the centre of the matrix by two servomechanisms which shift the image of the pattern sideways and vertically until appropriate sums of x signals are balanced. The central position can be made stable for continuous patterns by introducing appropriate correction networks and a particular pattern always moves to the same stable position on the matrix. In *fig.1* the letters are shown in the central stable positions.

Three methods of controlling the relative motion of pattern and matrix have been investigated. In the first method the camera tube or matrix of light sensitive elements is mounted on a platform that can be tilted or rotated by two motors controlled through suitable amplifiers by the error signals. The main disadvantage of this system, namely its large size and slow response, is overcome in the second method by using instrument servomotors to drive a mirror in the optical system.

In the third method of position control the position of the camera tube scanning raster relative to the light sensitive surface is controlled by currents supplied to horizontal and vertical deflection coils. The choice between these methods depends on the application and in particular on the speed of response and tracking angle required.

CONCLUSIONS

The synthesis procedures described in this paper lead to analogue operated pattern recognition machines that are considered to be particularly well adapted to visual signal inputs. The most important characteristic of the machines is that the y signal produced by distorted versions of a sample pattern is only reduced in approximate proportion to the amount of distortion and still holds the same output active unless the distorted patterns have new significance or the distortion reaches the 100% level. A special condition arises, however, when automatic tracking is in operation. The distortions must then be made in such a way that the error

signals of the horizontal and vertical servomechanisms remain approximately zero if the decrease in the y signal is to be proportional to the amount of distortion. A circular letter O for example with a circumference of sixteen units may be distorted, without changing the stable position appreciably, by removing eight equally spaced units of circumference to leave a dash line circle. The y signal is then reduced to approximately one half the original letter O value and produces the same output unless a separate y-network exists for the dash line circle. If, however, the eight units of circumference are all removed from one side of the letter, reducing it to a semi circle the error signals become non-zero and in the new stable position the y signal is reduced to a much lower proportion than one half of the complete sample value. It is also characteristic of human recognition that evenly distributed distortions, such as might be produced by printing or writing on rough paper, lead to less errors than concentrated distortions in which an equal but continuous area of the pattern is missing or obscured.

The important characteristic referred to above emerges quite simply as a consequence of the analogue signal addition methods described in this paper and is not easily incorporated into machines *(refs.4 and 5)* in which binary signals, derived from analogue visual signal inputs at the outset, control the states of logical gating circuits that supply binary output signals.

REFERENCES

(1) TAYLOR, W. K. Electrical Simulation of Nervous System Functional Activities. Information Theory, Edited by Colin Cherry. *Butterworths Scientific Publications, London,* (1956), p.314.

(2) TAYLOR, W. K. Pattern Recognition by Means of Automatic Analogue Apparatus, *I.E.E. Monograph, in preparation.*

(3) 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).

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

(5) Electronic Reading Automation. *The Engineer,* 1957, **203**, p.414.

DISCUSSION ON THE PAPER BY DR. W. K. TAYLOR

MR. R. H. TIZARD: There are in my opinion two purposes, and only two purposes, in making the sorts of machine which we have been hearing about in the course of this symposium. One reason is because one expects to learn something as a result of making the machine. The second purpose is that it has some practical use. Dr. Taylor's extremely interesting work in this field probably satisfies both of these objectives. But I wish to talk only about the second one; and although he has not written very much about this in his paper, from what he has said I deduce that he is not ashamed of this practical use.

I think Dr. Taylor's contribution here is a very considerable one and does lead to possibilities of character recognition by very simple means which have not hitherto been discovered. I would suggest that the question of analogue versus digital means is not fundamental, and that the basic and really important point is the method of determining what I might term a probablistic coincidence between the observed word and the pattern word. This is done in fact by summing the coindences of digits between the word as measured from the character and the pattern word, and in this way it is possible, without very complex logical networks, to show coincidence which is not exact but is very near. In order to produce a simple machine to do this, there is a fair amount of design work required to obtain the parameters for the specified type of characters. After talking to Dr. Taylor about this, I was interested in the possibility of doing it with a computer programme, and I have now made such a programme. It is a "learning" type of programme, in which the machine is fed with characters of the type the finally-designed machine will be expected to read. In this way it can be used as a design tool, but it also has the advantage of being able to prove to what extent one can cut down on the amount of information needed, for instance by using only the quantised levels 0 and 1 for each of the black and white squares. I only completed this programme the night before last, and it has not had much time to be tested, but I think the results shown so far are encouraging.

The first test was intended to show whether this sort of method could be used for handwritten characters which vary very considerably, as compared with typewritten or printed characters. The characters were drawn

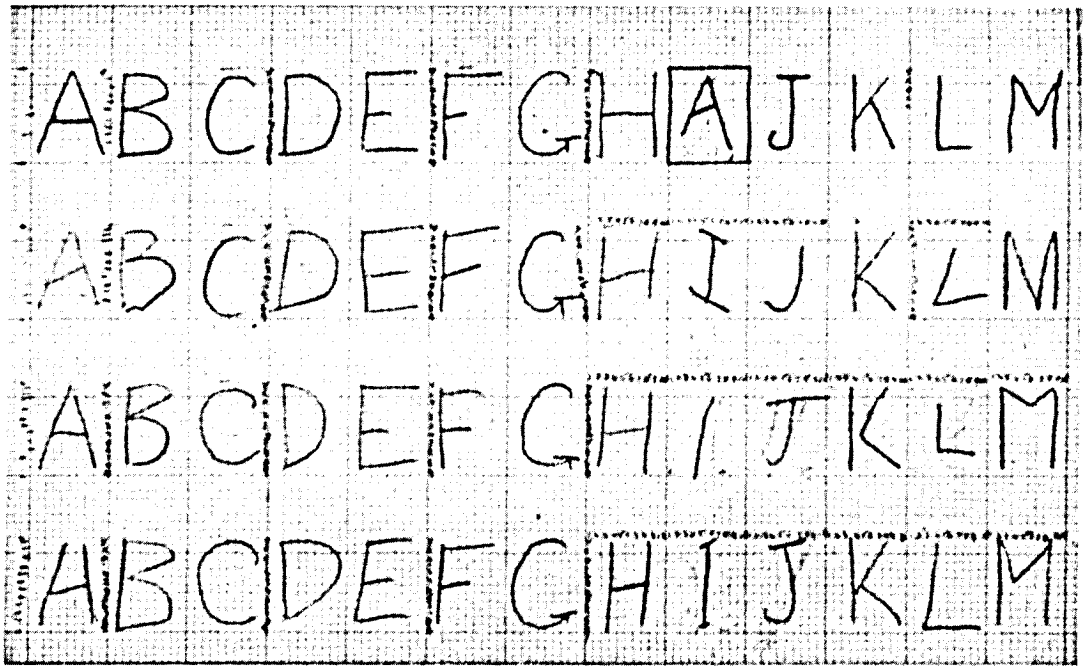


Fig. 1.

out on squared paper (*fig. 1*), and since I did not have a photocell reading device of any sort there was a manual translation from this form to a punched card form, in which, simply, a hole in the card corresponded to any square through which a line passed, and no hole in the card corresponded to a square through which a line did not pass. The letters A to H were fed into the machine, four samples of each, and these were labelled with the normal code, and the machine learned them. Then the same characters were put in again without labels. The programme is designed so that the machine produces one answer, and then if it is told that is wrong it produces a second choice. It can also, by a weighting technique, decide automatically the degree of certainty of the character chosen.

When these characters, A to H, in their four different forms were fed through the second time without labels, it made only one mistake, and that was on one of the G's. As you can see the G's are very similar to the C's, particularly when you consider the coarseness of taking only 0 or 1 for each of these squares.

I was then kindly presented with some assistance by the staff of the Laboratory to draw out some more letters and punch up the cards, which is

rather tedious work. I gave them the first sample set and asked them to continue the rest of the alphabets. Unfortunately I have not had time to test them, but I have made the machine learn all four alphabets and then go as far as G again, and the fact that it had a lot more characters to distinguish from did not make it very much worse - in fact it made a second mistake. There has been only one case so far in which the machine has been required to recognise a letter which it had not seen before, and that is the A shown in the square. It is quite a lot different from the other A's but it was recognised without difficulty. To indicate the loss of information content *fig. 2* shows what the A in the second row looked like when punched on the card. The shape of course is distorted, because the Hollerith cards' columns and rows are not to the same scale.

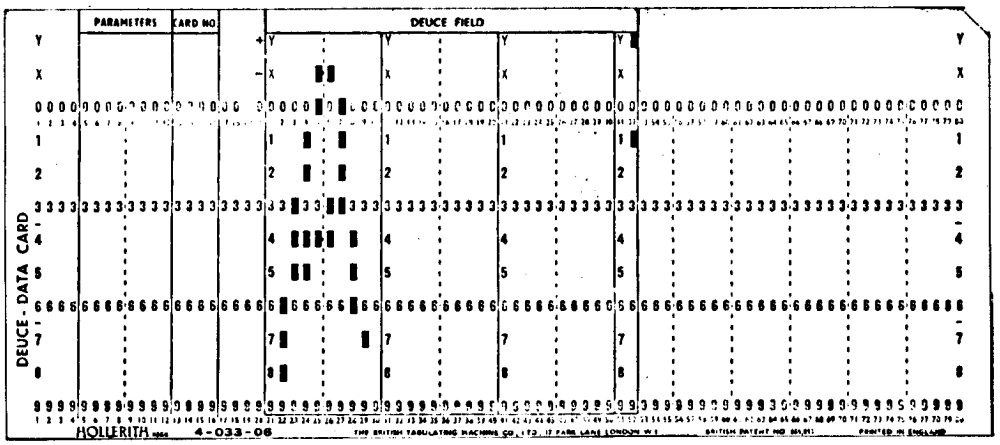


Fig.2.

The results so far appear encouraging for the very short time that this has been used, and I think it does lead to some hope of being able to read this sort of handwriting with a machine which is relatively cheap. One way in which it might be used, is for the sorting of postal letters. There is now a machine in existence for the mechanical sorting of mail, in which an operator punches keys according to the address. I think that the process could be made fully automatic provided the addressing of the postal towns were done in the way shown in *fig. 3*. It should not be difficult to persuade the public to do it in this way on a pre-printed form of this nature; and as I am, like many other people in this conference, a great believer in

A. N. Eskimo,
The Igloo,

N	O	R	T	H		P	O	L	E	

Fig.3.

kicks and rewards, I would suggest that the right way to do it is to say that anyone who writes characters which can be read automatically will have his letter delivered sooner. This should be made possible by arranging that during the development period of this device the postal service is suitably deteriorated from its present high standard.

MR. W. T. BANE: I would like to comment on this question of whether or not to quantise the input signals before they are processed and to draw Dr. Taylor's attention to some work on a problem which has certain similarities to the one we are discussing. I refer to the problem of processing the output signals from a search-radar so that targets are automatically detected.

The design of a search-radar is often such that as the antenna sweeps across a target, a whole series of echoes is received. If there were no noise in the system, the problem of detecting a target would be simple; one would merely have to examine each interval of range and determine whether or not the particular pattern of pulses has occurred.

Noise however is present and its first effect is to distort the pulses that comprise the pattern. Ideally, one would base one's decision of whether there is a target present on a consideration of the particular set of pulses presented. Alternatively, and more simply, one can first quantise the pulses before processing.

A fair amount of work has gone into certain areas of this problem and I should like to quote a three-year-old paper by Harrington (*ref. 1*) analysing the process. Although he uses some heavily simplifying assumptions, I think the results are of interest; he concludes that the effect of quantisation is effectively to raise the lower limit of signal-to-noise ratio (the smallest useable one) by about 2 d.B. This is not much, and he further points out that even if the ideal process were carried out by analogue equipment, imperfections in the equipment might well contribute a similar loss.

I do not claim that the results apply directly to the problem we are discussing; I merely want to point out that in this case, which may be regarded as a rather specialised pattern recognition problem, the effects of quantisation are not great.

REFERENCE

1. HARRINGTON J. V. An Analysis of the Detection of Repeated Signals in noise by Binary Integration. *Trans. I.R.E. Prof. Group on Inf. Theory*, 1955, I.T. -1, 1.

DR. W. K. TAYLOR (in reply): I would like to go back to the question of clipping level because Mr. Tizard points out that it can be varied automatically. The point is that there is only one clipping level per character and no matter where it is the signals below it are neglected or insignificant signals are given full value. By keeping to the analogue signals the correct weighting of all the available information is preserved. The two-state devices required for forming binary signals are far more complex and costly than the resistors that handle the analogue signals.

The simplest way to obtain the x-signals is to construct the transducer matrix and measure them when the characters are presented. The analysis of the patterns is the real difficulty and the transducers eliminate the problem of manual measurement that Mr. Tizard found so tedious, even for binary estimations. It is a simple procedure, once you have the x-signals, to design and construct the y-networks in order to recognize them. One simply starts with all the non-zero x-signals and eliminates the smallest values one at a time until the y-signal starts to fall. The main advantages of the new machine over a computer simulator is it's small size. cheapness and high speed. There is every reason to believe that a machine with sufficiently low resistor values would recognize a million characters/second, assuming that they could be supplied to the transducers at this rate.

SESSION 4B

PAPER 6

AN ANALYSIS OF NON-MATHEMATICAL DATA PROCESSING

by

E. A. NEWMAN

BIOGRAPHICAL NOTE

Mr. E. A. Newman, born 1918, was educated at Kingsbury County School and obtained his degree at University College, London and Birkbeck College, London.

From 1940 to 1947 he worked on radar control mechanisms and television at Electrical & Musical Industries Ltd. Since 1947 he has worked at the National Physical Laboratory on various aspects of automatic digital computors and their uses.

4B-6. AN ANALYSIS OF NON MATHEMATICAL DATA PROCESSING

by

E. A. NEWMAN

SUMMARY

THERE are classes of data processing problems which are more complex than those normally done by automatic digital computors, and which can occur in clerical work.

This paper gives precise definitions for a number of concepts which are necessary in discussing such problems, and does a certain amount of preliminary analysis.

AUTOMATIC digital computers are in wide use for scientific computation. For such work they are most valuable.

Recently much attention has been given to the possibility of mechanising the kind of data processing which occurs in clerical work.

Such data processing can differ from that involved in scientific computation, in two ways which are almost the converse of each other.

First, in some clerical data processing the amount of input data is quite large compared with the amount of processing to be done on it and output is similarly large. Secondly, the data processing is often exceedingly complex. In fact it can range over almost the complete field of human thought.

So far therefore, clerical mechanisation has not progressed very far. Mechanisation has only been attempted of clerical processes which are arithmetic or pseudo-arithmetic in nature. Such processes are just those that tend to give input and output problems, problems which can only be solved by apparatus which is only slowly coming into use in this country.

Mechanisation of the more complicated clerical processes has not been tackled since they involve data processing of a kind that is little understood, and data manipulating processes have to be fully understood before they can be mechanised.

This paper is written against this background and attempts to look at data processing in a very generalized way. It is difficult when talking or

writing about the more complex kind of data-processing problem, to be at all precise. This arises from two related causes - first, few of the ideas involved have been precisely defined, and second, there are rarely words available for the ideas. For this reason this paper contains a large number of definitions.

Most of the data that interest man are about systems, or the interaction between systems. It can relate to influences acting on systems, to internal events occurring within the system as a result, or about the influence the system has on other systems. Always the data must compare or contrast one situation with another. Data can be defined as any entities or concepts which show if situations or experiences are the same, or, if not, specify their differences, and data-processing as any operation on symbols, where a set of symbols is defined as a set of entities of such nature that a certain system can distinguish any one from all the rest, a symbol being one of these entities. A set of such symbols can be called a language. It will be noticed that this set of definitions directly implies the existence of a certain system, and indeed includes this system. This is fundamental; the idea of a symbol is meaningless unless one also includes the system.

To say that a system can distinguish between input patterns either means it is aware of the difference, or that a second system can tell that the input patterns cause different internal or output patterns in the first system. The second system can only do this if the input and output patterns of the first, and the internal state patterns of the first, are in fact input symbols of the second. There are great difficulties here. They are essentially matters of philosophy but these are not matters that can be lightly dismissed on that account. At a simple level it might be argued that it is sufficient to say that a system can distinguish between input patterns if these produce differing patterns within it. But in the last resort this means very little. A man examining the system might come to any one of very many different conclusions about the system, depending upon the sensitivity and kinds of detecting system he uses. Further as Mr. Pask has pointed out *(ref. 2)* it is not possible to discover anything about a system without altering it at least in some degree. To assume a detecting system does not resolve the dilemma - it moves it up a stage. Here, to make the chain conceptually sound one has to assume a little more than this, and assume the existence of some system that has self-knowledge. Systems are able to distinguish between symbols, if this self-aware system can tell that the response to the symbols is different. This leaves the difficult problem of what is meant by awareness unanswered.

It is easy to see that the language of a system is dependent on the way other systems react with it.

It is evident that a great deal of data-processing involves the recognition of pattern, and judgement as to whether patterns are alike or not. This involves the question of what a pattern is, and what is meant by

likeness. These are not simple questions. In particular it must be noted that two patterns which would be judged very like in one context would be judged as quite unlike in another. However, starting from a few more definitions one can build up an idea of pattern, and likeness between patterns which on the one hand is mathematically precise, and on the other hand is consistent with intuitive ideas. It is possible to have symbols which are built up of more elementary symbols. For example, associated with a machine with 10 input wires, each of which could have a potential on it or not there would be 20 elementary symbols. The machine might be able to recognise combinations of these and so distinguish 1024 "built-up" symbols. Elementary symbols one can call "marks" and built-up symbols patterns. A mark in one pattern might be itself a pattern made of more elementary marks. It should be noted that marks can only make a pattern if each one bears a special relation to at least one other such as, for example, a spatial relation. One can define a symbol P (a pattern) which contains a symbol Q (a mark) as being of higher rank than Q. If system A divides a sub-language T of a system B into sets, S_1 of symbols containing no other symbols, S_2 of symbols containing one or more of S_1, S_3 of symbols containing one or more of S_2, and so on, then one can define these sets as being of symbols of steadily higher rank. If two patterns have a common mark as part, one can define them as alike.

One can also define sub-rank within rank if one says that the highest sub-rank symbol of two symbols of equal rank is that which contains most marks of one rank less. One can make similar definitions of sub-sub-rank.

The following formula, based on the above conceptions can be used to give a mathematical definition of likeness.

If the ranks of two patterns are R_1, R_2, and that of the common marks are R_a, R_b etc. the degree of likeness can be defined as

$$L = \frac{R_a}{R_1 + R_2 - 2R_a} + \frac{R_b}{R_1 + R_2 - 2R_b} \text{ etc.}$$

This also gives a definition of X is more like Y than Z is like Y. Where patterns are the same this formula gives a likeness of ∞, where they are quite unlike it gives a likeness of 0. The formula will be found to fit in with intuitive ideas of likeness. Very like things have major sub-patterns in common and only differ in small points of detail. Very unlike patterns only have small points of detail in common.

It will be noted although the likeness so defined is quite definite provided the systems A and B are both defined, it is not independent of context. The calculated ranking scores of symbols of B will depend greatly on which sublanguage of B is taken, and this will affect the likeness score. This is also consistent with intuitive ideas. The estimated likeness of two patterns depends both on the observer system - be it man or machine - and on the context, and this is in fact due to different observer systems using different sub-languages of the patterns.

One possible kind of data manipulation task - one which on reflection will be found to cover a very wide range particularly in more difficult data-processing tasks - is that of one system, the learner, discovering the language of another, the subject.

There are two quite different kinds of task that could be involved, one in which the learner is solely an observer, and the second in which the learner and subject are part of an interacting system *(ref. 2)*.

In the first case one must assume that the symbols of the subject change from time to time, due to some cause outside of the control of the learner. The learner must now note what changes in the system occur as a result of these changes of the symbols. In the other case the learner in fact itself causes changes to the symbols of the subject, in which case it can be called a participating observer. It is convenient to divide the subject system into three parts, although such a division is a little artificial. These parts are the input organs to the system, the body of the system and the output organs of the system. The only relevant input, output or body states of the machine in the context of the learning situation are those which are symbols of the learner system. If the learner is a participating observer, then some of the states of the subject are output states of the observer.

At first sight it might appear a simple matter for a learner to discover which are input, output and body states of a subject. But if the learner has no prior knowledge the last is in fact quite difficult. The problem is best considered in terms of direction of information flow.

A number of symbols of an observer might occur coincidently; when this is so it happens that the probability of B occurring given A is 1, A implies B. If at the same time the probability of A occurring given B is less than 1, A contains B. If one assumes that a number of symbols can occur coincidently with A or B or both, and there exist a number of sets of symbols S_1, $S_2 \ldots S_n$ (including the set of no other symbols), then if for each S the probability of A given S but not B is the same as that given S and B, A is independent of B. It can be seen that the task of finding that A is independent of B is more difficult than that of finding A contains B. There is the similar concept that A is partly dependent on B. This is so where A is not independent of B, but the probability of A given B is less than 1.

If A contains B and is independent of it, then there is information flow from A to B.

If in a system there is no information flow to A, A is an input state, if there is no information flow from A, A is an output state, if A is neither an input state nor an output state, it is a body state.

It can happen that an observing system will find that, in a subject system the same input symbol always causes the same body state and this in turn the same output symbol. In order to make this general, the various

symbols might contain a time element and be a time sequence of more simple marks. In what follows in this paper it will be said that, if an observer finds a subject's behaviour to be invariant, it says the system is static. A system which is not static in this sense will be called dynamic.

As can be seen the property of a subject system being static is dependent on both it and the observing system. Another observer might say the subject is not static. For example, if an observer treats numbers in serial form as entities, it will class a serial adder as static - if on the other hand it only reacts to separate binary digits it will classify the system as dynamic.

A system can fail to be static for two reasons. First, its behaviour can be dependent on the past history of input or body states - where this is so it is said to contain storage - or secondly the body or output states can contain (or appear to contain) an element of randomness, when the system can be said to be noisy.

All learnt properties of systems are partially those of the learner. Hence one learner might find a subject to be noisy when another finds it not to be. For example if the two observers can observe the same body and output states of a subject, but there are input marks of the subject which can be observed by the first observer, but not by the second, then the second will believe the effects of input symbols containing these marks to be subject to noise, while the first observer will not.

It is evident that there must be a 1:1 correspondence between input symbols and body states, and that the number of output symbols must be less than or equal to the number of body states.

In so far as the machine state depends on history of input symbols one can say it has accessible storage. In the context used previously, a learner machine must have storage. The task of the learner machine in discovering the language of the subject is to find which body state changes of the subject correlate with input changes, and find which input symbols cause which of these body state changes. It also has to find which body states corresponds to which output states. A further task is to find existing hierarchy amongst input symbols, body states and output states but this in general a means to an end rather than an end in itself. It is evident that the complexity of the learning task depends greatly on the nature of the subject. If this is static, the task is relatively easy. If it is noisy, and contains inaccessible storage, the task is very much more difficult.

To learn fully the language of a subject, a learner must find every correspondence between possible input states and body states. Any combination of input mark symbols might prove to be an input symbol. The learner must wait until every possible combination has occurred, and store the result.

In practice if one calculates all the patterns that could be made from a set of marks one finds in general it vastly exceeds those patterns that are in fact liable to occur. This is another way of saying there is a great

deal of order in the world we know. Further those patterns we meet tend to be very much alike.

A learner machine that makes use of these things when studying a subject will in general do immensely less work than would one that did not. If the system under study is at all complex the total number of possibilities is very great. A pattern containing only 100 different marks is quite trivial as patterns go. Yet the total number of patterns which could be built from 100 marks is of the order 10^{30}. The time taken to examine this number of possibilities - even if one did them at 10^{9}/sec is very large even compared with the probable age of the solar system. It is thus impossible to consider all possible symbols.

Only very few of the possible symbols are of interest, and further even these few tend to be very alike. If one learns from experience in one field what kinds of patterns do occur, one can usually assume that patterns occurring in another field will be similar.

If a learning system starts from no knowledge, there is only one course open to it. It must at first limit its input symbols to a very few. All patterns then appear to it to be relatively trivial. It can study these by considering all possibilities, and eliminate from further consideration those that appear unlikely to exist.

It can then increase the number of its own input symbols in easy stages, and so study finer and finer detail. At each stage it must use what it has learnt at an earlier stage in order to cut out of consideration most of the patterns which are theoretically possible. Thus in due course it builds up a complete hierarchy of pattern.

In the general case, the order in which detail is introduced is important. One will only converge to a complete ranking if it is correct, and there is no way of knowing correctness. However, in practice where only a very few of all possible symbols exist, and where all complex symbols can be divided into a few groups of similar ones, the order is much less important.

When a learner system has a great deal of information about the structure of probable patterns built in, it can study a new system in various ways. One way is to investigate the subject system using only a limited number of symbols, and to extract those stored patterns which, when examined using the limited number of symbols, are very like that being investigated. There should only be a few of these. The selected patterns can then be examined and compared with the subject pattern using the full language of the learner. The probability is that the patterns will, even at the fine detail of inspection, be very like. Small variations can then be made to the selected patterns until identity with the subject is achieved.

A slightly different technique is to examine the subject pattern for marks of the lowest rank, and see how many stored patterns in the learner

contain this set of low ranking marks. If the number is large, examine the subject for symbols of next higher ranking, remembering that these must be built only from those lower-ranking symbols already found. Again compare these with stored patterns. Continue this until the number of found stored patterns is small. When this stage is reached, make small variations to the found patterns to achieve identity. It will be seen that when we say we use argument by analogy, we in fact are usually using one of these two methods.

One kind of task that is involved here is the manufacture of a pattern from marks, or from another pattern and marks. For example if a machine is attempting to find, in its store, a pattern which is the same as a given one, but instead finds all the relevant marks, and a pattern which is very like the given one, it might prove quicker to form the pattern, than to find it out of store. It should only use relevant information. The following definition fits what is usually meant by the term relevant. Let us assume that a problem involves certain symbols of a system A. In the context of the entire language of A these symbols will be ranked in a certain way. In general sub-languages of A will not give the same rankings, but there will exist sub-languages of A which will. In particular there is a minimal sub-language that will do so. This sub-language consists of those symbols relevant to the problem. The finding of such a sub-language is a relatively small extention of ranking of group of symbols.

It is easy to see that in order to operate in the sort of way described, the learner must be a participating observer. It is also necessary for the learner to find if certain patterns are included in others.

In the generalized sense how can one find if one pattern is included in another - how can one rank symbols? In general one can never be certain. However the idea can be put in terms of conditional probability *(ref. 3)*. Let us suppose that the observer has continually varying input symbols. It might be found that pattern *A* is always, or nearly always found when pattern *B* is present, and *B* when *A*. Then *B* and *A* are the same. Another possibility is that although *A* is almost certainly present when *B* is, the reverse is not true. Then *A* contains *B*.

Supposing we have symbols *A B C D E*, where 10010 would mean an event where *A* and *D* occurred but *B C E* did not. Then if the following sequence occurs:- 11111, 11101, 01101, 01001, 10100, 11111, 11111, 01001, 11101, 00101 00000, 10100 it is possible to deduce that *A* contains *C*, *B* contains *E*, and *D* contains *A* and *B*.

In order to perform a learning process in the way so far discussed, the learner must make use of stored information.

It must either take the symbols it tries from its storage, or it must create them, or it must do a little of each. To create symbols to try, it must be noisy.

It is evident that the noisy system needs the less storage. It is not so evident that a learner which creates its symbols of necessity goes

through a far greater sequence of operations to learn the subject's language and response than does that which obtains them from a suitable organized store. The problem of good store organization is not trivial.

For some tasks such as wage accounting efficient techniques can be found even when using widely differing equipment. Other tasks are surprisingly difficult to mechanize efficiently, either with existing equipment or with any which can be foreseen. Such tasks are library retrieval and index problems.

The difficult tasks are found to involve very difficult information retrieval aspects and the easy ones not. The root of the problem of information retrieval is very difficult to think about: largely because of this, many people have looked for the panacea of an "ideal store". Unfortunately this cannot exist.

Information retrieval is always part of a larger problem, such as the problem we have been considering; some information is available and more is required. This can be obtained in some cases by working it out from information present, but usually it has to be got out of a store. To be efficient such a store must have a great deal of structure and the users must know this and take full advantage of it.

Suppose we have a store with much structure - enough to enable us to inspect every bit without duplication. If there are N bits in the store and we need P bits, the quantity of information we must handle is of the order of N^P bits ($\frac{1}{2} \times N^P$ if we know the value of P). If we can divide our store into B blocks each of P/a bits then we need handle only $(aN/P)^a$ bits of information at a cost of only $B \log_2 B$ bits of data processing. If we can organize the store so that the label is the route to the record, the extra data processing is reduced to $\log_2 B$ bits. We get an enormous saving in data processing if the block size is correct, with a spectacular increase if it is too small.

This implies (a) the designer of the store must know beforehand enough about all possible problems to know the best block size. (b) the problems must be such that the clue contains the information required to select the record and requires little data processing. (c) the designer of the store must know all possible clues so that he can juggle with the labels so that they have a 1:1 correspondence with the clues. (d) the clue should give the route to the record.

We can now see why the wages problem is easy, the others hard. In payroll the nature of every problem can be foreseen, the records can be arranged so that the clues get one exactly to the data needed. In library retrieval one cannot foresee the problems which will arise; one cannot break the records into small blocks. There is a possibility of using a 2-stage process, first selecting a smaller unit than a book or report, say a paragraph, then estimating the probability for the book.

Similarly in indexing we cannot always be sure what the most significant clue will be.

If we do not take our symbols from store, but create them at random, it is evident that we are in a slightly worse position than we would be with a store containing no structure, and a greatly wrong block size.

It can be seen that the learner that makes systematic use of a correctly organized store will need to do very much less work in a typical learning task than one that makes random shots. The storage system will have to be such that low ranking symbols used as clues lead directly to the more probable patterns containing them. If the information has been built up as a result of experience it is essential that the storage system is of such a nature that it reorganizes its internal structure to match any change of information resulting from new processing.

There is a class of problem which is inherently more trivial than the learner subject problem. This is that which can be performed by a static noiseless machine. In this sort of machine the body state is uniquely determined by the input pattern, and in turn, uniquely determines the output pattern.

It is easy to show that any pattern can be built of binary marks, to any specified degree of accuracy, given sufficient of them. Hence if the static machine can be made to transform an input binary pattern to a required output binary pattern it will be able to approximate with any degree of closeness to the transformation of any pattern to any other pattern.

These are many proofs that it is possible to make a machine which will cope with any definable binary transformation. One is illustrated in *fig. 1.*

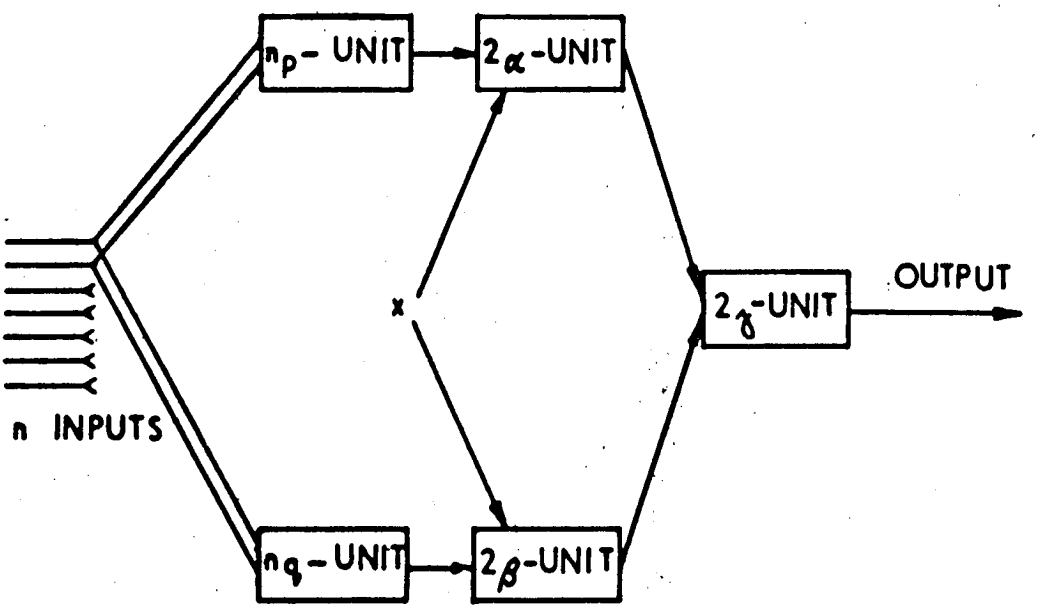


Fig. 1. Build-up of $(n+1)$ - unit from n-units and 2-units.

Let us assume a unit with n binary inputs, and one output. The output will be 1 or zero specified for each of the 2^n possible values of the n inputs. There are 2^{2^n} such units.

Now let us assume we wish to construct one of the similar units for $n+1$ inputs. This can be done from n-input units and 2-input units. Let the extra input be x. The output of the unit $n+1$ must be that of a unit n_q if x is constrained to 0, and of another unit n_p if x is constrained to 1. We can take a unit n_p and a unit 2α which with inputs a and b, has an output a when $b = 1$, and zero when $b = 0$.

With x and the output of n_p as inputs to such a unit we get out n_p when x is 1, and zero when $x = 0$. Unit 2β, output zero when $b = 1$, and a when $b = 0$, in conjunction with n_q gives the other half of the result. If the outputs of units 2α and 2β, are fed into unit 2γ, the required result comes out, provided 2γ is a unit that has the output 1 when its inputs are different, and 0 when they are the same. Since it is trivial to show that all the units 1 can be made (there are four of them) the theorem is proved.

It is also trivial to show that if any unit with n inputs and one output can be made, then any unit with n inputs and m outputs can be made, by joining the appropriate m n-units together.

There is in practice only one kind of technique possible to transform one set of patterns into another. The input patterns have to be broken down into their component parts - parts of lower rank and the higher-ranking output patterns then have to be built out of the parts. It is not in general necessary that the original patterns should be broken down to zero rank marks, although this is always sufficient.

The synthesis part of the task involves building up patterns of steadily increasing rank. This is essentially a sequential process. On the other hand, the build-up of a number of lower ranking parts, all required in the final pattern, can take place in parallel. Some aspects of a pattern transformation task can therefore be done in parallel - others are essentially sequential. If one inspects a typical overall pattern of a static converter, one frequently finds that it contains within it many copies of quite a few basic patterns. This often occurs at quite a few levels. For the overall pattern often contains several copies of quite high-ranking patterns, these in turn several copies of more simple patterns, and so on. One can make use of this if one uses a dynamic machine instead of a static one.

It can be seen that the power even of a static machine, suitably designed, is very great. It can for example, translate from the language of one machine to that of another - from English to French. However we must remember that a language, not only in the sense we have defined, but, in the widest sense, is a function of a particular system. Tom Smith's English is not that of Bill Jones.

It is not necessarily possible to transform one set of symbols into another without being given all in the first place. Human languages have

been developed so that to a considerable degree each little bit of language pattern is complete in itself. But this is not and cannot be entirely so. Human language is also designed that certain low ranking symbols are each contained only in one particular very high ranking symbol. Thus the amount of information carried in the written symbol 'dog' is quite small - of the order of 15 bits. Yet the information contained in the symbol it represents is very complex, and the information needed to describe every aspect of dog or dogness to an intelligent being with no knowledge of life or phenomena on the earth is tremendous. When we write or speak we use the fact that the reader or listener already knows a great deal about the objects and ideas we refer to. Without the background knowledge, the message would be virtually meaningless. For this reason translation of the word 'dog' in the language of our non-earthly being might take thousands of words, or might not be possible.

Although the manufacture of a static translator is possible in theory, in practice it would be next to impossible, because of very great number of units required, and the very great difficulty of designing it. It is even uncertain that a first rate translation could, in practice, be obtainable using a dynamic machine. A large static machine if built of simple units is, in sense previously discussed, a large pattern. In general it contains within it various patterns of lesser ranking. For example a static multiplier designed to multiply a 100-digit number by a 100-digit number could be so organized as to contain 100 100-digit adders, each adder 100 single-digit adder, and so on. Essentially a dynamic machine saves equipment by utilising this fact. Where the equivalent static machine has a suitable ranking pattern the dynamic machine is simple, otherwise it is not. The necessary analysis of a translation process from one human language to another has not yet been done. When it has been it might be found that at times groups of words which cannot be broken down to short serial smaller groups is very large, and that the amount of information about the pre-knowledge used by the writer of one text and the reader of the translated text might prove great.

The powerful electronic computing machines now in existence are dynamic noiseless machines, for which many of the essential parts can be readily and cheaply made by man. These parts are in fact the specified program steps. In getting out the program in the first place the programmer is acting as the learner machine.

There would seem to be no good reason why a digital computer could not be programmed to be an efficient learner machine. It would either have to be fed initially with a great amount of information about the inter-connections and probabilities between the symbols of man, or else it would have to pick these interconnections up by acting as a learner machine in at least as wide a context and for as many problems, as does man. To do this it would have to have tremendous storage capacity, presumably as much as

man has. Furthermore in a machine that relies almost solely on experience the storage in it would have to be arranged on a very flexible basis, since simple clue symbols must always lead rapidly to the most probable stored complex patterns. As the machine gains experience the probabilities alter, and the store organization and structure must be modified to correspond.

If the class of subject the learner is expected to examine is suitably restricted, the storage required might possibly be relatively small, and where experience is built in initially the store has less need to be flexible in organization.

REFERENCES

1. DAVIES, D. W. Switching functions of three variables. *Trans. Inst. Radio. Engrs.*, 1957, EC6/4, 265.
2. PASK, G. Physical analogues to the growth of a concept. Session 4B, paper 7.
3. UTTLEY, A. M. The classification of signals in the nervous system. *E.E.G. Clin. Neurophysiol.*, 1954, 6, 479.

SESSION 4B

PAPER 7

PHYSICAL ANALOGUES TO THE GROWTH OF A CONCEPT

by

GORDON PASK

BIOGRAPHICAL NOTE

Gordon Pask was born on 28th June, 1928. He was educated at Rydal School, Colwyn Bay, North Wales, where he developed a strong practical interest in Geology.

He went on to study Chemistry and Biology at Liverpool Technical College, and commenced informal research at home on organic analogues. He then read Physiology in Natural Sciences at Downing College, Cambridge, and graduated in 1953.

System Research Ltd., a Cybernetics Consultancy firm, was founded in the same year. Since 1956 Mr. Pask has been acting as Cybernetics Consultant to the Solartron Electronic Group Ltd., for whom he has developed the family of teaching machines.

He has published a number of papers on automatic teaching techniques and teaching machines.

PHYSICAL ANALOGUES TO THE GROWTH OF A CONCEPT

by

GORDON PASK

1. INTRODUCTION

IN this paper I discuss the circumstances in which we can say a machine "thinks", and a mechanical process can correspond to concept formation. My point of view about this question is as follows. It is reasonable to say that a machine does or does not "think", in so far as we can consider the working of the machine as in some way equivalent to a situation or an activity, (for example, riding a horse), which is familiar, and in which we ourselves are used to taking a part. Thus, when I speak of "thought", (as when saying a sonata is written, or a hairpin is invented, as a result of "thought"), an end product is introduced on which to hang the thinking process. The process itself is a descriptive expedient, a kind of analogy. Clearly the sonata was not written "by thinking", (in the sense of "by magic" or "by using a computor").

Thus, my view of thinking can be expressed in terms of the concepts "participant observer" and "external observer", as these terms are used by Colin Cherry *(ref. 6)*. If we assume that such an "external observer" watches the process of writing a sonata he will seek to describe the stages of the process and he will have no need to speak of the "thinking". On the other hand, if an observer does speak of "thinking" in such a context he wishes to assert, according to my view, that he was not purely an external observer, but to some extent participant.

Since it is the participant observer who, by the present hypothesis, uses the term "thinking" correctly, let us consider his description. For him thought is taking place about some end product, and although the nature of the end product tells us very little about the "thinking" as such, it does say something about the way that the observer examined the subject, (or going now from our common examples to thinking machines, about the way he examined the machine submitted for test as a thinking assemblage). Moreover, the particular observer conceives that the sonata and the hairpin were constructed as he, or we, might have constructed them, though he will be unable to say, in so many words, how he *would have* constructed them himself.

I take the construction of a new concept as typical of effective thought, and propose to use the experimental material provided by Bruner, Goodnow and Austin *(ref. 5)*, because it bears out current views on concept formation, is in a form appropriate to the present needs and because their whole descriptive technique is in terms of the theory of games.

Very roughly, at the partly introspective level, these experiments suggest that a thinking process boths builds up and employs conceptual categories. These categories are defined in terms of attributes, which may be common to a number of objects in the environment, or to other categories or to both.

At each stage in the thinking process a decision is made about whether an object should be placed in one or another of these conceptual categories. Such a sequence of decisions is a thinking strategy. The human being tends to regard these conceptual categories as definite and well bounded. But, objectively the categories are not clear cut, and decisions appear to be made between imperfectly specified alternatives. The categories are learned, or equally well they grow as a result of the strategies adopted, and it is not possible to extricate the category building from the decision making process.

The authors cite the case of a histologist, who is learning to categorize microscopic structures into those which are or are not a corpus luteum. He starts off with attributes like colour, and shape, which somewhat inadequately define the category of corpus luteum structures. He adopts certain strategies in his search, and as a result of these he modifies the original categories so that the objects are now specified in terms of a structure appropriate to his particular approach. Eventually he acquires what could equally well be called a mode of search behaviour or a "labile category". Bruner, Goodnow and Austin *(ref. 5)* call it the concept of "Corpus Luteumness", and liken it to a "gestalt". The overall process is the growth of a concept.

The experimental and descriptive techniques used by these authors and the connection between the technique and the process of concept formation enables us to understand the action of a participating observer when the "thinking system" is a machine. Bruner, Goodnow and Austin started off by examining a lot of subjects without any particular bias, and arrived at a method for describing the thinking process. They decided upon a method of describing it in terms of thinking strategies, the alternatives in the choice sets in the game being "conceptual categories". They then formulated a number of matrices, and a kind of "calculus", whereby these matrices could be treated like the payoff matrices in a partly competitive game. The entries in these matrices are those elements like "hairpin" and "sonata" which one agrees to treat as concepts. The formal mathematical operations with these matrices, (which are those operations studied in the theory of games), are those operations an *external* observer would recognise as played

according to the rules of the game, i.e. according to strategies he might have adopted. These strategies are then to be related to the thinking strategies which the thinking subject actually indulges in by recording his decision concerning the objects that have been agreed to represent concepts. If the solutions follow any of the courses set by the formal mathematician, it is argued that the subject is adopting a strategy more or less like this strategy or that.

Bruner, Goodnow and Austin are talking about real subjects with whom conversation in the normal sense is possible, and who can discuss details of experiments. Their arguments would not necessarily apply if the real subjects had been replaced by mechanisms. An essential feature of this argument is the tacit assumption that the entries in the matrices correspond realistically to "concepts". This assumption is made because of evidence which assures the observer that he and the subject are comparable, and which, in the sense of belonging to the same species, and therefore presumably of having a large fund of experiences in common, we conveniently summarize by saying that the subject "thinks". According to the present hypothesis such a similarity has to be inferred between an observer and any assemblage he may hope to describe as a "thinking assemblage". We must now ask what sort of evidence is needed in order to establish this similarity for the observer has no "culture" in common with the machine.

Now I have already assumed that it is possible to attribute concept formation to something outside of myself, if, and only if, there is a field of activity common to myself and the system concerned, and that if, for example, a chimpanzee has "grasped a concept", it is because I can imagine myself having learned from experience in somewhat the same way. In the case I have already mentioned of the horse and rider, again, the rider might say the horse "thinks" because he participates with it in solving the problems that are set by a common environment, namely the topography of the place in which the horse is ridden.

When, however, we want to discuss observers - those that are external to the systems observed and those that participate - what is the "common environment" or field of study that is presupposed? I suggest that it is the whole of what we know - vaguely as well as precisely - about The Brain. Indeed, I think to get an idea of the participating observer by constructing machines, you are bound to copy the way one looks at brains. You must, somehow, keep the brain in mind, and in this sense you do copy the sort of relationship we have with brains. There is no question whatever of copying the detailed anatomy of a brain, or the detailed physiology of a brain. Therefore, it is of interest that when we have copied, in this not very explicit way, how we look at brains, in order to construct an assemblage we find that the assemblage *is* rather like a brain in these respects.

I conclude this introduction with a definition. If an observer, by participating in the action of a mechanical assemblage, on the supposition

that he is to compare the assemblage with the action of a brain, and comes to attribute concept formation to the assemblage in this way, I shall say that the observer is in an E. relationship with the assemblage.

SECTION 2

2.1. Using the analogy of a piece of brain, what considerations will influence our choice of an assemblage? The assemblage must certainly satisfy two distinct sets of criteria. The first set of criteria stem from the requirements of any scientific observer, and are needed in order to make the assemblage worth observing from his point of view, namely, the viewpoint of someone examining brain-like-artefacts. The second set of criteria are those required by a 'Participant Observer' as already defined, and which must be satisfied (in his view) if he is to establish an E.Relation with the assemblage (and thus to regard it as a structure able to form concepts, in the sense that assuming this, and acting accordingly, enables him to control the assemblage).

The first set of criteria have been discussed by Beer. *(ref. 3)* in the context of Industry and general cybernetics and by Ashby. *(ref. 2)* in connection with 'Black Box' theory. Since they must be expressed, for the present purpose, in terms of conditions upon the working and structure of a physical assemblage which is constructable, rather than given in nature, these criteria will now be listed in the manner required.

2.2. *The first set of criteria, as required by a scientific or, 'External' observer.*

1. Since the assemblage purports to be a constructed mechanism it must be made of components which have one or more possible functions which are known about, and which are put together in a way which is revealed to the observer.

2. The behaviour of the assemblage must always be observable. Since the structure of the assemblage has been taken as known only the state changes of the assemblage are in the field of possible observations. Thus, the above requirement means that the assemblage must continually change state.

However we may invoke the general principle that a real observer has a finite capacity for observing an assemblage (namely the idea of quantised observation as considered by MacKay, *(ref. 9)*) to relax this condition, so that it will be sufficient if the assemblage changes state within each of the shortest intervals in which an observation may be made.

3. The observer must have reason to believe that underlying the state changes of the assemblage, there is something describable, a sort of consistency, or, in other words, that it would be possible, if he were a good

enough observer, to recognise invariant features of the behaviour, sufficient for him to make sense of it.

Such a description (or 'Model' as the term is used in 'Black Box' theory) could, if available, be isomorphic with the assemblage in the sense that there could exist a one to one relation between entities in the model and the assemblage. Thus manipulation of entities in the model would provide an accurate image of the assemblage and vice versa. However the finite capacity, or quantising condition, noted in (2) above implies that an isomorphic model will not be available to a real observer because he will be unable to distinguish sufficient observable states.

In this case, the consistency condition asserts that the imperfect model which is described should -if possible- be homomorphic with the behaviour of the assemblage. Such a model is obtained if the states, discernible to the observer represent a certain kind of partitioning of the ideally observable states.

Thus an observer might, ideally, be able to distinguish between the states α, β, γ and δ but due to his imperfections he may, in fact, be unable to distinguish between α and γ, or β and δ which we symbolise as a partition and by writing.

$(\alpha \cup \gamma) \subset X$ and $(\beta \cup \delta) \subset Y$ for the observable states X and Y.

But only certain kinds of imperfection, and partitioning, are allowed if a construction in terms of the observable states X and Y, is to be the homomorph of the ideal constructions of the states α, β, γ, δ, In general, it is sufficient to insist that the transformation which maps the ideal states of α, β, γ and δ, into the imperfect observer's observable states X and Y, is a partition which maps α, β, γ, δ, into non-overlapping sub-sets of themselves. In this case, suppose that the set of state transformations which specify the behaviour of an assemblage as it would be described by an ideal observer, (with unlimited access to its interior), form a group, and that this group is specified by such an ideal observer, (possibly with some conditions applied), as representing the behaviour of the assemblage, i.e. as a model of its behaviour. If the imperfections of an imperfect observer, which will, in any case, make the ideal model unavailable, are of the particular kind noted above, it will be possible for the imperfect observer to achieve a model which, though less informative than the ideal model, is consistent - which does not contradict though it may not always provide reason for - assertions made by the ideal observer and which is mathematically a group homomorphic with the original group specified by the ideal model.

4. The groups, noted above, must be finite. If they are, the possible outcomes of state changes in the assemblage will be predictable, so far as the observer is concerned, and in this case the assemblage is considered as recognisable, in the sense that the observer can talk about it as an entity in its own right, as something with a consistent pattern of behaviour, and a function relative to other entities.

5. Finally, there is an overall requirement of non triviality, which is best exemplified by reference to redundant and non redundant data. Thus, having agreed to a certain reference frame, namely in this case, having agreed to concentrate upon the state changes of an assemblage, being assured about its structure, the observer has every right to expect that the possible observations he can make are not redundant, within this agreed reference frame. If, for example, the structural specification allowed him to deduce with certainty that if any state changes occurred, there would be an observable sinusoidal fluctuation in some measured quantity at a point "X". Though not, perhaps, allowing him to specify its frequency or amplitude, the fact that fluctuations at "X" are sinusoidal is called redundant data and its observation is not counted for the purposes of 2 above. The observable state changes of 2 are such that they may not be predicted by deductive manipulations of the a priori data. The frequency at "X" and the amplitude at "X" might be admissible measurements to make in the sense that they might indicate state changes which are not redundant, but even if they are admissible in this formal sense the observer will not necessarily regard them as relevant. Thus, in order to be nontrivial the observable state changes must satisfy another and very important condition, namely that their observation implies making measurements directed towards answering the enquiries which appear (to an observer who has agreed to adopt a certain frame of reference) as relevant enquiries.

2.3. *Reference Frames*

In Section 1, we described how, to assert the property of thinking in a system of any kind, it is necessary to have in common with the system some sort of context or common field of experience. We now have to make the idea of context or common field of experience mechanically tractable by describing and defining "Reference Frames." A reference frame is a region of knowledge or a region of connected and tentatively confirmed hypotheses. Thus, for the immediate purpose we assume that any observer has some initial knowledge of the assemblage which he is observing, (say, data about how it is built), and that he has an objective, to achieve which he must reduce his uncertainty regarding its behaviour. In this case he reduces his uncertainty by making experiments which involve trials or enquiries and will continue so long as -

(i) The results are self consistent, in the sense of 3 and 4 above,

(ii) The results agree with predictions based upon his initial knowledge which for the moment we assume well founded.

The kinds of enquiry and, in particular, those attributes of the system which an observer deems important, depend not only upon how much he knows of the assemblage, but also upon -

1. How this initial knowledge is distributed.
2. His objective in making the enquiry.

The set of all possible enquiries which is defined on specifying the details of 1 and 2, as above, will characterise a reference frame.

Some reference frames, for example, "electronics" where we always measure the capacity, rather than the colour, of a condenser, and "mechanics", where we examine well specified parameters of distinct parts in a machine, i.e. the intake rate at a carburettor, are well specified reference frames in the sense that the set of enquiries relevant to all possible objectives of an observer is unambiguously defined. Because the observer is aware of what is relevant and thus, of what may be regarded as extraneous and of what imperfections may be allowed, his precision need not be great. Although a less precise observation loses specific points of detail, the approximate statistical results remain consistent, as in 3 and in 4, and this is of the utmost importance when, either due to his own limits, or to extraneous disturbances the results are necessarily rough and ready. The limiting case of imperfection occurs when the assemblage is a machine, (with its parts well defined), intentionally built to prevent the observer having access to its state, (and this system is usually called a "chance machine"). The observer refers to the behaviour of such a machine as producing a non stationary, or an indeterminate sequence of distinct events.

Knowledge of those enquiries which are relevant for a number of objectives which he *might have* adopted implies that an observer may, in the first place, communicate the result of his immediate enquiry to other observers, with possibly different objectives, and secondly may combine the results from a specific enquiry to substantiate or deny hypotheses of a more general character. (This process will be illustrated with reference to *fig. 1.* I shall call a reference frame in which this process is always possible a "well specified reference frame").

There are many systems where the process is impossible and the reference frame is not well specified, and which, as a result, appear more or less indeterminate to the observer. True, the indeterminacy is due to some kind of ignorance, but whilst in the case already considered which in its extreme form leads to a "chance machine", the observer was unable to obtain precise knowledge about a state of the observed assemblage, there are other cases in which he is ignorant of what states it would be relevant to specify, regardless of whether he could specify them precisely enough if he tried. An economist, for example, is usually unable to indicate the "appropriate" measures of society and has no satisfactory model to represent its behaviour and in examining a brain we encounter the same difficulties as the economist. Because of this we shall investigate firstly those features of an assemblage which prevent an observer knowing what enquiries are relevant and secondly the design of a machine or assemblage, in which relevance criteria are made difficult to come by.

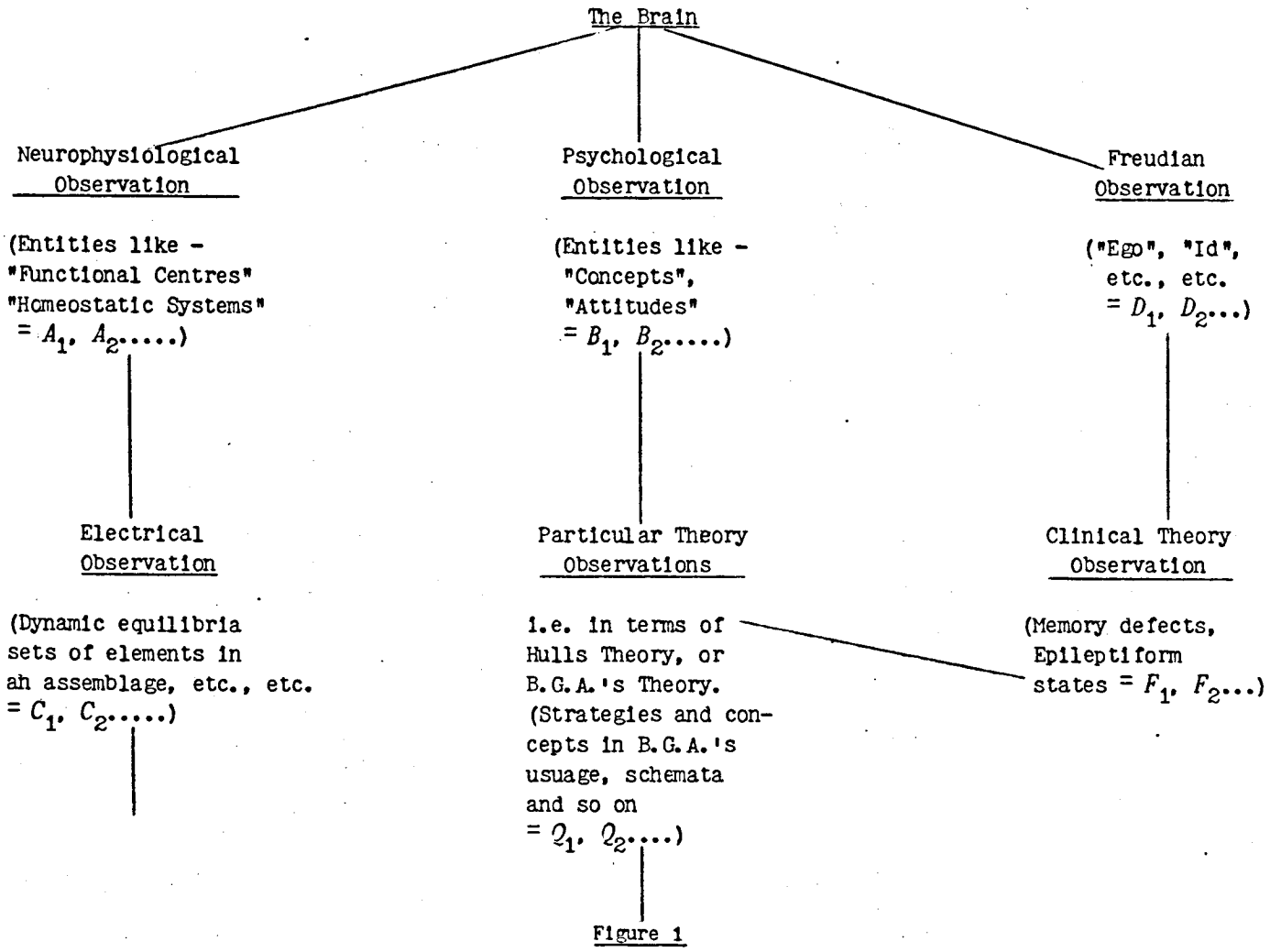

The Brain
Neurophysiological Observation
(Entities like - "Functional Centres" "Homeostatic Systems" = A_1, A_2.....)
Psychological Observation
(Entities like - "Concepts", "Attitudes" = B_1, B_2.....)
Freudian Observation
("Ego", "Id", etc., etc. = D_1, D_2...)
Electrical Observation
(Dynamic equilibria sets of elements in an assemblage, etc., etc. = C_1, C_2.....)
Particular Theory Observations
i.e. in terms of Hulls Theory, or B.G.A.'s Theory. (Strategies and concepts in B.G.A.'s usuage, schemata and so on = Q_1, Q_2....)
Clinical Theory Observation
(Memory defects, Epileptiform states = F_1, F_2...)

Figure 1

2.4. *The subdivision of reference frames.*

Any reference frame may be broken down into a number of regions of knowledge which are self consistent and which will be called "Sub Frames'. The reference frame which has been selected for the present discussion, namely "the Brain' may be reduced to 'Sub Frames' like 'Electrical observation of the brain' (characterised by those enquiries possible for an observer who is provided with electrodes, an amplifier, and recording equipment and who may both stimulate the brain and move his potential sensing electrodes about its surface) and 'Laboratory psychology of the brain' (characterised by all those enquiries possible for an observer who is able to employ physical and psychological tests of the whole organism).

Certain of these sub frames include others, and all of them are included by the original reference frame. A few such relations are shown in *fig. 1* where the entities, in terms of which an observers objective is specified and about which enquiries are made have been defined as A_1, B_j, ... and so on according to the sub frame *(A), (B),.....* to which they relate. The results of actual observations are denoted, again according to the sub frame in which they are obtained, as a_u, b_v,..... and sequences of such observations as $a*$, $b*$,..... and so on.

Certain sequences of observations are taken to confirm hypotheses which propose the existence of the entities A_1, B_j which have been defined above. The sequence $a_1^* = (a_{u,t}, a_{v,t+1}, \ldots\ldots a_{s,t+\tau})$ at instants $t, t+1, \ldots\ldots t+\tau$. might, for example, be taken to imply A_1.

Because any real observer is limited as in condition (2) he will not be able to make direct enquiries about the brain as a whole. However, he may submit hypotheses about the brain as a whole, namely an hypothesis in the reference frame of the brain but the evidence which confirms or refutes it must be obtained from experimental results in some sub-frame such as 'Electrical observation of the brain' and the process of using such specialised evidence to confirm a more general hypothesis is, according to the previous argument, characteristic of (and only possible within) a well specified reference frame. Thus we further characterise a well specified reference frame as one in which there exist arguments relating each A_i in *(A)* to some B_j in *(B)* and some.....U_k in *(U)* that are stated explicitly and unambiguously for all sub frames *(A), (B),.....(U)* included in the reference frame.

It is possible to provide a mathematical foundation in terms of which we can be more precise about the situation described intuitively by this *(fig. 1)*. The mathematical foundation centres upon the idea that what we tend to recognise in any system of the kind observed in a sub frame like *(C)* is a stability condition or dynamic equilibrium. Such a condition is to be identified with the appearance of a cyclic group of transformations relating successive results in an observed sequence. We then imagine these cyclic groups embedded in a more general field of transformations in which

the various recognisable features correspond to abstract symmetries preserved invariant by the various groups.

To relate this notion to the work of Beer and Ashby which has already been noted let us examine a specific case, namely, the electrical observation of a brain in the sub frame *(C)*.

In this case a sequence of observations-.

$c_i^* = (c_{u,t}, c_{v,t+1}, \ldots\ldots c_{s,t+\tau})$. is physically represented by a sequence of usually vector quantities which specify electrical states-for example-the vector of the potentials manifested at a number of different sensory electrodes held, in known spatial relationship to one another, on the surface of a brain or assemblage. It is necessary, in order that an observer shall regard these observations as consistent that they satisfy the previously outlined conditions and, in particular, that τ is finite and that he should be able to obtain, from inference upon the observed sequence a transformation M such that an unknown subsequent state, namely, at, t, the state c_{t+1} may be obtained knowing the state at T by using the relationship

$$c_{t+1} = c_t . (M)$$

If, for some finite τ we have-.

$$c_{t+\tau+1} = c_t = c_t . (M)^\tau$$

the sequence is generated by successive transformation by M (that is, the sequence is characterised by a cyclic subgroup of M). The most elementary sequences c_i^* are thus thought of as generated in this manner by corresponding transformations M_i included in sub groups say g_i which characterise the possible dynamic equilibria in this sub frame (and thus the possible corresponding entities C_i in the sub frame). The g_i are regarded as sub groups of some group $G_{(C)}$ such that all g_i and thus all C_i c $G_{(C)}$.

As noted in condition (3) the group G_C and the included transformations will not, in general, be isomorphic with a behaviour of the assemblage. However, the observer may manifest a particular kind of imperfection which allows him to have a homomorphic model of the assemblage, and in this case $G_{(C)}$ is a homomorphic representation of an original group $G_{(C)}$. But, to secure this degree of consistency, the observer must, when selecting those variables which he observes, as components in the vectors C_u. in terms of which he specifies the states of his system, know which of the possibilities are relevant.*

From the fact that observers are able to make useful and apparently consistent observations in many sub frames for example, that the relations of the F_i in *(F)* are deemed clinically useful, it is argued that similar relations between observable entities and the underlying state changes must

* An extension of these ideas to the more useful region of probabilistic observations where (if the model is consistent) the elementary dynamic equilibria are represented by fixed point vectors of a stochastic matrix, is possible, but will not be attempted in this paper.

exist, also, in sub frames other than *(C)* but that they may not be so readily expressed.

An equivalence relationship, $, is thus defined as meaning that, if A_i \$ B_j the entity B_j in *(B)* is causally related to, or determined by the entity A_i in *(A)*. and the existence of a consistent structure (whether readily expressible or not) in all sub frames of a reference frame is taken to imply and be implied by a set of relationships \$ between the entities A_i, B_j,..... included in the reference frame.

Thus, if B_j = An experimental pattern viewed by a subject and if C_k = A particular dynamic equilibrium implied by an observable sequence c_k^* (such as several, coincidently recorded, impulse sequences in some region of the subjects brain).

The relationship B_j \$ C_k would exist if C_k and B_j occurred as a pair under similar circumstances on other occassions-namely-with the same pattern and with the electrodes in the same region of the brain. As noted already in slightly different terms, at the start of the mathematics, this kind of structure is taken to characterise a well specified reference frame.

2.5. *Interaction and participation.*

At this point let us recall the idea, introduced in Section 1 of an external, or unbiassed and scientific observer and a 'Participant' Observer. In any specified reference frame (for the purpose of the demonstration it will be best to keep the sub frame *(C)* in mind) these observers are two extremes, and most observers adopt a position somewhere between them. The External or The Participant approach is favoured according, in the first place, to the objective which an observer seeks to achieve, and secondly, to the character of the assemblage itself.

Thus someone who wishes to dominate an assemblage, to achieve a particular dynamic equilibrium say, will be unable to do this by an external approach unless he has a mass of a priori knowledge about the assemblage to help him. Lacking this he is bound to interact with it and, in doing so as well as in order to do so, he is bound to participate. In other words, if he seeks a relation with respect to the assemblage which maximises his chance of dominating its state change, this relation will necessarily, also, be one which maximises the effect which his activities exert upon its behaviour (and, in the case of certain assemblages like brains, the effect which its activity will exert upon him). Thus, any descriptive model he provides is biassed, since it describes a combined system-he and the assemblage interacting very closely-rather than the assemblage itself. His observations whilst personally useful, will be taken from a viewpoint which changes to maximise the original objective and thus will neither be of much use to other observers or have the calibre of scientific results.

Because of this there is a tendancy to favour the External approach in which interaction is deliberately minimised, to keep the observers relation well defined and repeatable, and to keep the assemblage unmodified by his activity.

But however desirable, this external approach may, as noted above, prove impossible (both because of the type of enquiry which is made and because of the character of the assemblage). The assemblage appears indeterminate in its behaviour to an observer who does not interact with it (that is to say, his observations fail to satisfy the consistency conditions which have been examined).

We have considered two reasons why an assemblage should appear indeterminate, and if the assemblage is brain-like the indeterminacy will be due, largely, to the second of these - namely - lack of relevance criteria. In other words given that B_i may be related to some observable sequence and corresponding entity in (C) there is no means of telling what kind of sequence c^* it would be appropriate to examine. Thus the process of building up a descriptive model, which requires a set of assertions, like $B_i \$ C_j$ proves impossible.

All the same, if the assemblage is brain-like, the observer does not regard it as a 'Chance Machine' which is the limit case encountered when indeterminacy is due to the first cause. If it were a 'Chance Machine' any kind of observation would be fruitless - for example - it is only necessary to examine the bearings of a Roulette Wheel with sufficient accuracy in order to predict its state. But the machine is built so that the accuracy may never, by definition, be achieved, even though, the appropriate kind of observation is completely explicit. On the other hand, if the assemblage is brain-like, we use the fact that people do make sense out of particular kinds of interaction which brains encourage but roulette wheels do not encourage to define the kind of constructed - rather than natural - assemblages which might behave as brains, namely, those assemblages which permit an observer to interact with them and which, if he does interact, make sense but if he does not interact with them appear indeterminate.
The relation of such an interacting observer to an assemblage of this kind is the E.Relation which has been defined in Section 1. It implies that the observer is prepared to infer a similarity between himself and the assemblage in the sense that certain states of the assemblage appear to act, in its workings, in the same way that concepts (and certain other entities) work in his own thinking process. Because he has inferred this similarity the observer may be able to regard entities A_i, B_j,..... and so on as being equivalent even though the argument which asserts why they are equivalent is not available. This special kind of equivalence will be denoted by % so that if a pair of such entities say C_i and J_j are equivalent-.

$$C_i \% J_j$$

To exemplify the relation, imagine the observer is training an animal (a dog or a horse) and that he sets up an E.Relation with the animal - as he would have to - in order to train it. For this purpose we use a sub frame *(U)* including physical stimuli and observations appropriate to animals ile. observations of movement, implying predictable attitudes of the animal. As part of the training we wish to predict the occurrence of a behaviour sequence u^*_{II} which implies U_{II}. given an already observed behaviour sequence u_I which implies some attitude of the animal U_I. The relation of U_I to U_{II} is unknown and unavailable but a trainer will often establish the equivalence -.

$U_I \ \% \ J_I$ and $U_{II} \ \% \ J_{II}$ in which J_I and J_{II} are "concepts", in the functional sense, described. Given U^*_I which leads to U_I the trainer employs, in the same functional sense, an argument like 'If $U_I \ \% \ J_I$ then given J_I I know what I would have done - namely - J_{II}, and this allows him to predict U_{II} and from this u^*_I as an expected pattern of behaviour.

Notably, the enquiries which are made to confirm this hypothesis (in general, whether or not the prediction is successful) have nothing to do with the mechanism inside the animal or with its logical characteristics. Rather, one asks whether the assumption of similarity (which implies using oneself as a kind of dynamic model) maximises the chance of achieving the required objective, and in general makes it possible to interact more effectively with the assemblage.

Under these circumstances it would be fruitless to ask whether the trainer, by continual training, had imposed his way of thinking upon the animals decision process or whether due to continual proximity the man had horse-like or dog-like thoughts in his head. It seems impossible to usefully separate the two components of the interacting system which have become functionally indistinguishable.

2.6. *Second Set of criteria.*

We now come to the second set of conditions which were required, namely those which a 'Thinking' assemblage must satisfy. First of all, in the sub-frame *(C)*, rather than the sub-frame *(U)* any 'Thinking' assemblage must, at least, behave like the animal considered above with respect to a human operator. This much is open to empirical test and the manner of testing will be described in 2.8.

For the moment we require a physical condition which may be used in constructing such an assemblage and which will make it behave as required.

It must, in the first place, be possible for an observer to interact with the assemblage using stimuli or trials and using observations or measurements which are reasonable in the selected sub-frame *(C)*

It is not difficult to ensure that an assemblage is responsive to an observer and modifies its characteristics according to his behaviour. We

may refer to the first of these requirements as Condition (6) and the second as Condition (7) and, if both are satisfied, the assemblage will be able to interact with an observer.

However, assuming this, an observer is disinclined (for the reasons we have examined) to interact with an assemblage and, in general, he will only interact with it if (using the method of an external observer) he is unable to obtain a consistent model. This will occur when the reference frame of his observation is badly specified.

Thus, an admissible assemblage must satisfy a further condition say, Condition (8) which asserts that an assemblage must force the observer to interact with it, in the sense that interaction yields benefits. It must be an assemblage for which the reference frame is badly specified and we are seeking a physical condition on the assemblage which makes a well specified reference frame difficult or impossible to construct.

It may be impossible to derive such a condition in an entirely general form. The issue of what the observer is willing to call 'Entities' and 'Attributes' is involved. On the other hand the position is a little clearer within a particular, sub-frame, say *(C)*.

Thus, thinking of brain like assemblages composed of many similar elements connected together the reference frame of an observation is only well specified if there are definite regions (like the auditory region of the real brain) which relate to the different enquiries (namely enquiries, in *(C)* about the issue of 'Hearing'). If these exist it will be possible for an observer to maintain a known relationship with the assemblage and to regard entities as $ equivalent. The functional specificity need not, of course, be regional. It might equally well be histological, for example a statement like "All pyrammidal cells are motor neurones" specifies the kinds of object with which electrodes should be associated when an enquiry is made about motor activity. But it will avoid confusion to keep the idea of regions principally in mind.

When such definite regions fail to exist the assemblage is necessarily observed in a badly specified reference frame. In this case $ equivalence is unachievable, an external observer is unable to make sense of the behaviour, and interaction is favoured. Any assemblage - in *(C)* - which satisfies Condition (8) is of this kind.

The condition for a constructed assemblage is thus that no region in the assemblage shall be assigned a specific function to serve. The term 'Region' must be taken to include the smallest possible region, namely an element, that is, one of the components, from which the assemblage is built up.

If the Condition (8) was applied strictly each element in the assemblage would be able to serve the same set of functions as any other element - in other words elements would be regarded as completely undifferentiated raw material such that it might form amplifiers, storage devices, or switching

relays, and if it did form one of these functionally distinct entities, such that it might change into another. An assemblage of this kind, which will be defined a Pure E. Assemblage is almost impossible to describe because, in the first place, it could only be observed by an E. Related and interacting observer and secondly, when he did observe it, his interaction, in the absence of any internal constraints, would determine the function of the elements and the state changes of the assemblage. However a 'Pure' E. Assemblage is not so much practically difficult to make (it may, indeed, be approached quite closely) as logically difficult to manipulate. All the same, the idea of a Pure E. Assemblage provides some insight into the character of the E. Relation and those features which are present even in the majority of E. Assemblages (such as real life brains) where Condition (8) is applied with reservations. In other words, any E. Assemblage includes something akin to raw material, of elements, which is unstable until some kind of interaction introduces a pattern.

The pattern, namely a set of constraints which may have a transient existence or may persist, can arise due to the interaction of an observer. In this case the observer characterises the assemblage according to its existing constraints, but equally, he modifies its character according to the constraints imposed upon his own activity by his objective in making the observation.

Alternatively, the pattern of constraints may be built up internally, by interactions between components which are indistinct regions in the E. Assemblage. It will be possible to illustrate the existence of these regions and to show that there is no essential difference between such regions and the apparently well defined regions called observer and assemblage. The overall process of development is the Growth Process which according to the present argument yields 'Concepts' or entities which are functionally identical with 'Concepts'.

2.7. *Existing Constructed Assemblages which satisfy some of the conditions.*

There are a number of already constructed and familiar assemblages which satisfy these conditions with the exception of condition 2 and condition 8. The conditional probability machines developed by Uttley and Andrew *(ref. 11)*, are, for example, in this category if we regard them as associated with a control mechanism and able to interact with an observer who forms part of their environment.

Such a mechanism builds up a model of its environment which is, ideally, homomorphic with a pattern of behaviour in its environment. But, in order to do this, the machine must have a number of constraints imposed upon its structure, so that at least the state changes in the environment which count as relevant events are well specified.

Suppose that the machine is now associated with a control mechanism, and allowed to interact with its environment, including, perhaps, an observer.

The resulting behaviour will not, because of the initial constraint upon the kind of model it must build, satisfy condition 8. Further, suppose that it encounters no state changes which are deemed relevant events, it must, (unless provided with some arbitrary rule to deal with the possibility), stop learning, and thus it fails to satisfy condition 2.

In order to satisfy condition 2, without introducing an arbitrary rest restriction, a different principle of learning must be introduced. MacKay *(ref. 10)* has described a trial-making servomechanism which does satisfy this condition. It is a machine which continually makes trials which are intended to modify its environment and to elicit an event which it is able to recognise. A rule is applied such that, if a trial is made, the probability of its being made upon subsequent occasions is reduced. This rule is rescinded if, and only if, an event is elicited by the trial and this event falls into a rewarded sub-set of events, (such that all included events indicate some desired objective or state of the environment). Such a machine will retain, in its trial probability registers, a model which specifies those states which it assumed and which gave rise to events in the rewarded sub-set.

There is, of course, a sense in which a model of this kind may be regarded as a model of the environment, but it is a quite different model from the homomorphic image already considered. A machine like the trial-making servomechanism is a relatively inefficient control system, which does, however, seek out the best kind of representation for achieving the objective. Further, in the absence of any recognisable event it will continue to make trials and will satisfy condition 2, although these trials will become increasingly autonomous and equiprobable.

George *(ref. 8)* has envisaged a system which, in its trial making, scans a variety of possible relations between itself and its environment.

If the environment failed to yield any relevant and rewardable events this system would make different kinds of trial. The pattern of behaviour noted by Grey Walter *(ref. 12)* when a number of his conditionable tortoises interact in their scanning activity, is possibly due to the fact that the tortoises form such a structure under these circumstances.

None of these mechanisms really satisfy condition 8. The scanning device might do so in the sense of assigning different functions to its sensory and motor elements, but there is the over-riding objection that these functions are preprogrammed in a scanning rule. Thus, we are led to consider an assemblage which is less of a machine and more of a plexus of elements, these elements and their connections being specified to satisfy the conditions for an acceptable assemblage.

2.8. *The Choice of Physical Assemblages.*

To satisfy condition 2 the assemblage may not be energetically closed, since it is required to change its state continually. On the other hand,

to satisfy conditions 3 and 4, it must, in mechanical terms, approach at each instant some dynamic equilibrium. From the requirements of condition 1, the elements must have well defined functions, but from condition 8 no element has a unique function. Thus, we specify the elements, (and sub-sets of elements), as performing a number of, (in the pure case, performing all possible), functions according to parameters which are determined by the remaining elements in the assemblage, and in order to satisfy condition 6 and condition 7, any structure interacting with it.

Choice of a quantity which is employed as a measure of the state of an observeable assemblage and another quantity which is the variable modified by an observer when he interacts with it, determines the physical form of assemblage which satisfies the above conditions. This choice is a matter of convenience and a state specifying measure of resistance, and a state modifying variable of current passed, were selected for the demonstration. Thus, the elements of the assemblage are resistive elements which undergo a lagged decrease in their effective resistance when current is passed through them.

Two kinds of assemblage will be examined and both of them appear in the demonstration. The elements in the first kind of assemblage are thermally sensitive resistances, (the temperature of which is increased by passing a current), which have a negative temperature co-efficient of resistance, and which, (due to their thermal inertia), preserve a decreased value of effective resistance after the current which heats them up has ceased to pass. We envisage an indefinitely large symmetrical plexus of such elements, so connected that a potential difference is maintained across it to satisfy condition 1, and such that the current passing through any element affects all of the other elements, and all of a symmetrically related sub-set of elements in a well determined manner. The overall effect, summed over the sub-set must result in "no change" on the average, (i.e. if some elements are made to pass more current, others are made to pass less current).

The least recognisable assemblage would be a region within this indefinitely large plexus of elements in which the measured variable is conserved, i.e. the average resistance value is constant, (and, since the assemblage is to introduce no special kinds of structure, we also require that the "average value" of effective resistance of each element in this region is constant). To satisfy this and the remaining conditions, we require a limit which may either be provided by conditions on the indefinitely large plexus, or more practically by introducing constant current mechanisms at the boundaries of some observable region in the plexus. It is worth noting that without these mechanisms the current passing through the region will increase indefinitely and that with constant current mechanisms at the boundaries of the region alone, the result will be that some paths in a plexus will pass an increasing current, (for the elements

included in these paths will undergo a decreasing resistance), at the expense of the other possible paths which will thus be starved of current.

To overcome this difficulty we may arrange non-linear current amplifiers, which receive as an input, the effective resistance value between a pair of nodes in the plexus and cause a larger decrease in resistance, (by passing current), in two or more symmetrically related pairs of nodes.

The structure is illustrated for some of the symmetrical plexi which have been exhibited by Corbett *(ref. 7)* in *fig. 2.* The effect of such a feedback loop is summarised in a rule which says -

"If, in a finite assemblage, a change occurs this change may be perpetuated, (by such a feedback loop), in some other part, (or strictly in all other parts), of the assemblage. The ultimate result of this procedure will be obliteration of the original change.

Thus, if we regard the allowed current as a limited amount of currency with which structures, (i.e. patterns of elements with different effective resistances), may be built, there is not sufficient currency to permit building a structure everywhere in the plexus. The amplifiers, (by their feedback connections), initiate its construction at many points, and each

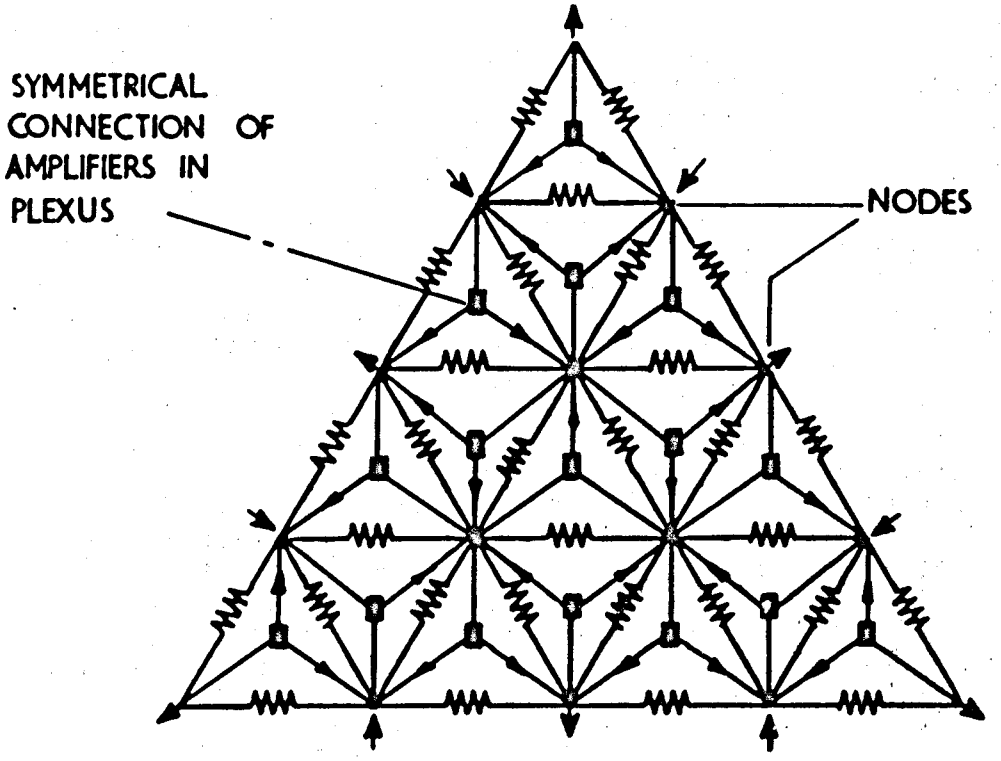


Fig. 2.

of the building schemes must compete for the available currency. If the plexus is connected symmetrically, and if the gain of the amplifiers is sufficient, (which is ensured in the constructional plan), the initial "building scheme" is least likely to have success in this competition, (since the feed-back process involving the amplifiers is cumulative). Thus, other things being equal, which they will be if the assemblage is undisturbed, the feedback process tends to oppose the original sequence of events, (namely increasing path current leading to decreasing effective path resistance), which, on its own, determines that the assemblage would be stable with one path conducting and the others starved of current. Combination of the cumulative feedback process with each original sequence, (i.e. with each possible path), specified a set of dynamic equilibria and there is one such set of dynamic equilibria for each cumulative sequence. The assemblage will approach each of these dynamic equilibria, namely each member of each set, with a probability of approaching any one, (at some arbitrarily selected instant), determined by the symmetrics of the plexus connections.

Such a system is a special case of the multistable and ultrastable systems which have been defined and discussed by Ashby *(ref. 1)*. The analogy appears if we regard each of the possible "paths" as specifying a set of "critical" points in the critical surface of Ashby's phase space, (the set of dynamic equilibria are specified by the set of parameter changes which keep the state representing point of the ultra-stable system in the admissible region of its phase space).

A system of this kind is also able to learn in the sense that, if it is disturbed the behaviour which has been described is modified, to include so far as possible, the disturbing effect. In general, the system becomes increasingly sensitive to any disturbance. Thus, new dynamic equilibria become possible, and since each of these represents a recognisable pattern of behaviour, the set of possible behaviours, (which an observer might discern and which are characterised with sequences like -

$$c^*(i) = (c_{u,t}, c_{v,t+1}, \ldots\ldots c_{s,t+\tau})$$ for the

dynamic equilibrium C_i) is enlarged.

The elements in the assemblage, with the possible exception of the current amplifiers, may not, after an interval of activity, be ascribed a particular function. The function of each element and each region of elements is continually and unpredictably changing, so that any assertion made about its function would be ambiguous.

A number of the possible functions which an element can serve will be indicated. In the first place any element has a thermal inertia which makes it a possible storage device. If current is passed its subsequent state is modified and although, if the current were entirely discontinued, the element would return to its previous state after an interval, the position

in practice is more involved because some current is being passed at each instant. Thus, the result of a current increase is to modify the current passing characteristics of some region in the plexus in a manner which depends upon the magnitude of the increase in current and upon the pattern in which each element is included.

Suppose a plexus in a plane, and a node in the plexus which receives only one connection from higher, (more positive), nodes, whilst sending, (by way of intervening elements), a number of connections to lower, and more negative nodes. In this case, let any one of these lower paths assume a low effective resistance, this will lead to a decrease in the chance of all of the other paths becoming low in effective resistance, since there will be a reduction in the potential across the entire set, as in *fig. 3*. Thus, one of the lower paths will tend to be current passing and the others will be high resistance paths. In this sense the elements act as non linear devices which determine binary events in a set of continuously changing variables.

In the same sense the elements may perform a binary transformation, that is to say, they may act as switching elements. Thus, in *fig. 3* the one lower element with a low resistance is the 'made' contact of a 'switch', the other positions of which would be selected by some other element being low resistance. The one connection from upper elements in the plexus may, in this manner, be regarded as the made contact of a higher switch. As indicated in *fig. 4* the switch may also be one to many or many to one. In *fig. 4* the current limitation is assumed such that more than one of the lower elements is possibly of low effective resistance.

It is possible to devise a kind of amplifying region in the plexus, in the sense that a small change in effective resistance in one element will yield a large change in the current passing through some other set of elements. This would remove the necessity for separate current amplifiers, but, in practice, the characteristic is difficult to achieve. A more sensible method of unifying the function of elements would be to redefine each element as including a local energy source, and to do away with the potential difference across the observeable assemblage. Plexi of a similar kind have been made and shown to have self organising and information organising characteristics *(ref. 7)*.

The really arbitrary feature of this plexus does not, however, reside in the character of its elements, but in the fact that a pure E. assemblage should be a completely connected plexus. This ideal is almost impossible to reach, but it is possible to see that if the various degrees of freedom used up in specifying the symmetries of a real life plexus were available, the elements would act like raw material from which any assemblage might be built.

Rather than consider approximations to this ideal, it seemed more profitable to see if the required characteristics were shown by a different mechanism. At any rate, I made a guess about this different kind of machine.

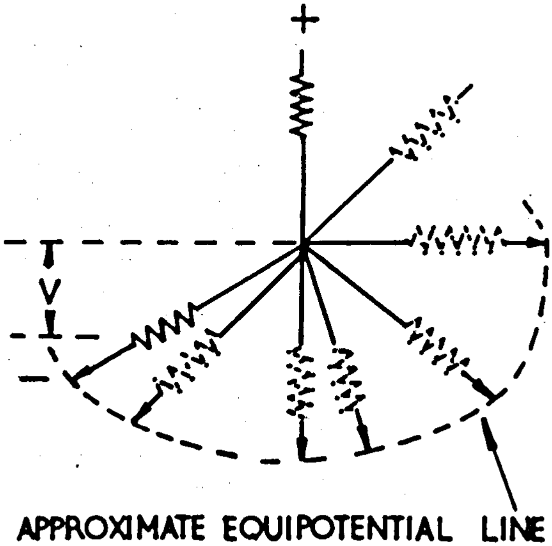


Fig.3

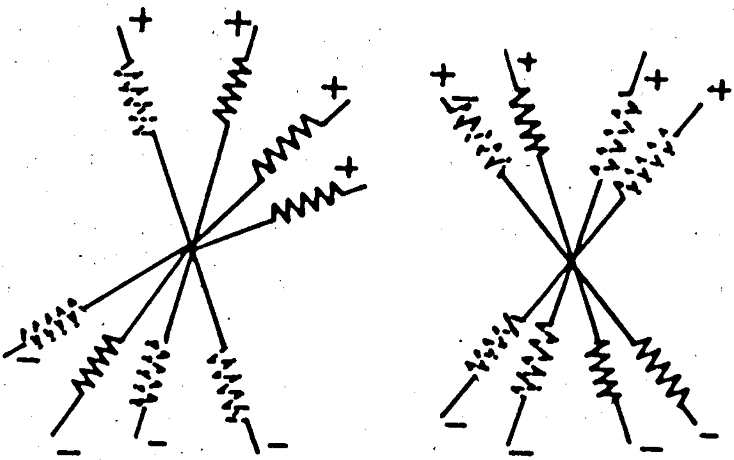


Fig.4

The guess was that the effect of adding further initial degrees of freedom to a plexus of parametrically variable elements is achieved, biologically, in a less clumsy manner, namely by providing raw material of unstructured but structureable elements, the surroundings of an embryo when it starts to grow, being a case in point. The surroundings of an embryo are disorganised elements, in the sense that within wide limits, its development is genetically determined, (and relatively unaffected by the parameters of its surroundings), and I regard these surrounding elements as an assemblage. The limited currency condition is a requirement, determined energetically, which limits the amount of organising activity which may take place in a unit interval. As the surroundings are organised, in other words, as elements which were initally raw material in the assemblage have some function determined, we say that the embryo grows, (and, looking at it at this stage in its development, we also say that it is now considerably affected by its surroundings which are, however, largely determined by the embryo itself). For the present purpose I regard the development of the embryo as equivalent to the growth of a concept in the assemblage, in the sense that I can assign to the continually changing entity called "embryo", at each instant, certain functional characteristics, (the uses of a concept). In this analogy either "the observer" or a "specialised region" which interacts with an assemblage is equivalent to the genetically determined structure which is the ancestor of the "embryo".

It is possible to make a mechanical analogue of such a process and this will be called the second kind of assemblage. Descriptively it has many advantages. Whilst the entity which represents, (and acts as) a concept, in the first kind of assemblage, is an organised region which is continually changing and may only be detected by using a rather involved electrical method, the entity which represents and acts as a concept in the second kind of assemblage is a solid object which, (although it is being continually rebuilt and reorganised) may be examined or photographed.

In an assemblage of the second kind the plexus is replaced by a conducting plane, with electrodes which correspond to the nodes in the plexus, and a conducting material which is a solution of metallic ions, (which are the elements). Whilst in solution, the elements have no function assigned to them. To have a function, they must come out of solution, and form part of a metallic thread which has, (compared with the resistance of the solution), a very low resistance. Such threads tend to develop along lines of maximum current passing between the electrodes, and these lines are determined by a field distribution in the conducting plane, comparable to the current path distribution in the previously described "plexus". Clearly these threads will tend to develop from the nodes where current passes into and out of the solution, and at which there are constant-current mechanisms which allow only so much current to pass per unit interval.

The initial behaviour of the system is similar to the previously described plexus, since the thread which develops between a pair of nodes across which there is a potential difference, tends to reduce the resistance between these nodes. However, the field distribution in the plane is not only determined by the potentials at the electrodes, (i.e. at the nodes), but also by the disposition of these electrodes. The threads which develop from the electrodes act, in this case, to extend the electrodes and thus to modify their disposition, and the process leads to a continual change. Further, the existence of a thread depends upon sufficient current passing through it, since there is a tendency for it to dissolve into the surrounding solution. Thus we may regard some threads as more stable than others, according both to their own form and the form of the surrounding threads, and if a thread tends to dissolve, it is not usually the case that its disappearance recapitulates its building.

The pattern of threads which exists at any instant is thus a structure in dynamic equilibrium. In the undisturbed assemblage the system will pass through a variety of dynamic equilibria which are stable under the current limitations.

The two kinds of assemblage are thus comparable, and if the first assemblage were made very large and completely connected they would tend to isomorphism. However, for all practical purposes, we may usefully distinguish the first assemblage as "learning network", ("the network problem" having been solved initially by the designer, who introduces certain symmetries in the plexus), and the second assemblage as a system in which the "network problem", (again of a "learning network") is solved as a part of the learning process. The distinction is not very sharp, but on common sense grounds I should call the second, but not the first assemblage, "self building", and say that it illustrates a "growth process".

2.9. *Experimental Hypotheses*

We are now in a position to examine a real life assemblage and to confirm or refute a number of experimental hypotheses. These seem to fall under two well defined headings.

The first set of hypotheses refer to enquiries about whether or not the assemblage, (which is available for demonstration), does, in fact, satisfy the conditions we have discussed and in particular does it exhibit the characteristics of a developing embryo, (within the terms of my analogy). If so, a second enquiry becomes reasonable, namely, is this biological analogy appropriate for representing the growth of a concept.

The hypotheses which refer to the second enquiry concern whether or not an observer may be E. related to the assemblage, and whether or not adopting an E. relationship yields any advantage in the sense of achieving a number of reasonable objectives, (dynamic equilibria in the assemblage).

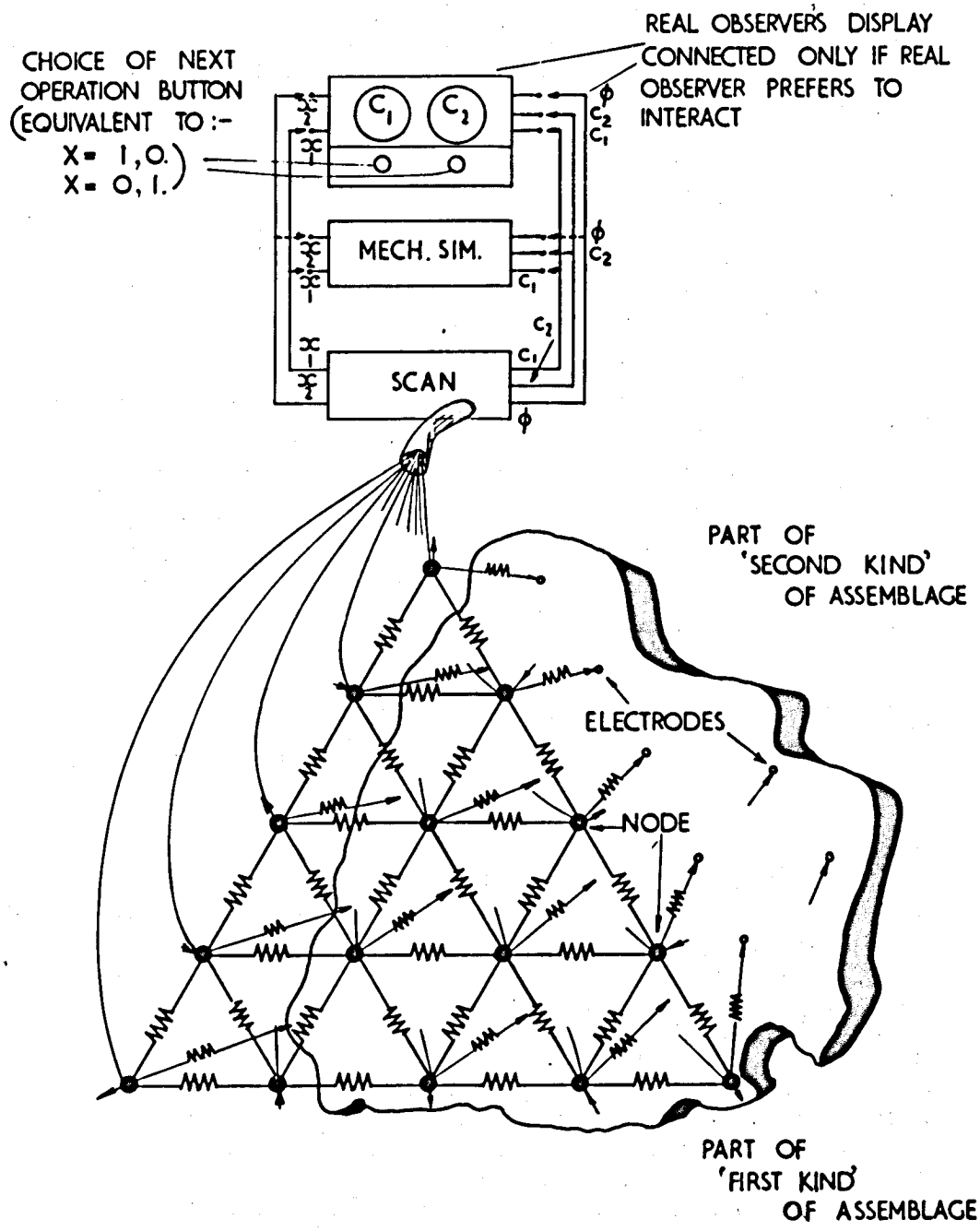

CHOICE OF NEXT OPERATION BUTTON (EQUIVALENT TO :- X = 1, 0. X = 0, 1.)
REAL OBSERVER'S DISPLAY CONNECTED ONLY IF REAL OBSERVER PREFERS TO INTERACT
C1
C2
MECH. SIM.
SCAN
PART OF 'SECOND KIND' OF ASSEMBLAGE
ELECTRODES
NODE
PART OF 'FIRST KIND' OF ASSEMBLAGE

Fig.5

The experiments, (which will be realised in practice) are described in 3.2., and involve the idea of a finite sequence of observations made by a real observer, i.e. any person who wishes. This sequence may be selected by the observer who must, however, choose between the alternatives of (i) making many different observations in a manner which does not appreciably affect the state of the assemblage and then providing some rule by which the parameters of the assemblage are modified to achieve the objective, or (ii) on the other hand, making fewer observations in an interactive manner, which does affect the assemblage. In this case the objective must be achieving as part of the interactive process.

The latter observer may be, whilst the former may not be, E. related to the assemblage. It is possible to demonstrate that the latter course of action leads to success, though the former does not achieve the objective in a finite interval. Further, it will be possible to perform an experiment which overcomes, to some extent, the comment that given this, and given a real observer, the issue of E. relations still depends upon personal evaluation.

2.10

The demonstration assemblage is of the second kind which has been discussed. The experiments examined in 3.1. are performed upon this part of the demonstration. In order to associate this assemblage with an observation sequence, an assemblage of the first kind, (namely a symmetrical plexus of elements), has been introduced and has exactly the same status, (in the demonstration), as a specialised region in a real brain.

It is, in other words, a region in which there is a certain amount of functional specialisation. An interacting observer determines the state of this region knowing that it means something to take current from, or to make an observation at, a specified node. But, as we shall see later - in the experiments of 3.2., his knowledge does not amount to certainty.

SECTION 3

3.1. *Experiments to demonstrate the physical characteristics of an assemblage.*

I. The assemblage must show a self building characteristic. If we regard the metallic thread as a decision-making device, in the sense that its presence gives rise to a current flow which selects one alternative, and its modification gives rise to a different pattern of current flow which selects another alternative, we require that if a problem is found insoluble using a specified thread distribution, the assemblage will tend to build itself into a new decision making device, able to reach a solution to the problem.

The experiment which is intended to show this characteristic is illustrated in *fig. 6*, where points 'X' and 'Y' are nodes, more positive than node 'S', so that if the intervening plane is an assemblage of the second kind, a metallic thread will tend to develop from 'S' to either 'X' or 'Y'. In the simplest case, which is obtained by making 'X' assume a high positive

potential, we determine an initial current path towards 'X' and thus ensure development of a thread along this path. Let this occur in an interval $t_2 - t_1$, and at the instant t_2 we change the parameters of the system so that 'X' and 'Y' have, with respect to the thread which is now terminating at the point 'P', an equal but relatively positive potential, so that the further path of the thread is ambiguous. Development of the thread in an interval $t_3 - t_2$ in which this new set of parameters apply, depends upon the form assumed by the thread, the current which it is able to pass, (due to the "currency" limitations), and the surrounding threads, (which determine details of the field in parts of the thread other than its terminal point 'P', and which may, for example, make the thread assume a positive rather than a negative polarity with respect to 'X' and 'Y' within this interval). We shall consider, for the moment, only four of the possible alternatives. -

(i) The thread takes an intermediate path, or
(ii) It approaches 'X', or
(iii) It approaches 'Y', or
(iv) It bifurcates.

Of these, the possibilities, (i), (ii) and (iii) may occur if little current is available, and might occur within any computing machine presented with this decision. We are interested, however, in (iv) which is most likely if the current is available and which is shown in *fig.6*.

If, at the instant t_3 the parameters are returned to the values assumed in the interval $t_2 - t_1$ the behaviour of the assemblage will be quite different. Since it is the behaviour of a double thread, (i.e. a bifurcated) assemblage which determines an entirely different field distribution, the behaviour in the interval $t_4 - t_3$ would not be predictable from observations made in the interval $t_2 - t_1$, when similar parameter values applied. Thus, an observer would say that the assemblage learned and modified its behaviour, or looking inside the system, that it built up a structure adapted to dealing with an otherwise insoluble ambiguity, in its surroundings, (i.e. the ambiguous parameter values, in the interval $t_3 - t_2$).

II. The assemblage must always exhibit this kind of behaviour unless its surroundings are entirely determined. To show this we determine a unique current path to the nearest practical approximation, and observe that the thread develops by a process of abortive trial, namely, it bifurcates continually, but most of the bifurcations are abortive, and the dominant bifurcation is predictable.

III. When we say that a system adapts to deal with, (or to assume a dynamic equilibrium with respect to), its surroundings we imply a certain foresight on the part of the system, (thus we imply, at least, a supposition that these surroundings will persist until the modifications are completed). An admissible assemblage should have a degree of foresight which increases as it develops. Although this cannot appear, directly, a similar characteristic may be shown -

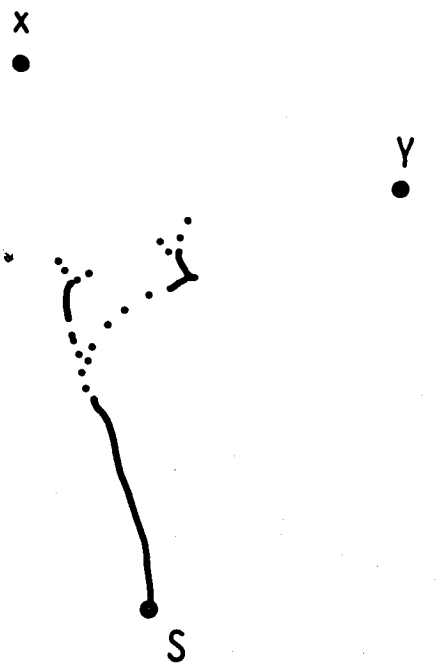


Fig.6

Fig.7

(i) Development of a structure of threads is a competitive process, by definition, and by examining the system.

(ii) In this case, if there are two structures of threads, say 'U' and 'V' in *fig.7* there may be a stage in their development at which one of these will dissolve in favour of the other, perhaps, in the manner indicated. The one which does not dissolve is said to dominate, or to be more stable than, the other. Suppose that 'U' is a structure which has been built up in one part of the assemblage and has been in equilibrium with a very variable set of parameter values, whilst 'V' has developed independently, (that is to say, in relative independence of the other, though it cannot have been completely independent by definition). Suppose, further, that the structure 'V' has developed in fairly invariant conditions. At some instant 'U', and 'V' will be in competition, due to the limitations, (both the current limitations and the spatial limitations of the plane), which are imposed by an assemblage, and that one or the other must be dissolved. Then the probability that 'V' will dissolve is very much greater than the probability that 'U' will be dissolved.

IV. If 'U' and 'V' had been structures developed in comparable surroundings, it is possible that they would have combined and that, on examining the system, we should have agreed -

(i) That the development of 'U' was assisted by the presence of the structure 'V' and vice versa,

(ii) That 'U' and 'V' were no longer distinct, but should be regarded as a combined system.

In terms of the theory of games, this process is "cooperation", and the combination is a "coalition". Further, in view of III we see that stable coalitions will only occur between, and will, thus, only accelerate the development of, comparable stable and dominant structures. In III and IV we have a selective principle which says that a self building assemblage tends to develop along a dominant pattern, but if several structures are dominant, a coalition is more likely to be stable.

Finally, some comment is needed regarding the sense in which an assemblage of this kind has a memory. In what sense, for example, is a pattern retained invariant, and would it be possible to say of such as assemblage, as it would be of an organic system, that it preserved an organisation even though the elements which mediated the organisation were continually changing.

V. The first part of the experiment which may or may not be convincing is to modify the assemblage by pouring away some of the solution, and showing that this does not greatly modify its behaviour. One might argue that there is no reason why it should, yet whilst this is the case, there is every reason why such a drastic modification of most decision making or learning assemblages would be important.

The second part of the experiment is to show regeneration of a thread. The experiment is indicated in *fig.8*, where the thread 'J' is assumed to have developed under conditions say, 'L', which have just been modified to other conditions, say the conditions 'M', such that, under the conditions 'M' an entirely different thread would have developed.

At this point the thread 'J' is cut and a portion is removed. The thread 'J' will now be regenerated by a process which involves dissolving away at the edge 'g', and deposition of elements dissolved into solution at the terminal point, 'm' of 'J'. The regenerated 'J' never catches up with the old thread, but, for quite marked differences, between the field distribution determined by 'm' and determined by 'L', its precise replica is produced, after an interval needed for regeneration to occur. In other words, the existence of the structure 'J' has constrained the assemblage so that even if the actual structure is modified, and the field surrounding it is modified, the pattern will be retained.

There is, in addition, a fairly good analogy between the various stages of "determination" in the biological system, and the various stages of modification and partial regeneration which occur, if the regenerating

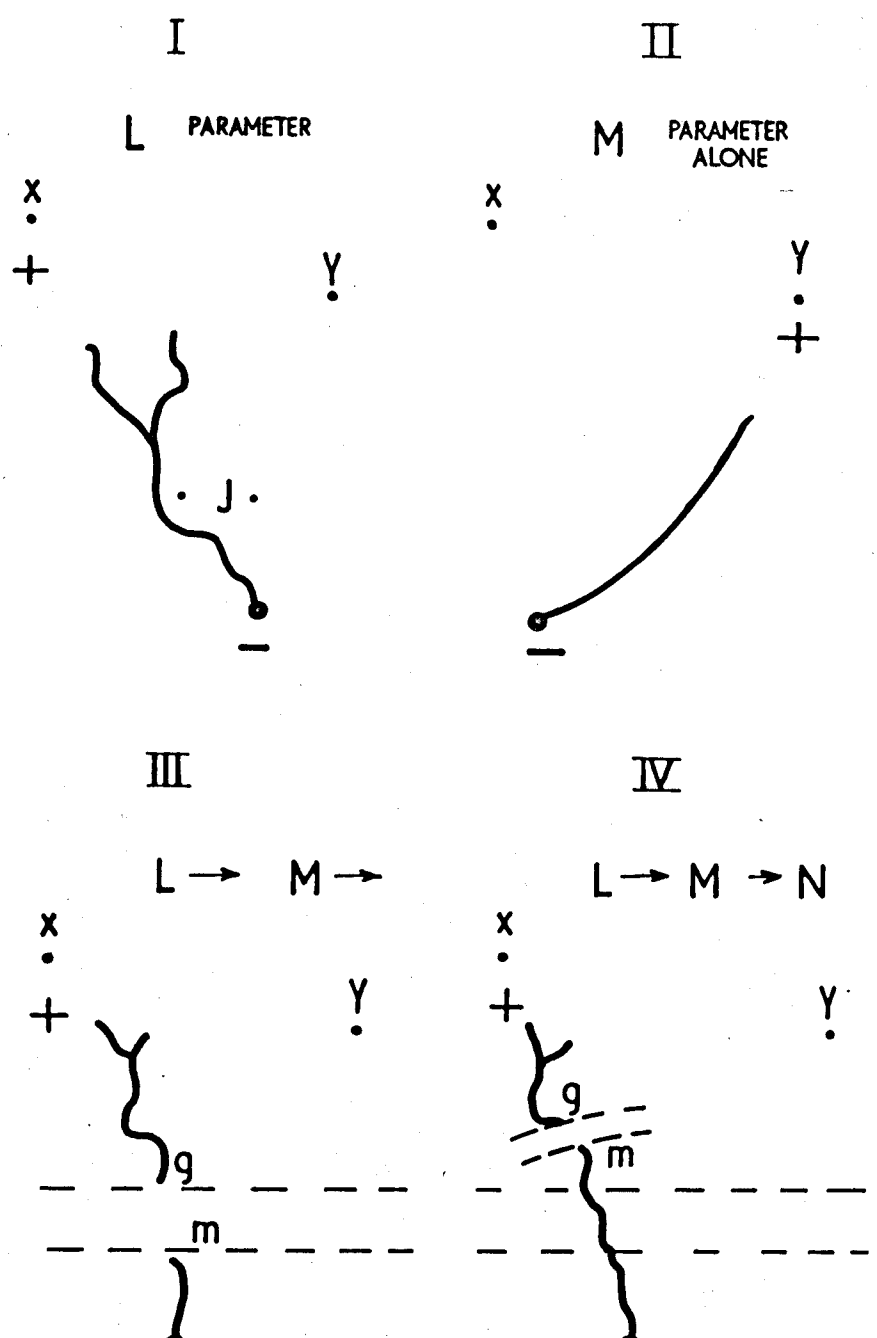

I
L PARAMETER
X
Y
J
II
M PARAMETER ALONE
X
Y
III
L → M →
X
Y
q
m
IV
L → M → N
X
Y
q
m

Fig.8

thread 'J' is subjected to an increasingly incompatible field distribution. The evidence taken as a whole, supports the view that regeneration and non specific forms of memory occur in an assemblage of this kind.

These characteristics may be described in terms of a sequence of constraints which are necessarily imposed upon an assemblage. It is clear that any constraint will initiate activity which tends to remove the constraint and to bring the assemblage into dynamic equilibrium with its surroundings. However, the self building characteristic implies that the modifications which occur necessarily produce further constraints and these function in a similar manner, as ancestors determining the next constraints.

Although, it is the case that constraints which determine a stable pattern tend to persist, and are recapitulated, the assertion that one pattern is more stable than another, may only be interpreted with reference to a particular environment in which the stability is achieved. Since the environment becomes increasingly determined by the constraints which are developed the interpretation is thus being continually modified.

3.2. *The Experiments which show the advantage of an E. relationship.*

The second set of experiments have already been introduced by the discussion in 2.9, and are performed upon the demonstration as a whole, which is shown in *fig. 5.*

An observer in *(C)* is required to achieve one or more objectives, (namely dynamic equilibria) denoted C_j and implied by the existence of observeable sequences $c_j^* = [c_{u,t}, c_{v,t+1}, \ldots\ldots c_{s,t+\tau}]$.

The vectors c_u have components which refer to different meters in the observer's display, (in the present machine there are four such components), and these meters indicate the effective resistances of the elements interposed between specified pairs of nodes.

In order to achieve the objective, an observer may either decide to adopt a non interactive or and interactive approach. If he prefers a non interactive approach as he would if he were an "external observer", he is allowed to select an observation sequence of n alternative sample loci each of which defines a different vector c_u. These sample loci are associated automatically with the meters via a scanning mechanism which moves on at each observation, *(fig. 9)*. The next observation, at each stage, is determined partly by the observation sequence selected initially, and partly by the observer, who is allowed to select one amongst a finite number P of alternative next observations, by pressing one of P alternative buttons.

Thus, the observer is able to modify his observations according to what he has already observed, within the limits of which he is aware at the outset. In terms of the theory of games, the observer is a player, his set of pure strategies the set of tours across sample loci, and the pure strategy he adopts the tour he determines by the procedure described above.

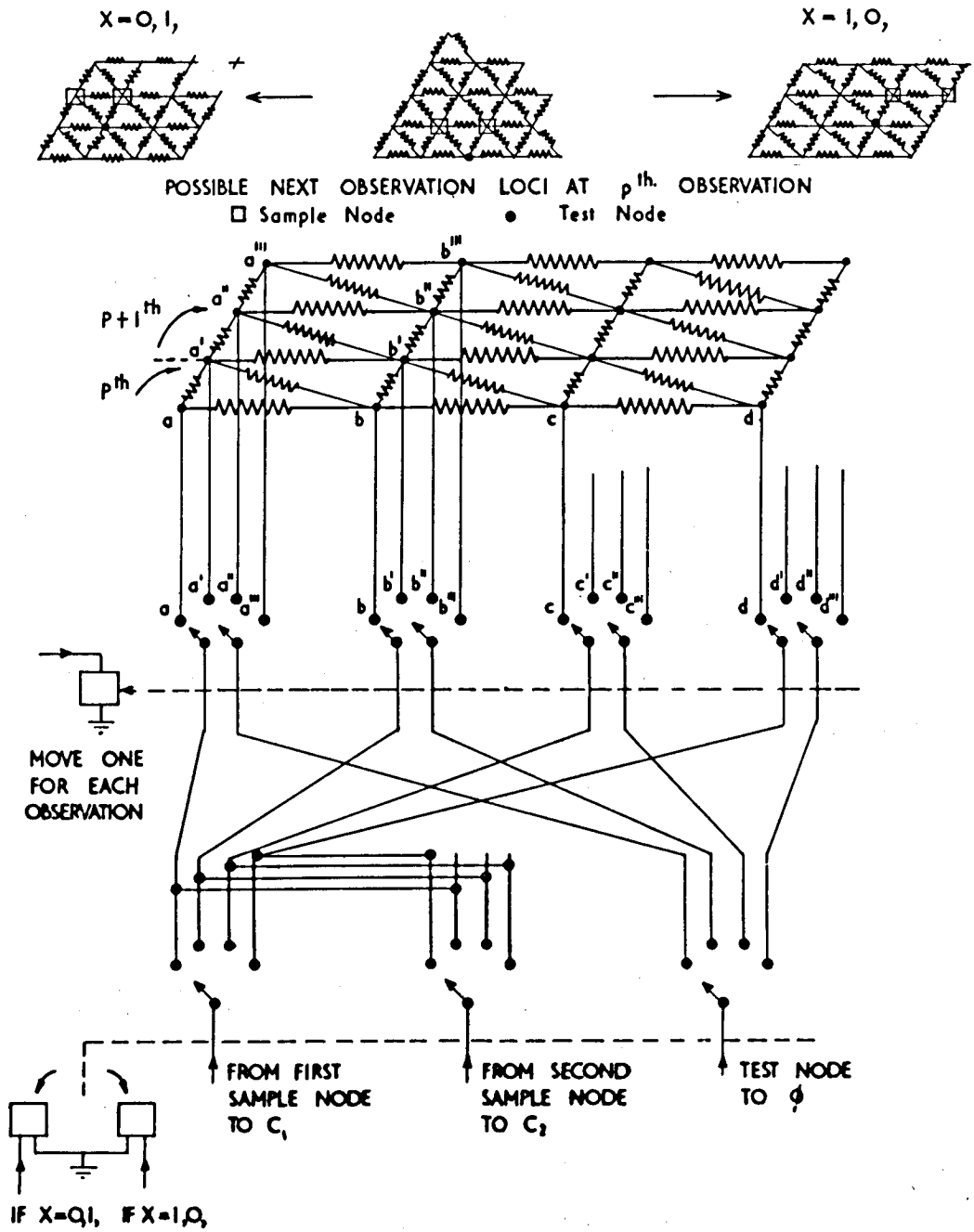

X = 0, 1,
X = 1, 0,
POSSIBLE NEXT OBSERVATION LOCI AT p th. OBSERVATION
□ Sample Node
● Test Node
P+1 th
p th
MOVE ONE FOR EACH OBSERVATION
FROM FIRST SAMPLE NODE TO C1
FROM SECOND SAMPLE NODE TO C2
TEST NODE TO ϕ
IF X=0,1, IF X=1,0,

Fig.9

If he adopts the interactive or 'participant' approach he has, with two exceptions, the same facilities. The exceptions are that his set of pure strategies includes only m sample loci with, $n \gg m$ and that, whenever an observation is made, current is taken, via a Test Node, from, the assemblage which thus modifies its state. This current which is taken may be regarded as the price which is paid for observing an assemblage.

If the assemblage behaves like most of the physical assemblages which are examined, the observer with m very much less than n would be at a disadvantage. He would have less chance of specifying a model adequate to determine a rule for achieving the objective. Again, he would always have to pay the price of modifying the assemblage and still further, reducing his chance of finding a real consistency in his observable sequence C_i^*. However, it may be shown that observers who prefer to interact succeed in achieving quite generally specified objectives C_i and report that they do this by using the ability to interact with such an assemblage in much the same way that an animal trainer uses his ability to interact with an animal.

In particular it is impossible, without further enquiry, to comment upon the relation between an observer in these experiments and a subject in the experiments performed by Bruner, Goodnow and Austin, *(ref. 5)* which were examined at the outset of the discussion. Some comment on this score is necessary. For example, it must be possible to say how an objective is related to one of Bruner, Goodnow and Austin's problems, and how finding an objective is related to finding the solution to such a problem. Given this, a calculus for describing and using these systems as thinking mechanisms is at least conceivable. Without it, the state changes of the assemblage show a close relationship to concept formation, but serve only as an analogy. Again, given this, we are in a position to set real problems and find, experimentally, if they are solved, but without it, a "solution" does not have the precise meaning of "solution" in the game of thinking.

SECTION 4

4.1. *Mechanical Simulation of the Real Observer*

As a first step in this direction I shall assume a particular interpretation of the game which these authors describe. In this interpretation, the game, (played by a real subject), is a competition of part of a man, (namely a part of the subject's brain which is aware of and trying to solve a problem), with the remainder of the man. Thus, the authors examine for each problem various logical strategies which might be adopted for solving it. Some of these, for example, require a good deal of memory capacity, some involve taking risks, and some are safe but slow. I am assuming that the problem solving part of the brain tends to adopt one or another of these strategies according to the facilities available, i.e. according to a bargain it is able to make with the remaining part of the brain. Thus, if it is possible to have memory capacity available, and if

the strategy which taxes the memory is efficient, this strategy will be selected. On the other hand, it would not be selected, however efficient, if memory were not available.

My main justification for adopting this view is the fact that it leads to a coherent picture in terms of the present argument. The interacting observer is clearly a player in the position of the part of the man which is aware of and trying to solve a problem. The assemblage is the remainder of the brain which may, (according to the play of the game), be used to serve various functions in solving a problem, (that is to say, in achieving an objective which implies some state of the combined system).

Assuming, for the moment, that these relations are justified we must examine the decision function which is used by an interacting observer. In order to do this we shall replace the real observer by a mechanism of the kind described by MacKay *(ref. 10)* as a trial-making servomechanism. Such a device will be able to construct, in the manner which we discussed previously, a decision function which is appropriate for achieving (i) maximum interaction with the assemblage, and (ii) the specified objective, providing that it is possible to define -

(I) A function θ which increases with increasing interaction, and

(II) A function η_1 which increases as an objective C_i implied by c_i^* is approached.

The function θ may be specified quite generally for the assemblage concerned, since (in order to modify the state of an assemblage), an interacting observer must be able to take current from the assemblage. It is also intuitively clear, (and it may be shown at least in particular systems), that this depends, in the case of an observer with a finite set of test nodes at which current may be taken, upon his previous behaviour. If, for example, he has adopted a strategy which has led to a set of low resistance paths which terminate at the sub-set of nodes which are visited, then he will be able, by taking current at these nodes, to exert a large effect upon the state of the assemblage. We thus, define the current taken as the "price" of an observation, as ϕ, and specify a constant current servomechanism, as shown in *fig. 10* which takes this amount of current from each of the test nodes visited. We then define θ as inversely proportional to the feedback needed in this servomechanism in order to take a current ϕ from the assemblage.

The function η_i is, however, restrictive, since it may only be defined for a few of the possible dynamic equilibria C_i, and this difficulty will be dealt with in a moment.

An appropriate kind of trial-making servomechanism is shown in *fig. 10*, and involves a few developments of the original device. It has been assumed in *fig. 8* that the vectors c_u have two components c_1 and c_2, and that a binary vector $Y = y_1, y_2$ is elaborated by means of a resolver circuit. A resolver circuit is the mechanism which embodies the rule, employed by a

trial-making servomechanism which we have discussed, namely the rule which asserts that if the input event $Y = 1,0$ occurs, its subsequent occurrence is made less likely, (and similarly for $Y = 0,1$). We restrict the set of input events to $(Y = 0,0)$, $(Y = 1,0)$, and $(Y = 0,1)$, by the condition that $y_1 + y_2 = 1$, and we make $Y = 0,0$ assume a probability of occurrence which is nearly 0 by defining a process which tends always to make both $Y = 1,0$ and $Y = 0,1$ occur, (this is achieved, in practice, by the mechanism involving the condensers). Since $Y = 1,1$ is prohibited one event inhibits the other. But, supposing one input event occurs, its probability of occurring upon subsequent occasions is reduced and thus the probability of the other occurring is increased. The input vector c_u is now applied to the resolver, as shown in *fig. 10*, so that it biases the chance of one or the other input event occurring, for without this bias each input event would occur equiprobably.

The scanning mechanism, shown in *fig. 9*, moves an observer's test node and sample nodes across the sub-set of nodes included in his set of pure strategies. The set of four storage condensers in the matrix $\xi\ (p)$ are specified differently for each position of the scanning mechanism. Thus, if there are α positions there will be $4 \cdot (\alpha)$ condensers in the matrix ξ corresponding to sub-sets of entries $\xi_{(p)}$.

The potentials associated with these storage condensers are the entries in a decision function matrix which is built up as a result of the interaction. Thus, at the p-th position of the scanning mechanism, some of the storage condensers are charged via a constant resistance from a potential of value $\theta_{(p)} \cdot [\eta_{i,(p)}]$ in which $\theta_{(p)}$ is the value of θ at p, $\eta_{i,(p)}$ is the value of η_i at (p) and in which $1 > \theta > 0$ and $1 > \eta_i > 0$.

The particular storage condenser in $\xi_{(p)}$ which is charged is in the column relating to the input event which occurs on the occasion concerned, and in the row which, (as we shall shownin a moment), corresponds to the output event or decision to which this input event gives rise. Thus, the entry in this position in $\xi_{(p)}$ is the average reward achieved, (by the trial making servomechanism assuming this particular input and output state), and the distribution of these entries is thus a decision function.

We assume in *fig. 10* that a decision is made between two alternative next observations one of which is selected if a binary vector $X \equiv 1,0$ and one if $X = 0,1$.

The vectors X occur as the output of a resolver circuit, shown as an "output resolver" in *fig. 10*, and comparable to the "input resolver" which determines the values of Y. The resolver would produce, without any bias, equiprobable output events, and thus decisions. It is biased, however, at the p-th position of the scanning mechanism by the quantity $Y_{(p)}\ (\xi_{(p)})$. Thus, the decision function determines the decision (for specified c_u and Y), and the decision made gives rise to a selection, (for specified c_u, Y, and X), of the entry in $\xi_{(p)}$ which is modified on this occasion.

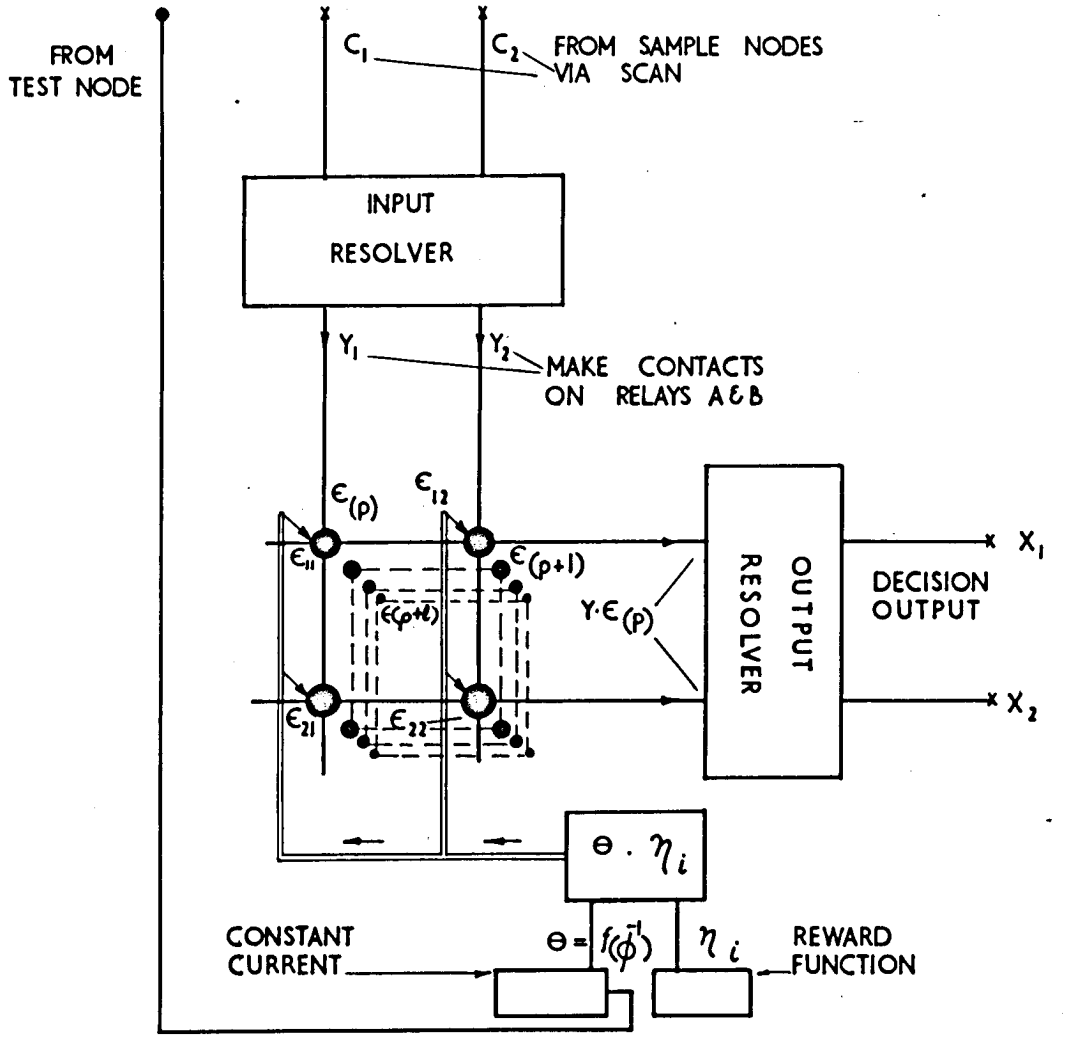

FROM TEST NODE
C1
C2
FROM SAMPLE NODES VIA SCAN
INPUT RESOLVER
Y1
Y2
MAKE CONTACTS ON RELAYS A & B
ε(p)
ε12
ε11
ε(p+1)
ε(p+1)
Y·ε(p)
ε21
ε22
OUTPUT RESOLVER
X1
DECISION OUTPUT
X2
θ · ηi
CONSTANT CURRENT
θ = f(φ⁻¹)
ηi
REWARD FUNCTION

Fig. 10

Although the potentials which form the entries of the decision function matrix ξ are continually changing, their values are, on occasion, statistically stationary, so that ξ may be represented as a stochastic matrix. At first sight we might expect to use this matrix in order to predict the strategic behaviour of the machine, (although, the reverse is assumed in a slightly different field by Beer) *(ref. 4)*. It is interesting to note that it is rarely if ever possible to make such predictions.

On the other hand, the machine appears to act in a well ordered manner, and it will achieve its objective. In retrospect, the stages of the process seem rational. The point is that the moves, (in achieving the objective), exhibit a rationality which is characteristic of a much larger machine, and which would be impossible for a mechanism of the properties described.

The difficulty is, of course, largely descriptive, but it is all the same of great importance. This machine is only able to achieve its objective if it imposes an organisation upon the assemblage such that the requirements of maximising θ, whilst using a (not usually unique) strategy, implies maximising η_i. The organisation which it imposes upon the assemblage, (which, again, is not unique), may equally well be regarded as a functional part of the machine - or, of course, vice versa, with the assemblage acting as a dominant player.

Indeed, as soon as the machine makes its first move, it becomes impossible to say that the particular set of elements defined as a trial-making servomechanism is, in fact, the *functional* machine. Rather the *functional* machine is something which extends, if successful, into the assemblage. A memory, for example, may equally well reside in a modification of elements originally deemed pieces of the assemblage, as it may in a specified condenser deemed a piece of the trial-making servomechanism.

What we have gained by introducing a mechanism which is able to interact with the assemblage, but which has an insufficient number of parts with a well defined function, (so that, in practice, it has to use other badly specified parts in the assemblage), is the advantage we gain from describing a piece of brain with reference to a highly specialised region, or, (in the experimental context), of employing the specialised conventions which make people accept certain objects as representing concepts.

4.2. *The Meaning of a Solution to the Competition of two Mechanically Simulated Observers*

There is still the difficulty that a η_i may only be defined for a few of the possible C_i and without a η_i the trial making servomechanism will not achieve an objective. It seems there is a possible way out of the difficulty which involves a fairly reasonable view of a solution to the game of thinking, and which will be submitted in conclusion.

If a pair α and β of similar trial-making servomechanisms are made to interact with an assemblage, both of them trying to interact maximally, but neither being restricted to reach a particular objective, it is possible to recognise increasing regions of organisation in the assemblage, which have α or β as ancestors. Eventually a metastable state is achieved and this will be defined as a solution.

Up to this point it would have been possible (and this may be demonstrated) to modify the availability of current in the assemblage, and to obtain two consistent kinds of response, one response for α, and one for β, (indeed, this is usually the only way in which α and β may be distinguished). After this point, although a change may occur, there is no consistent kind of response and I thus assume there is no difference in the preference orderings of α, and β. But the only distinction between α and β was of this kind. Thus, I assume that there is now one large coalition, or one combined system and in any case so far as the dealings I am allowed to have with the assemblage are concerned, the distinguishing of α and β is no longer useful.

A solution of this kind is a compromise effected between players which may be arbitrarily defined regions in the assemblage, (the introduction of the trial-making servomechanisms makes the process easier to describe and easier to demonstrate, but the argument applies to any region specified). The form of these regions which behave as players is determined by my own reference frame in terms of which I talk about problem solution. A sub-frame of this reference frame characterises the solution and a solution is said to occur when using the mode of interaction allowed in the sub-frame, I am able to make no useful distinction of regions in the assemblage.

Finally there is a way in which I can form a solution, or arrive at a compromise, or deal with a problem which is stated in my own terms. Namely, I can say what a solution means. This will be the case if, instead of talking about solutions and dynamic equilbria, I interact with the assemblage, regard it as similar in a functional manner, and employ it as an extension of my thinking process.

ACKNOWLEDGEMENT

I should like to acknowledge the very close cooperation of Dr. E. W. Bastin in writing this paper. He has clarified many issues which were obscure, and a number of the ideas which are submitted have arisen jointly in the course of our discussions.

REFERENCES

1. ASHBY, W. ROSS. Design for a Brain. *Chapman and Hall, London.* (1954).
2. ASHBY, W. ROSS. An Introduction to Cybernetics. *Chapman and Hall, London.* (1956).
3. BEER, R. STAFFORD. Industrial Cybernetics.
4. BEER, R. STAFFORD. The Scope of Operational Research in Industry. *J. Inst. Prod. Eng.*, 1957.
5. BRUNER, GOODNOW and AUSTIN. A Study of Thinking. *Wiley, New York.* (1958).
6. CHERRY, COLIN. On Human Communication. *Wiley, New York.* (1957).
7. CORBETT, B. D (Mullard Equipment Ltd.) Unpublished Data.
8. GEORGE, F. H. Probabilistic Machines. *Automation Progress*, 1958, 3, 1.
9. MACKAY, D. M. The Quantal Aspects of Scientific Information. *Brit. J. Phil Sci.*, 1951, 6.
10. MACKAY, D. M. The Epistemological Problem for Automata. Automata Studies p.235 ed. by C. E. Shannon and J. McCarthy. *Princeton.* (1955).
11. UTTLEY, A. M. The Classification of Signals in the Nervous System. *E. E. G. Clin. Neurophysiol.*, 1954, 6, 479.
12. WALTER, W. GREY. The Living Brain. *G. Duckworth.* (1953).

Reference is also made to:

13. PASK, G. The Growth Process in a Cybernetic Machine. *Proc Second Congress International Association of Cybernetics.*
14. PASK, G. Organic Control and the Cybernetic Method. *Cybernetica.*

APPENDIX - PHYSICAL ANALOGUE TO THE GROWTH OF A CONCEPT

by

GORDON PASK

Certain of the demonstrations which were not envisaged when this paper was written clarify essential points in the argument. The models concerned will be examined with reference to these points.

In order to develop a thread structure we require, in the simplest case, an assemblage of raw material, in the demonstration model of ferrous sulphate solution. The energetic conditions which have been described are applied to this assemblage and in order to realise these conditions in practice a learning machine must be introduced such that it distributes the limited current available in a manner which ultimately develops threads. The most general learning machine is a device able to pass current through each of a finite set of electrodes, by making connection with these electrodes. Connections are made and are regarded as trials. The result of a trial may lead to greater or less current passing via the electrode to which it refers. The learning machine is programmed to learn that trial making activity which maximises the current passed via each electrode subject to the limitation that only so much current may pass in the assemblage as a whole.

It has been pointed out in connection with a more specialised assemblage that the result of such activity on the part of a learning machine will be development of thread structures adjacent to the trial making electrodes. These modify the feedback obtained as a result of subsequent trials. Soon the characteristics of the learning machine are largely determined by the thread structures, and only trivially by the characteristics initially possessed by the learning machine. In particular the comment applies to the 'memory' of the learning machine.

Carrying the argument one stage further, it is unnecessary to introduce a formal learning machine at all. Rather, as shown in the demonstration, amplifiers may be associated with a sub-set of the electrodes and in the simplest case these are 'binary amplifiers'. A 'binary amplifier is a component which receives an input signal proportional to the impedance between the electrode to which it is connected and other adjacent electrodes. The signal is averaged and when the average exceeds a limit the amplifier delivers an output current impulse into its own electrode and this current impulse gives rise to a thread structure. The term 'electrode' used in specifying the input is, however, taken to mean not only the point of platinum wire, but the thread structure which is, at

any instant, associated with it. If we apply to the assemblage the condition that

$$\sum_{(1)} \varphi_{(1)} = \lambda$$

where $\varphi_{(i)}$ is defined, in Section 4, the current passing on the average, via the (i)-th electrode, we find that a thread structure develops which is functionally equivalent to a deliberately constructed learning machine.

If threads are equally likely to develop at all electrodes, and if current is passed through the assemblage a state of unstable equilibrium will develop ideally characterised with a set of current limited threads. Although the system is practically unrealisable, the concept is useful since it is clear that if the unstable state is disturbed a thread structure tending to obliterate the disturbance will be produced.

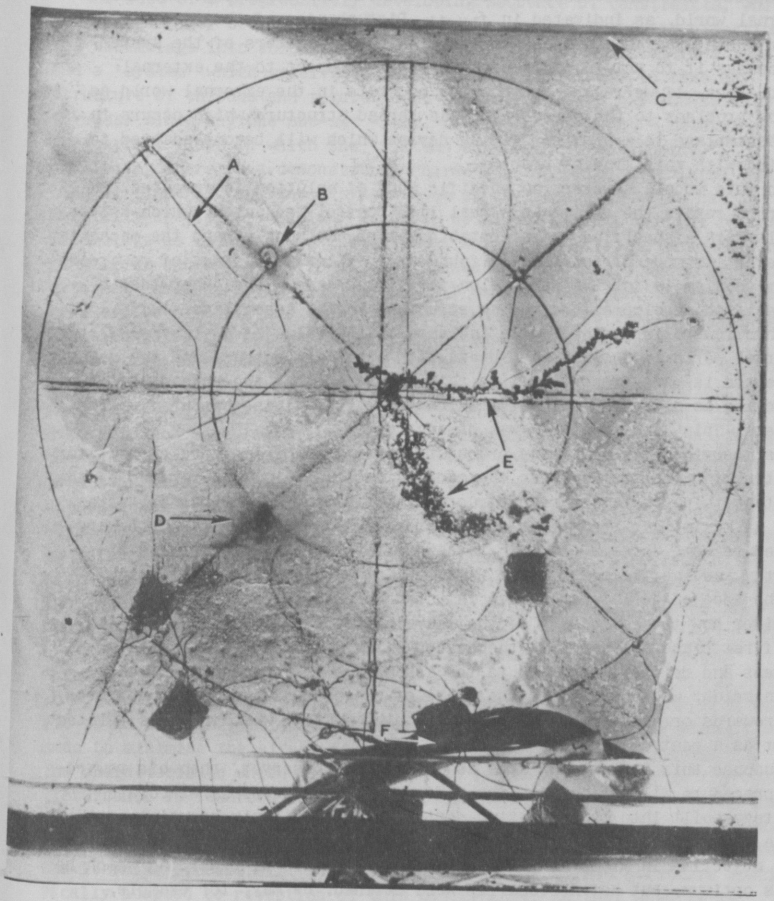


Fig.12. (A) Connecting wires for electrodes.
(B) Platinum pillar electrodes.
(C) Edges of glass tank containing ferrous sulphate.
(D) Chemical reaction in progress
(E) "Tree" threads being formed.
(F) Connecting cables.

(Reproduced from *British Communications and Electronics*)

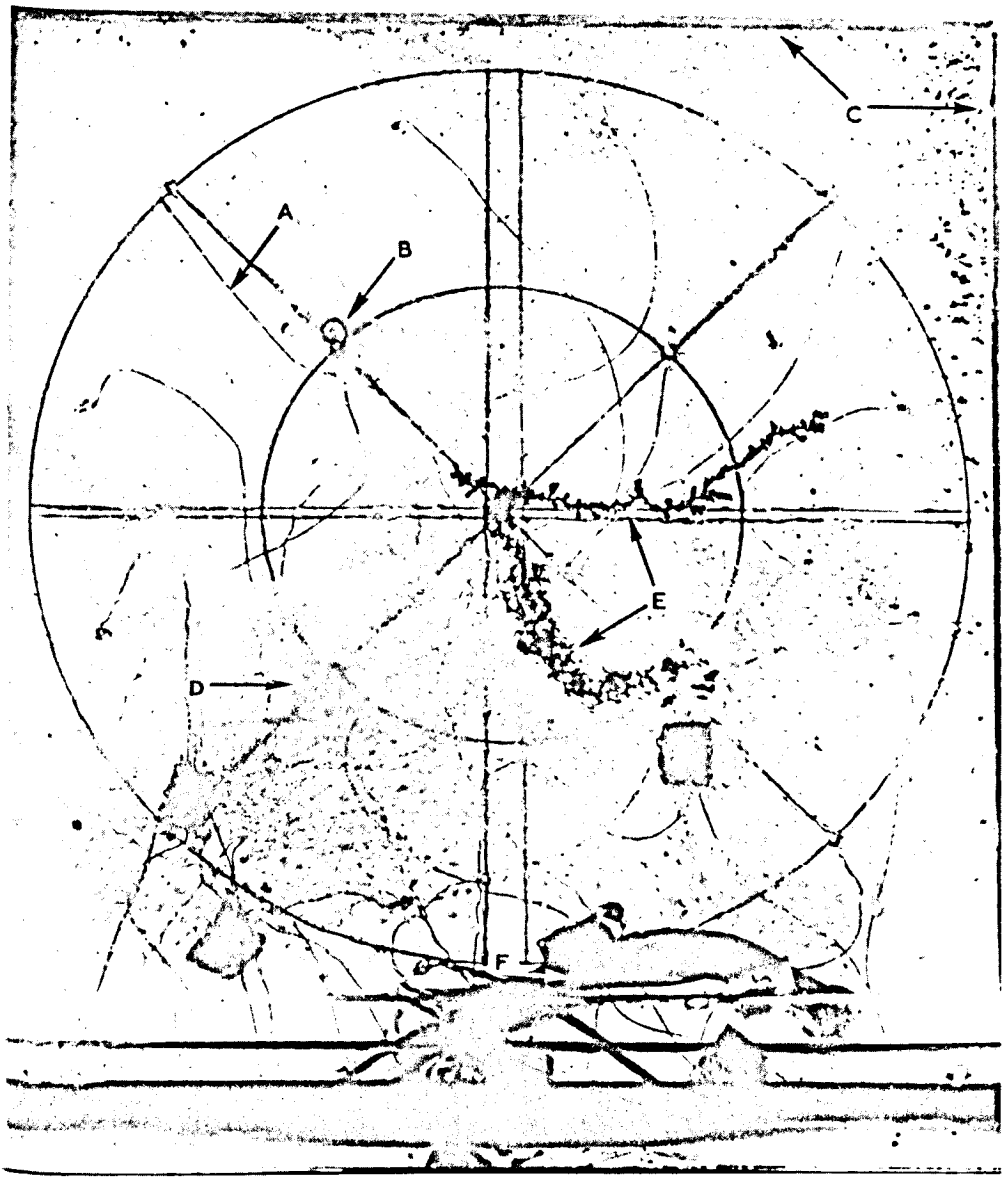


Fig.12. (A) Connecting wires for electrodes.
(B) Platinum pillar electrodes.
(C) Edges of glass tank containing ferrous sulphate.
(D) Chemical reaction in progress
(E) "Tree" threads being formed.
(F) Connecting cables.

(Reproduced from *British Communications and Electronics*)

Suppose that connections are set up between the assemblage and the external world, as indicated in *fig. 11*. Disturbances of the system will occur whenever a change of state modifies the parameters of the system, and causes it to act in a different way with respect to the external world. Thus, if we regard the changes of state in the external world as posing problems to the assemblage, the thread structure which occurs in the assemblage is a problem solving device which will become adapted to dealing with this kind of problem.

In this model, however, no specific mode of solution is favoured. In order to regard the system as a real life control mechanism, which solves problems in a specified and acceptable manner, we must reward the mechanism whenever its mode of solution is found acceptable. The effect of rewarding the mechanism is to give permission for an increased, but otherwise unspecified development of thread structure in the assemblage.

Thus reward is defined as an increase in the value of λ. In *fig. 11*, the external world is shown as a process with a certain output, and the value of reward is made proportional to this output so that the thread structure is acting as a control mechanism which tends to maximise output from the process. This kind of system was demonstrated.

As described the rewarding procedure acts by supplying more current for constructing threads whenever the mode of problem solution, implied by the existence of a certain thread structure, satisfies an external criterion, such as maximising the output of the process. In this learning by reward procedure some threads flourish, others will prove abortive. It is a lengthy and inefficient kind of learning not unlike natural selection.

The idea of an E relationship is introduced into the model when we consider how the efficiency of the learning process may be improved. In the first place suppose that the device which measures the output of the process and delivers a reward to the assemblage is replaced by a real human being, who scrutinises the state of the process and either approves and rewards or disapproves of the developing system, which he sees related to it as a control mechanism.

Suppose this human being acts as an external observer, then his rewarding procedure will be in no sense different from a possibly elaborate strategy built into the computing device which previously scrutinised the state of the process. On the other hand, if he agrees to participate with some objective in mind, he will adopt some procedure of rewarding which he finds optimum, but not necessarily well defined. He may, for example, discover that rewards should be delayed, or that rewards should sometimes be withheld in order to make the system learn efficiently. If the process starts off with an unstructured assemblage the learning efficiency depends almost entirely upon the extent to which the observer is participant and thus E related to the system.

The idea of a progressively developing sequence of constraints, which act upon further development is clarified by making imaginary bisections of the assemblage into components of an 'observer' who administers a reward and a 'system' which receives a reward. It is legitimate to take any structure of threads and regard these as rewarding another sub-set of threads, since any one sub-set - the 'observed' component - determines the current field in which the 'system' component is developing.

Finally there is a demonstrable characteristic which deliniates Condition 8, since it typifies a system made up of components which had no initially specified function. We have, up to the moment, thought of the assemblage as connected to the external world by electrical channels, and it will be convenient to regard the connection from the assemblage into the external world as still mediated in this manner. The input connection may, however, be of any kind. True there will be some electrical connections established at the outset, but it is not impossible that changes of temperature, chemical constitution, vibrations, magnetic fields, and so on, will affect the development of the assemblage and serve as inputs. In general one of these arbitrary disturbances will be an input, if it is relevant upon some logical ground, to the process, and if, in particular, the assemblage would be rewarded if it were able to sense the input in question. Thus, for example, a buzzing sound emitted by the process, although not explicitly included in the original set of inputs, as it might have been by provision of a microphone and an appropriate wire, will be relevant if reception of the buzzing signal without provision of a microphone, leads to a state of the assemblage which would not otherwise have occurred, and which in turn gives rise to a rewarded state of the process.

As a receiver of vibration, and thermal stimuli, a structure of threads is very inefficient. However, like any other physical system, it is to a certain extent sensitive to these variables and due to the fact that its parts have no initially specified function, no deliberate attempt is made to minimise the effect of such disturbances upon its workings. The characteristic which was demonstrated is that, supposing an arbitrary disturbance is rewarded persistently over an interval, and supposing that the assemblage is able to sense the disturbance occasionally, in the way that a computing machine might sense a vibration via a microphonic valve, the learning process will lead to a region of the thread structure specifically adapted to reception of this disturbance, so that when it is repeated it exerts an increasing effect upon the state of the system. Put crudely some region of the thread structure becomes adapted to react like a microphone, and in the particular case of a vibration it is possible afterwards to look back and see two conditions, namely -

(i) Terminal fibrils on a thread resonant with the vibration frequency, and

(ii) A field constriction such that these threads are in a position where a high current passes through the system.

There is no sense in which a microphone has been designed into the system. Rather, as a result of learning, the system has selected attributes of the external world which are relevant to its problem solving activity but which were not initially specified as part of its input set.

DISCUSSION ON THE PAPER BY MR. G. PASK.

MR. N. KITZ: I would like to ask some of the speakers to enlarge a little on the question of learning in computing machines, which seems a most difficult thing to achieve. The basic problem in learning really concerns, not machines, but beings such as human beings or animals with a will to live. The reason why, in my opinion, anybody learns at all - and I speak here as an engineer and not as a biologist is the instinct of self preservation. For example, if you are training a kitten and you want to have it house trained, which most people do, then if it misbehaves itself you so to speak rub its nose in it. You make it go through an unpleasant experience which it will remember and will try not to bring about again. Similarly, a small child will learn to keep away from the fire, not by virtue of the fact that its mother may have told it to do so, but because sooner or later it will burn itself.

As a computer engineer I am fascinated by how you could reproduce such a thing in a computer. I am not aware that any computer has ever wanted to work: this may seem funny, but it is perfectly true; and rather fundamental.

What kind of reward are you going to give to your computer to make it learn? It has already got a delightful environment - a nice big room and air conditioning - which is more than most human beings get. Yet this hardly makes it reliable. In fact very often you have to coax it to do what it is supposed to do, let alone learn new tricks.

I do not want to elaborate this too much and I am not really trying to point out the funny aspect of it. I think it is a real point. A computing machine which is essentially a device capable of doing at speed what it has been built to do seems to me to be fundamentally unable to learn. If you want a device to learn you must depart from the digital method and certainly the binary on-off type of method. A lot has been said about memories but the memories I understand are purely methods of recording information as it was at the time. You cannot get a computer to improve on its stored information at all. In fact, the only thing a computer will do is degenerate it.

How do the previous speakers feel about the problem of making machines that will want to learn and will therefore do things that they were not originally designed to do?

DR. J. McCARTHY: I think what the last speaker said contains a number of common fallacies. I shall discuss only the point about self-preservation

being the basis of learning. I do not think that you can explain the learning by children or animals entirely by self-preservation, as exemplified by a child learning to keep away from the fire by burning himself. That is an extreme example; children do not ordinarily face such severe and immediate punishment for doing wrong. I think, from some observation of children, that they have developed, by the time they reach the age of one year, a mechanism which contains some positive seeking of information and experience. If we try to base our methods of making a machine learn on self-preservation, we will construct something too crude to learn anything interesting. If we want the machine to learn anything substantial, we must build into it specific mechanisms for learning, that is, mechanisms for trying things and selecting the best of them according to criteria. Otherwise, the machine will not learn anything interesting until it evolves such specific mechanisms by more primitive methods.

DR. J. PATRY: I would ask the two previous speakers if the expression "learning machine" is exactly the right one. The speaker talked about a child and the fire, the child learning not to come too near because it is hot. Of course, a machine cannot learn from outside. We must put into the machine what we want from the machine and the machine can generalise, can make statistics, and make a choice out of the statistics, and we say "it is learning". I am not quite sure that the expression is exact. I think that the preceding speakers do not agree with each other only because of this special point.

MR. R. H. TIZARD: Since Mr. Kitz suggested I had been very clever in making a digital computer learn. I would like to take the opportunity to deny such a flattering suggestion. In fact the purpose of this computer programme was to show how simple a 'learning' type of character reader could be. I can only say that I plead guilty to following the very common practice of using a word which happened to be convenient for the purpose. I would add that the greatest danger in the whole of this subject is that of confusing very simple concepts with very high-sounding phrases.

DR. J. V. GARWICK: I will only make a short remark on how to teach computers to learn by rewarding or by kicking them. When we in this connection use the word "computer" we do not mean the electronic equipment alone, but this with a programme inside. The only reasonable "kicks" and "rewards" will therefore influence the programme in some way, not the physical machine itself.

DR. D. M. MACKAY: About four years ago, we made some experiments in King's College on the possibilities of electrolytic growth in a learning mechanism, using silver in silver nitrate; and some two years ago my assistant,

Mr. Pritchard, and I demonstrated this kind of process in a lecture to the Maxwell Society there.* But I want to tell you more about the snags than the successes. The first point is that if you grow dendrites in a liquid as Mr Pask has done, then of course the system is highly susceptible to mechanical disturbance: and while a good learning mechanism should not mind small disruptions, it is informationally wasteful if it has continually to relearn what it has built up. To avoid this, we first tried using gels, but these attempts were not particularly successful. Wet filter paper proved to be a much more suitable substrate. Despite a tendency for growth to be along the filter fibres, it was perfectly possible to 'condition' a connection by controlling the field distribution. What we did find disappointing, however, was the relatively small variation in (A.C.) impedance until connection was nearly complete. For this reason I think that electrolytic fibre-growth may be more suited to all-or-none types of conditioning, where one just wants a contact to be made or broken, than to systems where probabilities must be continuously varied.

It may also be worth throwing into this pool of ideas a fact well known to telephone engineers as a nuisance, namely that on solid insulators you can get dendritic growth, of silver for example, which is embarrassingly permanent. *(ref. 1)* It seems possible that a development along these lines, using a solid substrate, might provide the most suitable way of making relatively permanent self-wiring systems. In any case there can be no doubt that more basic research into the physics of electrically-guided growth processes could be rewarding in the field of automation. I should be interested to hear if Mr Pask has done any work of this kind.

DR. SELFRIDGE: This is a slightly orthogonal comment. I think that the beautiful experiments of Dr. Pask are indicative of a strong trend. Some of us feel very strongly about the next generation or generations of computers. We feel they are going to have to be in some sense more parallel, much more than they are now. To some extent there is the problem of finding people to wire them. I would like to suggest the wiring of a complicated machine, or a simple machine for that matter, under electronic controls. I am not proposing right now a machine could successfully rewire another machine or itself extremely fully, but nevertheless the notion of electronic wiring is a very valuable one for the future. People are getting sort of expensive.

* 'Intelligent mechanisms': 11th Feb. 1957.

REFERENCE:

1. Kohman and others: 'Silver Nigration in Electrical Insulation'. *Bell Syst. Tech. J.*, 1955, 24, 1115.

It would be very nice to have a machine build another machine electronically without any physical motion involved, and this is the second such mechanism which has actually worked. The first one probably most of you are, like myself, too young ever to have heard of. It was, I think the way the first radios worked, with coherers.

MR. G. PASK (in reply): If you don't mind I will take the replies backwards and deal first of all with Dr. MacKay's points. I am sorry I didn't know about work which he has done in this field, for it is clearly relevant, but judging from his statement we were aiming for quite distinct objectives, and I think it will be valuable to dwell for a moment on the differences. As I understand it Dr. MacKay has considered threads as a kind of probabilistic storage device, and has found them unsatisfactory for this purpose. It is not surprising, for although a thread may be caused to develop along a line of maximum potential and thus to vary the impedance to an A.C. sensory signal between pairs of electrodes, the impedance variation is non-linear, and the whole process of development depends upon many variables. For use in a conditional probability machine, I imagine, this sort of dependence, which Dr. MacKay has illustrated by referring to mechanical sensitivity, will be an embarrassment. However, I wish to emphasise that these very properties are those which, in the present application, we should try to encourage, in other words, I would like the threads to exhibit sensitivity to all sorts of disturbances, not only the particular set of disturbances which as a computor designer I might regard as relevant; for example, the set of disturbances which amount to changes in the field parameters. Again, I am far from embarrassed by the fact that threads may develop according to a variety of mechanisms, such that upon any particular occasion, I frankly don't know how a thread grows. True, it would be very pleasant to know, but one is looking for a device where there is such a variety of ways of developing that it will always be difficult to know, and possibly not worth while.

Because of this we have not paid very much attention to rendering these threads more stable. However, this is a matter of degree, and we have found techniques such as embedding the thread in silica, developing them upon filter paper, or blotting paper, as Dr. MacKay has done, useful upon occasion. I should like to add a framework of small glass beads to the list, since I think that this was the most successful method of restricting the threads. Furthermore this method is applicable to three dimensional rather than two dimensional thread structures, and these three dimensional structures having a more reasonable topology are those which are of practical interest. The demonstration aimed at two dimensional structures simply because these are more readily exhibited, but some of you will probably have noticed that it is difficult to restrict the threads to a plane since they tend to arch over one another at the slightest provocation, whatever is done to prevent it.

The possibility of solid state systems is very interesting indeed. I fancy these would be more useful for determinate applications, since there seems no very obvious way of getting rid of structures once they have been formed.

Returning to the actual use of these threads which was demonstrated and discussed in the paper, I should like to summarise the characteristics which are essential for any components used in an E. assemblage. The first characteristic is that raw material, in this case, ferrous ions in solution, shall be transformed by the expenditure of energy into a structure, in this case, a metallic thread. Secondly this structure, once created, must be able to serve a variety of different functions, and because of its origin, it will be legitimate to say that the function of any element in the resulting system is not determined a priori. A thread structure, for example, may serve as a conducting pathway, or as a register, or as an element which constrains the field, in which other threads develop, or even as an amplifier, as evidenced by the interference of two threads at differing energy levels.

The possibility of threads acting in this way, in particular, as acting as part of an amplifying system, depends upon the existence of a tendency for any thread, once constructed, to dissolve. A structure exists, in other words, so far as it is able to compete in a dynamic equilibrium against a tendency for return to the raw material out of which it was made. It seems to be a general property of systems which satisfy these requirements that they are all liable to change state in response to a wide variety of physical variables, and I do not think that this feature is accidental. It is, of course, possible to so restrict the activity of such a system that it would be inadmissible (in an E. assemblage, but useful in other applications. This has not, however, been the object of these experiments.

We have done very little fundamental physics. What experiments have been done, in my laboratory, on the organisation of a thread structure suggest that the process is a very complex affair. Although it is convenient to think of the development of further threads in a multiple electrode system as being chiefly determined by the constraints imposed upon the field by those threads which already exist, there are a variety of other factors to consider as well. For example, the collection of ferric hydroxide gell in the solution is important, and it is possible to discern quite complicated colloidal systems built up regionally - systems with semi permeable membranes, and so on - which occur due to local collections of the gel. This is not objectionable for the use which I am making of these threads. It does, however, prevent one saying precisely what mechanism is involved in mediating a change. Any change may be produced by many different mechanisms.

One may, of course, make an arbitrary distinction between regions in a dish full of a self organising system like this. If you do so you might

think of one region as the "system", and the other region as the "environment". I submit that the state changes which appear, using this framework of reference, are probably akin to those you had in mind when you were talking about two interacting systems, one of which became progressively unconstrained as it becomes able, by developing fresh constraints, to deal with its environment. Isn't this the sort of thing we are aiming at?

DR. MACKAY: Yes, but there is a difference of approach in that we would rather understand the system first and then apply it.

MR. PASK: So should I, and in the sense of knowing its logical characteristics as against its detailed physical properties, one does understand it. But the emphasis has been on getting something which works rather than finding the ideal. Further we have been mostly concerned with the philosophy of the system and its broad application. These metallic threads, for example, are only one of many systems which have been tried. I don't think they are even the best, but they are probably the most easily demonstrable.

In reply to Dr. Selfridge's question one of the first systems which I examined as a possible E. assemblage was somewhat like a coherer, in that a conducting pathway was made between semi conducting granules. However, in this case, and I believe I am right in saying in the case of a coherer, there is no sense in which the existence of this structure depended upon successful competition against an obliterating tendency. The thermal network described in the paper has been rejected, largely because it is too coherer like. In other words, it is easy to produce a conducting pathway, which once made, will be preserved, supposing that energy is supplied, but the kind of system I needed was certainly not just a device for ingraining a conducting path along lines of maximum energy flow however useful this may be in a different context.

Some of the comments about learning have already been answered. Whilst I agree entirely with Mr. Tizard, in the case of a learning by reward programme on a computor, there seems to be a sense in which a mechanism working in this entirely different field may learn in a more human like manner. It is in the nature of a self organising system to learn somehow. Further, it will necessarily learn by reward, since reward will generally mean that it has more energy to develop with. Because reward means this, and because the system may use this energy in constructing fresh parts, of its own structure, there is a perfectly good sense in which one says it is rewarded by self preservation. What I am trying to bring out is a basic distinction which exists between reward, as used in a computor programme to mean that a rewarded event becomes more probably, and reward as it used in a system able to create fresh components and parts of itself. In this latter case, reward means ability to develop, ability to expand, and ability to become stable by becoming a larger system. But, this point has been dealt with at some length in the paper.

LECTURE - DEMONSTRATIONS

LECTURE - DEMONSTRATIONS

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

A description of the demonstration is given on page 933.

2 The growth of concepts and their physical analogues.
MR. G. PASK, System Research Ltd., London

This demonstration is described in the Appendix to Mr. Pask's paper, Session 4B, Paper 7.

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

A description of the demonstration is given on page 945. Further details of the Conditional Probability Computer are given in the papers by Dr. A. M. Uttley, Session 1, Paper 5 and by Dr. A. M. Andrew, Session 3, Paper 5.

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

A description of the demonstration is given on page 947.

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

A description of the demonstration is given on page 951. See also the paper by Dr. Taylor, Session 4B, Paper 5.

6 Automatic Speech Recognition
PROF. D. B. FRY and MR. P. DENES, University College, London

See the paper by Prof. Fry and Mr. Denes, Session 3, Paper 2.

7 RRE Library Retrieval System
MR. S. WHELAN, Royal Radar Establishment, Malvern.

A description of the demonstration is given on page 953.

8 DEUCE: Experiments relating to visual perception and simple self-modifying programmes
CME Division, NPL

9 ACE: Automatic Computing Engine.
CME Division, NPL

10 Optimal Coding Device
MR. P. E. DONALDSON, Physiology Dept., Cambridge University

This demonstration is described in the Appendix (by Dr. Barlow and Mr. Donaldson) to Dr. Barlow's paper, Session 4A, Paper 1.

11 The Grouped Symbol Associator (an aid to Medical Diagnosis)
DR. F. A. NASH, Western Hospital, London

This exhibit is fully described in the Lancet, April 24th, 1954, pp. 874-5. See also Dr. Nash's contribution to the paper by Dr. Paycha, Session 4B, Paper 4.

MACHINA REPRODUCATRIX

AN ANALOGUE MODEL TO DEMONSTRATE SOME ASPECTS OF NEURAL ADAPTATION

by

DR. A. J. ANGYAN

MEMORY has always presented an important problem to physiologists and neurologists, and there have been many attempts to interpret the brain and its memory function by physical analogues or illustrations. The early ones included Dubois Reymond, Pavlov (*ref. 16*) who used telephone analogues for both unconditioned and conditioned reflexes, and the Hungarian neurologist, Jendrassik (1912) who used analogies of physical induction and resonance. More recent theories of conditioned reflexes Young (1938, *ref. 22*), Hilgard and Marquis (1941, *ref. 9*), Konorski (1948, *ref. 11*), and Hebb (1949, *ref. 8*) used similar ideas in attempting to make hypotheses on plastic adaptation using various electronic circuits. All-or-nothing features of synapses were detected by microelectrode studies and discussed by McCulloch and Pitts (1943, *ref. 13*), and by Eccles (1953, *ref. 6*). Cragg and Temperley (1954, *ref. 5*) used electromagnetic phenomena, i.e. field processes.

However it is clear that all these analogies are only very tentative ones, even if they do show some features of the brain. They are useful in helping to explain the meaning of biological terms, and, not less important, they may help to clear up errors and inconsistencies in our terms, and of the physical analogues associated with them. Physiologists are obliged to define clearly the concept of memory as a mechanism of a highly organized complex living system.

Machina Reproducatrix, which is shown here (photo, *fig. 1*) is a model which attempts to demonstrate some, but not all, of the recently formulated concepts of nervous adaptation. It is based on Dr. Grey Walter's Machina Speculatrix and its development Machina Docilis. It is a relatively simple analogue model of a simplified pattern of the innate and acquired reflex connections of a living being. It is a simple model compared with the more refined concepts of Dr. McCulloch (*ref. 14*) or Dr. Rosenblatt's perceptron, Dr. Uttley (*ref. 19*) and his Conditional Probability system or Dr. Ashby (*ref. 4*) and his habituation and homeostat models.

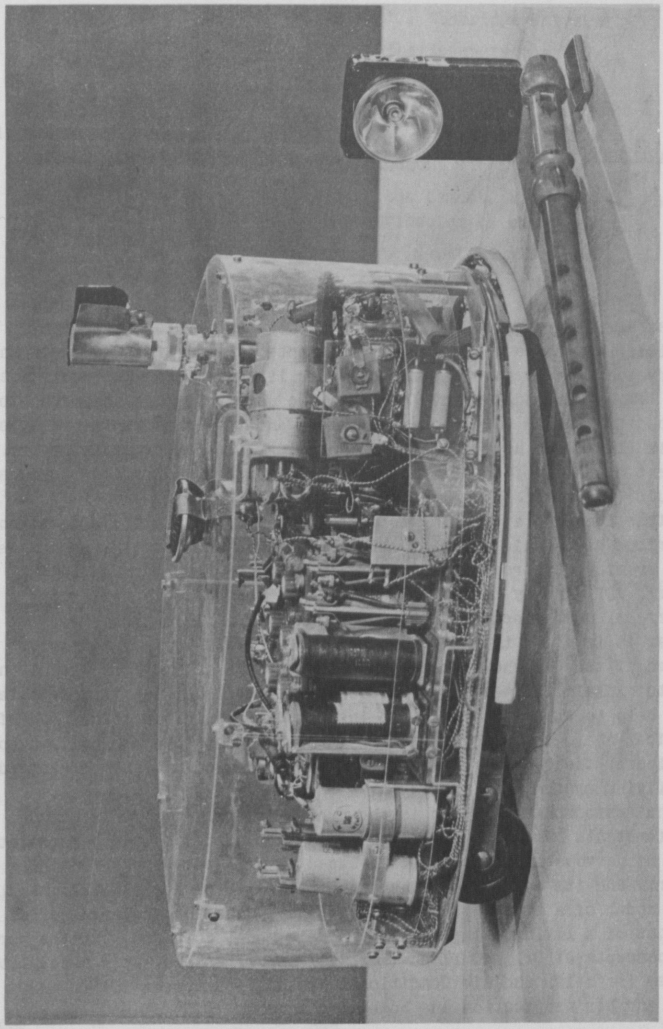


Fig. 1

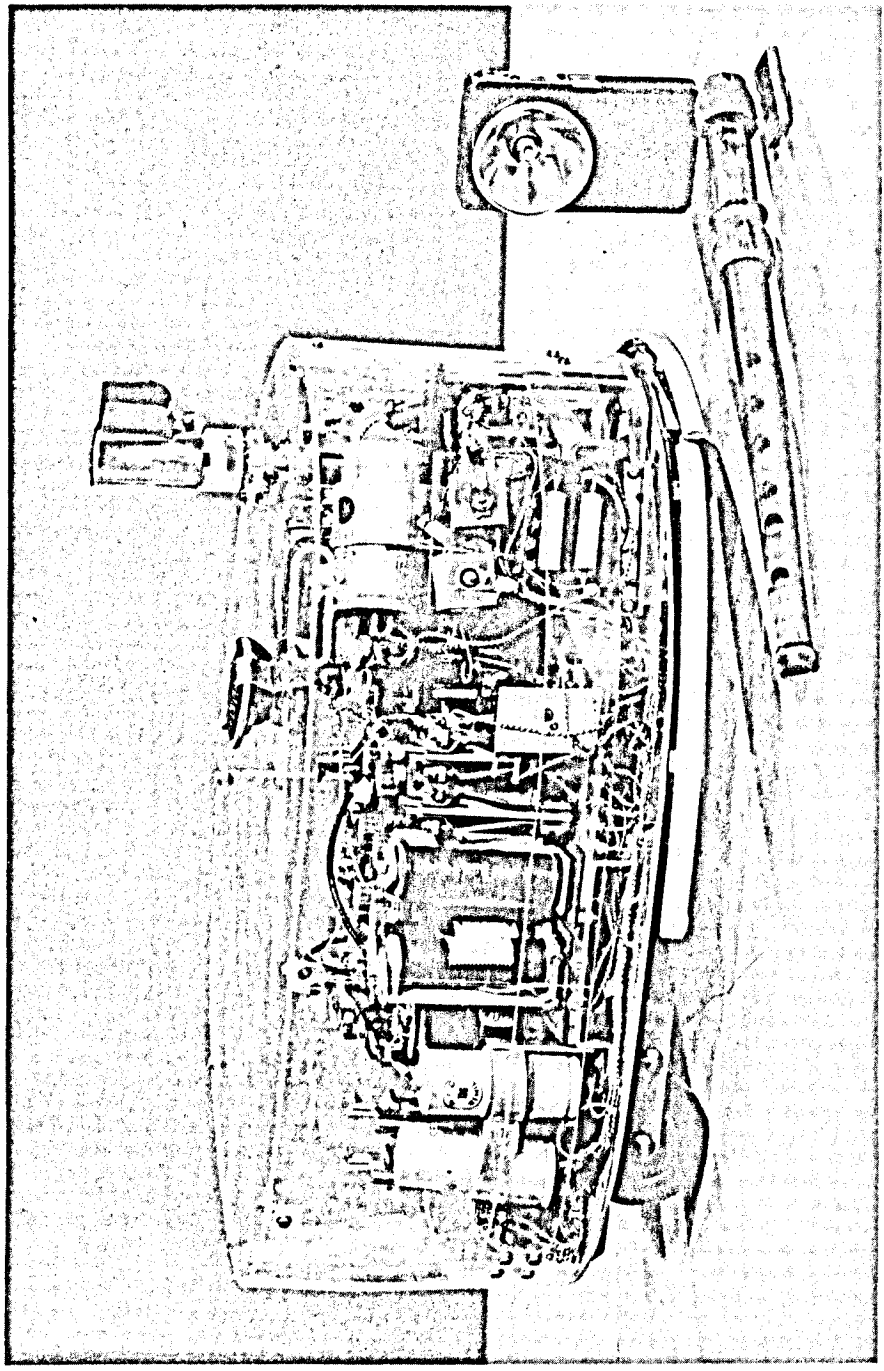


Fig. 1

The model has two or three receptors, for light, sound, and mechanical input. The signals from these receptors are relayed through two control "centres" each of which controls a motor; this dichotomous control serves to demonstrate a basic innate pattern of behaviour. (There is in fact a very important dichotomy inherent in all behaviour, between *well-oriented and directed goal-seeking and random searching.)*

Like most living things, the model, if once aroused, searches for a goal for one of its tropisms or inborn unconditioned reflexes. It searches for light because of a connection between its photo-receptor and effector motors. This behaviour, which is dominant in the system, is disturbed or inhibited by any other stimulus impinging on its mechanical or acoustical receptor systems.

But extreme strength or duration of the original dominant stimulus causes the model to seek a new dominant cue. This illustrates another principle of reflex activity which never allows an organism to follow any drive continually: by switching off the drive at some level or after some time it starts a search for a new stimulus. Analogies of this can be found both in whole organisms and in any artificially separated part of their nervous system. Two teleological principles seem to be involved, but by introducing the concept of positive or negative feedback the organization can be seen to be very simply determined.

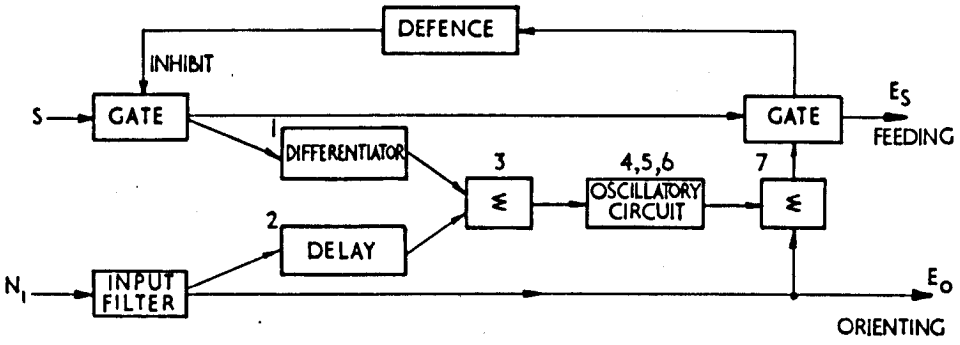


Fig. 2

A logical concept of *conditioning and learning* was built up by Dr. Grey Walter in his 'CORA' and we must study the action of 'CORA' briefly before going on to the next stages. *Figure 2* is a block diagram drawn in similar form to Machina Reproducatrix. The first three logical elements of Dr. Grey Walter's model can show a behaviour analogous to that which may happen in every instant of alert activity of the brain centres. Two stimuli, which originally were temporally or spatially separated, may meet on one or more synapse of the brain, ensuring alertness which is a basis for receiving communications from the external world. In 'CORA' this can happen if (a) the original driving or tropistic stimulus is differentiated(1)* corresponding to the on-off effects produced by appropriate synaptic connections at sensory inputs to the brain. (b) a second non-dominant or neutral stimulus preceding the dominant one is delayed in its effect(2) (corresponding to "after discharge" or widespread non-specific activation of synapses). (c) the two signals satisfy the necessary conditions, e.g. coincidence, summation(3).

So far there are no difficulties in explaining these features of neuron behaviour, but for building up a conditioning process a further logical step has to be introduced. This involves a temporal summation of the overlapping coincidence areas or the probabilities of the two stimuli. In the model this can be represented quite simply by a capacitance circuit(4); nevertheless in a nervous system it may involve a more refined mechanism e.g. it may work on principles similar to Dr. Uttley's conditional probability system.

Dr. Grey Walter points out that a nervous system, working as a learning box must rule out sheer chance and reach a threshold of coincidence for conditioning. There have been many hypotheses of conditioning, e.g. in the neurophysiologically almost correct work of Eccles (*ref. 6*) he attempted to localise it as a specific feature of the central grey matter or of the cerebral cortex because of the multitude of input and output connection in this region. But Pavlov pointed out (*ref. 17*) that in appropriate conditions the "Summation reflex" (i.e. association of two stimuli) may occur and give effects on any level of the nervous system, especially the less co-ordinated ones. Dr. Uttley (*ref. 19*) found that, in his conditional probability computer, the activation of a unit corresponding to the firing of a neuronal network does not suffice to distinguish actual occurrence from computed probability, and he introduces the postulate for a regenerative loop.

The next steps in Dr. Grey Walter's logical scheme are activation(5) and the preservation(6) of information by an oscillatory circuit with very light damping (an analogy of cortical alpha-rythym). These steps are

---

* Numbers in brackets refer to points on the block diagrams.

followed by gating[7] the oscillation and the original neutral stimulus to produce a response corresponding to the original dominant or unconditioned stimulus. The response is analogous with cortical neuron output.

At first glance this seems to be a correct representation of conditioning and of the memory function of the brain and *most hypotheses explaining conditioning on the basis of neural connections are similar.*

We decided, in our model Machina Reproducatrix, to reproduce conditioning in the same way as 'CORA' does, but instead of an oscillatory circuit we employed a neon tube with two relays and a thermistor (the slow rise of temperature of which produces a memory for paired stimuli for almost the whole running period). This solution was preferred since the original one was too sensitive to mechanical disturbances. This may also be said about the neural network which is exemplified by this circuit. In the introduction of the thermistor we may see a very slight analogy with the newly-established fact that the excitability and activity cycles depend on the activation system and thence the cortex on vascular effects. There is so far no justification for taking these analogies too far.

Though Dr. Grey Walter's learning box displays some surprisingly correct imitations of life it shows that *until the last few years the explanation of brain processes in conditioning and learning followed a very simple scheme.* Compared with any animal during conditioning, we must conclude that the CORA circuits omit some important characteristics. These are:-

(a) the conditioning trials do not lead to habituation of the neutral stimulus,

(b) other stimuli related to the conditioned one produce no effect, i.e. no generalization occurs.

(c) memory is lost as the oscillations gradually decay in the memory circuit and is finally extinguished. We must repeat the stimulus combination to obtain recovery which is never spontaneous as it is in animal conditioning experiments,

(d) a new disturbing stimulus may cause gross deficiencies in the function of such a delicate automaton depending on external influences. They seem to produce neurotic-like behaviour, but no marked external inhibiting or disinhibiting effects can be distinguished.

We therefore attempted to supplement CORA to overcome some of these defects. Machina Reproducatrix responds to three stimuli, light, flute and whistle. Six or eight repetitions of flute and light cause conditioning and the model turns it front towards the sound (just as the previous model did to light). But once the capacitive memory sustained by the thermistor circuit decays, no more conditioning effects are observed. If we repeat the combination of flute plus light conditioning occurs at once. If we now try the whistle, we may observe a conditioned effect i.e. a *generalization.*

But after several (about 18) repetitions of whistle without light no more such effects occur, whereas the flute (which with light was the original stimulus) remains a conditioned stimulus. This is simple *discrimination* and if we carry on fluting only, the effect of the flute is also extinguished or habituated and no tropistic behaviour is seen. But, in our model, the conditioned effects of the flute recover in a few minutes, but the whistle remains as an effectiveless stimulus. If a new sharp sound is intonated, even the flute may be depressed; this is *external inhibition*, and after a few instances the whistle will regain its effect, this is disinhibition. The flute also recovers its conditioned effect comparatively soon.

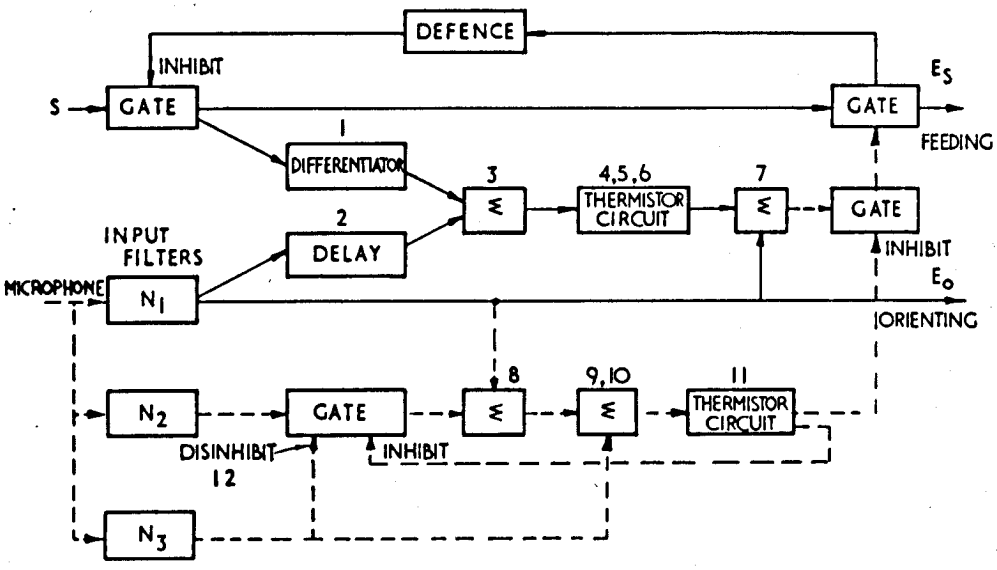


Fig. 3

Figure 3 is a block diagram of the original model, Machina Docilis, and the supplementary parts which have been added to make Machina Reproducatrix achieve these features. The three inputs N_1, N_2, N_3 have a common input to a counting device[8] connected with another thermistor circuit[9] [10] [11], the outputs of which are connected to the gating relay. If N_1 and N_2 occur together conditioned reflexes are obtained from any input which contains either, but stimulations due to N_1 or N_2 alone are counted and cause blocking for a given time of the conditioning effects. If a sufficient number of unpaired stimulations of N_2 occur it leads to permanent

inhibition of unreinforced stimulations of N_2 until N_2 is disinhibited. Thus we build up a discrimination. If a new stimulus N_3 is given, and it is strong enough, it connects a transistor circuit to produce disinhibition of N_2 and also inhibits the conditioning effects of N_1.

We realise that this model is somewhat crude, but it was conceived not only for demonstration purposes, but also to draw attention to some important questions. The supplement to "Machina docilis" was conceived to represent the phenomenon of *habituation* which results from any biological stimulus. It is a basic mechanism by which extinction of conditioned reflexes and their internal inhibition leading to discriminations is built up physiologically. But its effect is temporary, and may be thought of as a negative feedback to the changes in neural connections of every nervous adaptation. In fact, the conditioned reflex cannot be fully represented by taking account only of association, combination or summation of two stimuli, since the mechanism of habituation occurs in parallel and counteracts any accumulation of corrections. (Thus, it ensures the conditions necessary for the coexistence of a "conditional probability" and a "conditional certainty" system in the brain). Experimental observation on the mammalian and human brain especially in the Neurophysiological Institute of Pecs and also in the Neurophysiological Clinic in Debrecen, (Lissak and Grastyan *ref. 12*, Kajtor et al, *ref. 10*), clearly suggested that it is the hippocampal system which may act like a counting device for any sensory inputs. This agrees with Green and Arduini's electrophysiological observations. They observed that, following stimulation, the hippocampus may inhibit conditioned reflex activity, independently of the origin and dominance of conditioned or unconditioned stimulations, and this may be demonstrated by studying orientation reflex phenomena of the model. Now it seems that the supplementary parts of the model may be compared with the negative-feedback effect which the hippocampus has on neocortical and brain stem phenomena in conditioning. We should leave further neurological discussion of the structural analogy and point only to some observations by which the validity of this analogue can be tested. The electrophysiological observations of Morell, Jasper et al (*ref. 15*) have shown that a significant impairment of electrocortical conditioning occurs only with archipallial (hippocampal) lesions and Penfield and collaborators (1958, *ref. 18*) pointed recently to the fact that distinctive, bilateral lesions of the hippocampus only wipe out the recollection of recent memory experiences. But even if we control the effect of such lesions or that of any extended cortical lesion in a conditioned reflex experiment, we may find that the association process by itself is less impaired than discrimination and appropriate recall (Angyan, 1956, 1957 *ref. 2*). This somewhat too docile impaired learning mechanism is modelled by "Machina Docilis". We get the same result by comparing the brain adaptation with any model taking account mainly of the conditioned summation reflexes.

Our model demonstrates (1) that habituation and internal inhibition must occur *in parallel* and must to some extent inversely regulate the conditioned summation or probability computation process. (2) that *discrimination* is based on a *common mechanism with habituation* but it may be further improved if specific inputs with specific analysing mechanisms are considered. Some degree of discrimination may be obtained by filter mechanisms, but it is only effective if it is regulated by extinction based on habituation. (3) that *spontaneous recovery* (or recall of *forgotten or extinguished* conditioned reflex) cannot be based on preferential states for more recent stimulations (as in Dr. Uttley's concept). Our model though far from being a sufficient analogy, allows the repeated recall of an extinguished conditioned reflex during the whole span of its memory.

The fact that our model is still of insufficient complexity is shown by the effect of external inhibition. By causing a block through the orientation-habituation mechanism, this wipes out both the effects of positive and negative stimulations. To obtain a recovery of discrimination, we need to build up again the extinction or discrimination process. In our opinion, this problem can be adequately developed by constructing a model which incorporates *another* very important and experimentally well-founded fact of nervous adaptation - the Sherrington principle of *reflex antagonism* and mutual induction. Pavlov has pointed out that this is the third basic mechanisms (with excitation and inhibition) in the maintenance of dynamic equilibria of every complex of innate and acquired behaviour.

If we assume that, in parallel with the extinction process or the development of any internal inhibition, that the stimulus to be discriminated build up a connection with another specific stimulus in opposition to that already conditioned (e.g. summed with a negative stimulus), it would not be too difficult to build a model which demonstrated electronically this mutually or reciprocally inductive antagonism (perhaps in the sense of Wiener's 'moth and bedbug' model combination). If the two summation systems are coupled together mutually in an inverse feedback manner, an automatic mechanism is obtained which resets the original state of discrimination immediately after the disturbing stimulus disinhibited it. In fact, this occurs in uninjured, normally adapting, animals, with a speed and repeatability which is very characteristic of the individual. It is one of the most important transformations which sometimes seems to almost entirely override the rules of conditioned association in everyday psychic activity. It is usually the first mechanism to be impaired following functional or structural disturbances of the brain. We are convinced that this mechanism of 'direct' inhibition is an inherent structural feature of nervous organization. Every unconditioned reflex has a positive 'to' and a negative 'fro' aspect (seen in defence, feeding and sexual activities). These form the basis of unconditioned reflexes in their lower and higher manifestations.

Plasticity (acquired individual adaptation) is however maintained by habituation, or by indirect inhibitory mechanism, which never allow 'summation reflexes' to build up above a certain limit on the basis of the former ones as a result of new stimuli impinging in their networks. Plasticity also counteracts the development of fixed correlations between the summation reflexes, in contrast, e.g. with the spinal cord's reflex organization. In an ideal model system which also contains the development analogies it should be shown, as in the mammalian brain and in general in animal of the main phiietic line, that (a) these mechanisms are acting along spatial axes according to Child's gradient (b) the habituation mechanism in its interaction with a summation of probabilities is never allowed to be static (c) the opposing coupled basic reflex systems cannot be static. If it satisfies these conditions the model system fulfils the requirements of Dr. Ashby's concept of homeostatis maintained by ever-changing dynamic equilibria and regulated by partly deterministic and partly probabilistic percetion systems. But our supplemented model has functional correlates in the high organization level of the mammalian brain. The dichotomy of its design is also based on simple experimentally-demonstrable properties of neurons just as in the hypothesis of Eccles - *impulse treshold changes, synaptic use and disuse, reciprocal antagonism and feedback principles.* There are also probably differences in the axosomatic and axodendritic connections of neuronal adaptation similar to those of summation and habituation on the higher organizatory level.

CONCLUSIONS

Our model cannot exclude the possible generalization that learning is simply due to a coupling of two mechanisms - say a positive or excitory feedback and another negative feedback on inhibition. This coupling can occur at every structural level of biochemical, electrophysiological, or functional anatomical organization, and to show its special adaptive features we must find out the principles for further development which are directing the mechanism whose nervous system quickly adapts to its environment. Recent observations of the author (Angyan et all 1957, 8 *ref. 3*) on the behaviour of flatworms during regeneration of the cephalad and caudad parts of their primitive nervous system shows that the two aspects of nervous function referred to above may be separated from each other functionally and structurally by simple experiments. It seems that both developmental and behavioural mechanisms show interesting polarization along the developmental axes of Child's gradients. Perhaps the conditioned reflex is no more than a temporal expression of an axial developmental mechanism propelled by an excitatory summation which shows more general rules of polarization and is limited by inhibitory habituation.

Since in complex organisms we are dealing with at least three special dimensions or axes (and perhaps a temporal one as well) we are inclined to suppose that a more complete analogy of learning and conditioning would be represented by a model in which three antagonistically-coupled CORA or Reproducatrix systems were linked and their varying dominance cancelled by an adequate internal scanning or temporal summing mechanism. This idea forms the basis for further development of our model (Angyan, Zemanek and Kretiz, 1959).

In our opinion, models are only useful in that they help to draw a better concept of present knowledge of brain function the realization of which may show whether our ideas and terms may correspond or should be excluded from the interpretation of life.

REFERENCES

1. ANGYAN, A. J. : The role of the orientation reflex in the connecting and signalisatory function of higher nervous adaptation. Dissertation. *Budapest, Hung. Ac. Sc.* (1955).
2. ANGYAN, A. J. and NEMETH, J. : Observations on the conditioned reflex behaviour of Turbellaria during and after regeneration. *Abstr. Hung. Physiol, Congr. Kiserl. Orvostud.* (1957).
3. ANGYAN, A. J., GAUTIER, B. and NEMETH, J. : Effect of neurotropic drugs on the behaviour of flatworms. *Abstr. Hung. Physiol. Congr. Kiserl. Orvostud.* (1958).
4. ASHBY, W. R. : Design for a brain. *Chapman & Hall, London* (1952).
5. CRAGG, B. G. and TEMPERLEY, H. N. V. : The cooperative analogy of neuron organisation. *EEG. Clin. Neurophysiol.*, 1954, **6**, 85.
6. ECCLES, J. C. : The neurophysiological basis of mind. *Clarendon Press, Oxford* (1953).
7. GREGORY, R. L. : Models and the localisation of function in the central nervous system. Session 4A, Paper 5.
8. HEBB, D. O. : Organisation of behaviour. *John Wiley and Sons, New York* (1949).
9. HILGARD, R. and MARQUIS, F. : Conditioning and learning. *Appleton-century-Crofts, New York* (1941).
10. KAJTOR, F. : Epileptic manifestations of the human temporal lobe. *Orvosi Hetilap*, 1955, **16**, 421.
11. KONORSKI, J. : Conditioned Reflexes and Neuron Organisation. *Cambridge University Press*, (1948).
12. LISSAK, K. and GRAST YAN, E. : Comm. of the 1st Int. Congr. of Neurological Sciences, Bruxells, 1957. *EEG Clin Neurophsiol*, 1957, **93**.
13. McCULLOCH, W. S. and PITTS, W. : A logical calculus of ideas immanent in nervous activity. *Bull. Maths. and Biophys.*, 1943, **5**, 115.
14. McCULLOCH, W. S. : Agathe Tyche - of nervous nets the lucky reckoners. Session 4A, Paper 3.
15. MORELL, F. and JASPER, H. H. : *EEG. Clin. Neurophysiol.*, 1956, **8**, 201.
16. PAYLOV, I. P. : Lectures on conditioned reflexed. *Internat. Publishers New York*, (1928).
17. PAVLOV, I. P. : Wednesday Conferences. Russian Edition. *AMN. Moscow*, (1952).
18. PENFIELD, W. and MILNER, B. : *Am. J. of Neurol. and Psych.* (1958)
19. UTTLEY, A.M. : Conditional Probability computing in a nervous system. Session 1, Paper 5.
20. WALTER, W. G. : A machine that learns. *Scientific American*, 1951, **2**, 158.
21. WALTER, W. G. : The living brain. *Norton and Company, New York*, (1953).
22. YOUNG, J. Z. : Essays presented to E.S. Goodrich Evolution. *Oxford*,(1938)

THE CONDITIONAL PROBABILITY COMPUTER

by

DR. A. M. ANDREW

THE Conditional Probability Computer constitutes a versatile "learning machine". It simulates certain features of animal learning and embodies principles which are important in the development of adaptive control mechanisms for industrial applications. A brief account of the operation of the computer is given below, followed by descriptions of the particular applications demonstrated.

Conditional Probability Computer

The computer shown has five input channels labelled j,k,l,m,n, but computers with any number of input channels are possible in principle. It consists of 31 similar units which count the number of occurrences of the patterns of input activity presented in these channels. When some statistical data has been accumulated, and when one or a group of the input channels is activated, the computer determines, on the basis of its stored statistical information, whether this group is usually accompanied by activity in another channel or channels. If it is, the computer makes an *inference* of activity in the other channels (shown by *red* illumination of the panel above the computer). The precise conditions for an inference are described below.

The 31 units are essentially counters; five of them count the numbers of occurrences of activity in the respective input channels, while a further ten units count the numbers of occurrences of simultaneous activity in each of the ten possible pairs of input channels, and so on for groups of 3, 4 and 5 channels. The counts are represented in the units by the amount of charge on a capacitor, which is deliberately made to leak towards the level corresponding to zero count. The introduction of leakage ensures that the inferences made by the computer are governed more by recent events than by less recent events.

The conditional probability of activity in the l channel, given that there is activity in the j and k channels, is given by:

$$p_{jk}(l) = \frac{\text{count stored in } (jkl) \text{ unit}}{\text{count stored in } (jk) \text{ unit}}$$

Whenever the j and k inputs of the computer are activated simultaneously, it computes the quantities $p_{jk}(l)$, $p_{jk}(m)$ and $p_{jk}(n)$. If any of these exceeds a predetermined threshold value, an inference is made of activity in the l, m or n channel. The operation is similar when other groups of input channels are activated.

A SIMPLE COMPUTER FOR DEMONSTRATING BEHAVIOUR

by

DR. W. ROSS ASHBY

SINCE any imaginable behaviour of a system can be represented as a sequence of transitions from state to state, any behaviour can be represented isomorphically on a machine if the machine is designed or programmed to perform the appropriate transitions.

The demonstrated machine has a set of states (lamps lit or unlit) and a set of input states (switches open or closed.) When the handle is given one turn (representing the passage of one unit of time) it will change to some state that is a determinate function of what state it is at now and of what state is (at the moment) on the input. Repeated turning of the handle will thus generate a sequence of states, i.e. a trajectory, or line of behaviour.

Abstractly, it may thus be described as a machine that, with a set I of input states, and a set S of system states, can be programmed to give (within limitations of size only) any desired mapping of I x S into S. *Which* mapping is to be used is controlled by a plug-board. A particular setting on the plug-board makes it (functionally) a particular machine with input, with a particular way of behaving and of responding to stimuli.

Its chief use is the purely illustrative one of demonstrating how various forms of behaviour that are well known in the biological world appear after they have been analysed in accordance with the logic of mechanism and behaviour. It is especially suitable as a teaching device for making clear the principles of mechanism and of Black Box theory *(ref. 1)*.

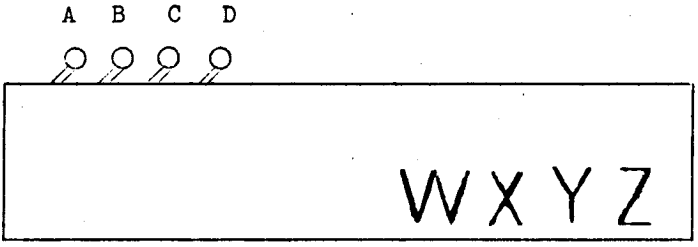
Over forty well known pieces of biological behaviour (often simplified for reasons of size) have already been programmed on to it. They include:-

Simple reflex with sustained response (e.g. the pupillary.)
" " " transient " (e.g. the knee-jerk.)
" " " latent period.
Accumulation of drive.
Displacement activity.
Response only to a pattern.
" " " patterns of particular type.
" " " a sequence of patterns.
Conflict leading to oscillation.

Conflict leading to compromise.
" " " "catatonia", with protection and cure.

DEMONSTRATION

The demonstration set up for the Symposium shows the last named. The behaviour may be thought of as analogous, in animals and man, to certain reactions evoked by situations or stimuli that arouse conflict between two ways of behaving.



The system's resting state is with lamp X lit. If switch A is closed, the lighting moves over to W (as if drawn by A.) If switch C is closed, the lighting moves over to Y. Repetition of A and C shows that the system's responses are regular and reproducible.

To B and D it gives no apparent response.

If now A and C are put down together (so that if the system were living we might say that the light was in a conflict about whether to go to the right or left) then the system responds by generalised extinction. No further stimulation by A or C (even if applied singly in the normal manner) can elicit any further response from it. Application of B does nothing to change the situation.

Closing D, however, has a "curative" effect, for subsequently the original behaviour is restored.

B, however, is not wholly without effect; for if the conflictful situation (of A and C being presented simultaneously) is combined with the application of B, the extinction does not occur.

Thus the behaviour shows the features that:-

(1) The effect of applying stimuli A and C simultaneously need bear no natural or simple relation to their effects when applied separately.
(2) A transient event (the simultaneous application) may have lasting effects.
(3) Other actions on the system (closing B or D) may have a "preventive" or "curative" action. The one action does not imply the other.
(4) A variety of other deductions could be made.

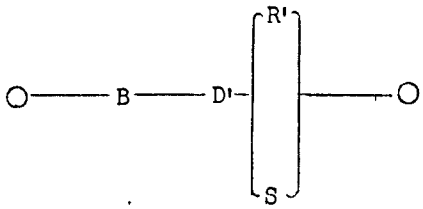
EQUIPMENT

The machine uses relays, whose power comes through their own contacts so as to make them self-locking.

During one cycle of the handle, the first quadrant makes contacts so that the states of W, X, Y, and Z are copied on to other relays P, Q, R, S. W, X, Y, Z are then cleared. Then power is applied so that (e.g.) W is energised if and only if a way exists through the network on the plug-board, which tests various series and parallel combinations through the contacts of A, B, C, D, P, Q, R, and S. Thus if W is to become energised if and only if:

(1) B is energised,
and(2) D is not energised,
and(3) either R is not energised or S is energised,

then the plug-board would use the net



(where a prime indicates a break-contact.) Similarly, three other networks control X, Y, and Z.

Once the behaviour is clearly specified in a canonical representation, the arrangement of the plug-board can be found by a purely formal process (i.e. calling for no further insight.)

REFERENCE

1. ASHBY, W. ROSS. An introduction to cybernetics. *Chapman and Hall, London*, (1956).

AUTOMATIC PATTERN RECOGNITION

by

DR. W. K. TAYLOR

The demonstration illustrates the method of constructing pattern recognition machines described in Paper 5 of Session 4B. To obtain the highest character recognition speeds that the networks are capable of it is necessary to use a parallel input system such as a 9 x 9 matrix of photomultipliers. The speed can then be up to 10^6 characters per second in a practical system, since it is only limited by the finite rise time imposed by stray capacitance. Photomultipliers are expensive, however, and for the purpose of demonstration a similar set of signals are obtained in approximately 1/50 second by arranging for the single output of an electromechanical scanning system to charge a matrix of capacitors to voltages that are proportional to the light flux falling within corresponding elements of the image. The characters to be recognised are typed by a standard typewriter and illuminated by a spotlight. A small lens projects an image of the character onto a perforated scanning drum and the light passing through the holes is converted into a voltage by a single photomultiplier. This voltage is connected to the matrix of capacitors through a synchronous switch and after one revolution of the drum the capacitors are charged to the appropriate voltages which are called x-signals in the paper.

In designing the demonstration machine a letter A was typed onto the paper and adjusted to be within the field of the scanning system. The x-signals were then measured with a voltmeter and the set of signals that gave the largest value of y in equation (1) was easily determined by combining them in order of magnitude. A y-signal network was then constructed as shown in *Fig. 3* by connecting resistors to the bus-bars that correspond to the required set of x-signals. It should be noted that this procedure could have been repeated with the letter A in a set of sample positions within the field of the lens, which is approximately 20% larger than the letter. The number of samples required to give recognition of the A in any position is not large since the x-signals do not change appreciably until the letter moves through a distance of one matrix element. In *Fig. 1a*, for example, the A can only occupy approximately

10 discriminable positions and by constructing this number of y-networks the machine would operate correctly with the A in any position. The extra expense involved in having a number of y-networks for each character is negligible since the resistor networks can be produced very cheaply in a printed circuit form.

In the demonstration apparatus there is only one y-network for each character and in consequence the alignment has to be preserved to within approximately half a matrix element. The mechanical accuracy of the typewriter is within this limit.

By repeating the design procedure for B, C and D the apparatus reached the stage shown in the demonstration. Unfortunately time did not permit the construction of y-networks for the entire alphabet but the 19" x 30" x 10" cabinet contained ample space for them, in addition to the 26 output stages and amplitude filters.

For applications such as computer read-in and letter or cheque sorting the only output required is the closure of a different electronic gate or switch for each character or pattern class. In the demonstration model the output switches select recordings of the letter names from magnetic tape. Thus as each letter appears under the lens it's name is reproduced through a loudspeaker. By extending the size of the input system n-times in the direction of writing it would be possible to accommodate words of up to n-letters and hence to produce a realistic reading machine. The speed would be limited by the rate at which recordings of the words could be selected from the store and a special parallel recording system giving rapid access would be desirable.

LIBRARY RETRIEVAL*

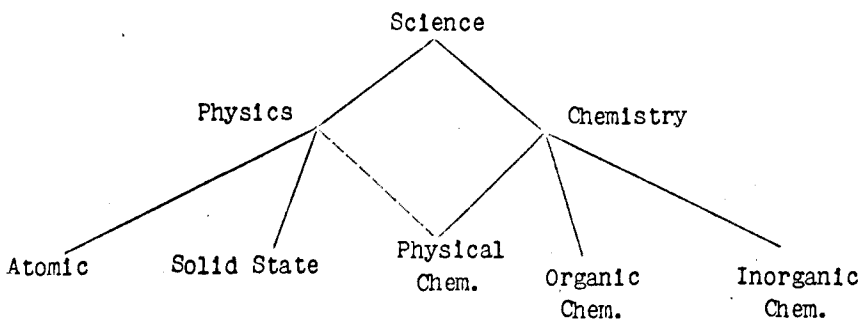
by

S. WHELAN

1. INTRODUCTION

The object of a retrieval system is to organize the indexes of a collection of documents in such a way that any documents requested can be found algorithonically.

Undoubtedly the simplest form of retrieval system is the ordinary alphabetical index. Where small collections of documents are involved, such a scheme works moderately well - most office filing systems are of this kind. However, such a method is of little use when dealing with large collections (say 50,000 documents), for then each letter of the alphabet will have coded under it a large number of documents and to retrieve any one document from the sub-collection associated with any one alphabetical letter will need some further retrieval system and, hence, some modification of the original alphabetical index is necessary. As the library grows (and libraries do!) the modifications super-imposed on the original (relatively simple) system will be such as to complicate the system and lead to more sophisticated retrieval methods such as the Universal Decimal System (UDC) for example. Despite apparently superficial differences, all such methods have one thing in common. They treat the subject matter of the library (field of knowledge) as being capable of subdivision into a tree-like structure; thus, e.g.:-

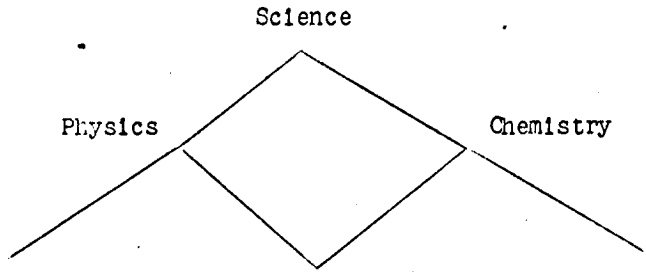


* Reprinted from R.R.E. Journal, Oct. 1958.

Such a subdivision is essentially arbitrary and becomes more so as the library grows. After only two subdivisions we are in difficulty straight away. How, for example, are we to classify a document on Physical Chemistry? Do we classify it under Physics or under Chemistry? Does it deal with Physics in Chemistry or Chemistry in Physics? At the very start, a decision of some sort is forced upon us. This would not be too bad if, having made the decision according to some principle or other, we could be certain that when a similar decision is to be made in the future, it will be made in conformity with the same principle. Unfortunately there is no procedure which can guarantee such conformity. Indeed there cannot be, for it would be impossible for any principle or procedure to take account of all the consequences inherent in arbitrariness.

Various subterfuges are introduced to overcome such difficulties but they end by making the system they were designed to correct unwieldy and, in some cases, unmanageable. The net result is that, sooner or later, such systems fail to retrieve wanted information and, very often, a high percentage of the library's documents are effectively lost. As far back as 1945, Dr. Vannevar Bush *(ref. 1)*, when reflecting on this matter remarked, "Even the modern great library is not generally consulted..... our ineptitude at getting at the record is largely caused by the artificiality of systems of indexing".

The question arises, therefore, as to whether there is any other structure which will enable us to code Physical Chemistry under both Physics and Chemistry, which is logically where it ought to be coded instead of under either. The answer is that there is such a structure - a lattice, thus:-



The notion that library language is a lattice occurs in the writings of Fairthorne *(ref. 2)* and Mooers *(ref. 4)* and the researches of the Cambridge Language Research Unit have established that a thesaurus is also a lattice, where thesaurus is taken to mean the grouping of words of similar or related meaning into notional families after the manner of Roget. In other words a thesauric or lattice classification is not arbitrary.

2. THESAURUS

By 1955 RRE felt that Retrieval Research for the most part was doing little more than invent systems which, while certainly an improvement on existing systems, did not materially differ from them and, in any case, were not sufficiently fundamental. It was this consideration together with the intrinsic merits of the scheme which decided RRE to adopt what is essentially a thesaurus approach and to embark on a pilot experiment designed to test this approach. The RRE scheme consists in choosing a limited number of basic or elementary terms (100-200, say) which, both singly and in combination, cover the subject matter of the library of documents. These terms should be as fundamental to the library as possible - indeed its ultimate constituents. They need not even be dictionary words, but it follows that they will be as exclusive as possible. The exclusion property follows from the lattice property of the thesaurus but it can also be argued from first principles. For, if an additional term is added to a list of thesaurus heads and if this additional term can be accounted for by a combination of one or more existing heads then, clearly, it is not a head. Choosing these heads (terms) is not easy; nor is it clear on what general principle it should proceed - still less is it clear whether the process could be mechanised by, e.g., machine searching of dictionaries. Obviously synonyms and near synonyms would be grouped in the same head, as would also words of cognate meaning such as - e.g. - "export" and "import". In this way we avoid the risk of non-retrieval of, e.g., a document on "Imports to Y from X" when the enquirer wants documents on "Exports from X to Y". We can always choose heads to ensure that the system discriminates to any extent we wish and the RRE scheme further ensures that the frequencies of recurrence of the heads are contained within certain limits (referred to as bandwidth). Clearly "electro-magnetics" is far too general a head for a library dealing with electromagnetics.

Furthermore the thesaurus must be constructed before embarking on the experiment. Indeed the system would not be thesauric if the terms (heads) are allocated after reading documents. This latter scheme would be more like Uniterm.

The following more or less random sample of terms used in the RRE experiment shows the scheme to be a thesauric one:-

No.	*Head*	*(Synonyms, Cognate words etc.)*
2	Add	(gain, superimpose, sum, application, join, towards)
10	Calculate	(compute, analog, digital, count, enumerate, numerical, determine, estimate, value, error)
28	Generate	(excitation, construct, make, produce, prepare, design)
37	Macro	(large, excessive, increase, amplify, wide)
73	Star	(solar flares, prominences, eclipse, meteors, sun)

Should the library extend in some unforeseen way either by storing books on some completely new subject or books on a subject not new, but which the library did not hitherto deal with, then it is always possible to add to the list of heads. The RRE scheme uses 75 heads. In practice, assuming one has a fair knowledge of the library to be coded, it is not difficult to discover the first 20-30 heads. To arrive at more is usually a difficult procedure. Another possible way is to make a library lattice, which is what the members of Cambridge Language Research Unit have done. Either way the task is not easy.

When the list of terms is complete they are then placed in some convenient order (RRE uses the alphabetical order). It is to be remembered that there is no essential significance - beyond one of mere convenience - attached to this ordering of the terms amongst themselves. When ordered, the terms are now numbered from zero upwards. This list of numbers (heads, terms) is now used to code the documents for storage and subsequent retrieval.

3. CODING, STORAGE AND RETRIEVAL

The document to be stored is given an accession number, read and abstracted.

(a) *Coding.* To facilitate the operation of the scheme, (even by persons unacquainted with its basic principles), an alphabetic dictionary of terms which occur in the reports and in the library requests has been compiled. These terms give reference to as many of the listed thesaurus heads as is necessary. Consequently, when a report is abstracted, the numbers associated with the heads which cover its subject matter are noted. This may be done either by reference to the list of heads or, more easily, by reference to the dictionary.

(b) *Storage.* Metal plates (about 14in. square), capable of being punched with 10,000 holes each, are used for storage. There are as many plates as there are thesaurus heads and the coordinate position of a hole in a plate represents the accession number of the document being stored. If, for example, the document to be stored has an accession number 509 and the heads number 5, 17, 24, 36, 61, 83 cover its subject matter, then the plates corresponding to heads 5, 17, 24, 36, 61, 83 have each a hole punched in the 509th position.

(c) *Retrieval.* To make retrieval easy and speedy all the plates are stored together and each plate has a tag on it which is easily visible and identifiable and which bears the number of the thesaurus head to which it corresponds. A request for a document is broken down into the heads which cover its subject matter (again, either by using the list of heads or by means of the dictionary), and the plates (heads) which apply to the document in question are then pivoted about one end, thus bringing them

simultaneously in register in front of a parallel beam of light; clearly the document having the hole (and hence the accession number) common to all the plates is the document sought after. More sophisticated means for reading the coordinate positions of holes easily suggest themselves.

4. EXPERIMENTAL RESULTS

A first experiment on a very small (randomly chosen) document-sample (110 reports) gave 100% success and encouraged us to increase the sample to 1000 documents. These 1000 documents consisted of documents currently received at the RRE library. The fact of currency eliminated bias in choice and so the sample can be regarded as sufficiently random.

Several lists of questions were prepared to test the scheme. The lists were prepared in the following way. A list of the titles of the documents in the sample was handed to members of the library staff who were asked to compile requests similar to the requests the library usually receives. The requests were to have no more vagueness or precision than the usual library request. Typical questions asked were on such subjects as: Butterfly Circuits, Brightness of the Atmosphere, Properties of Oxide Cathodes, etc. (see Appendix 1, which gives a sample extract from the experimental results).

In all cases where the request was moderately specific the document in question was retrieved without any unwanted or irrelevant material. As the precision of a request decreases, giving way to vagueness, then the document or documents in question are still retrieved, but so also are other documents in descending scales of relevance corresponding to the increasing scale of vagueness. This, of course, is what one would expect a thesaurus retrieval system to do and it becomes more clear when one regards the thesaurus as a lattice. For a thesaurus system will always retrieve something and this something will be what is most relevant to the request *(ref. 3)*. The interesting result is that the relevant material is always retrieved. In the RRE scheme the worst proportion of relevant to less relevant materials was found to be of the order 1:3 and this only in cases of extremely vague requests. In no case did the system fail to retrieve the wanted material. While perfection is not claimed for the RRE thesaurus, it gives an amazing degree of success (see Appendix 2).

5. CONCLUSIONS

The research makes clear that the thesaurus approach to information retrieval is the most promising approach made so far. The theoretical and experimental fact that it always retrieves the material most relevant to a request is the best guarantee of this approach. It means that none of the library's material is ever 'lost', as indeed a fair proportion of it is when other retrieval systems are employed.

From our studies we are convinced that any further research on retrieval will have to proceed along thesauric lines if any worthwhile results are to

be obtained. Nor is there any objection to extending the scheme to cope with the largest libraries. A larger thesaurus (lattice) will, of course, be necessary and obviously the scheme would have to employ digital computers but the rules for finding the elements of a lattice are precisely the rules which computers are designed to obey.

6. ACKNOWLEDGMENTS

Acknowledgments are due to Dr. A. M. Uttley of the National Physical Laboratory for discussions on the thesaurus heads and for having designed and constructed to retrieval equipment; to Messrs. P. M. Woodward and B. W. Hodlin, both of RRE, for many useful discussions, especially on the subject of the thesaurus heads, and to members of the RRE Library Staff, notably Miss K. Duncan and Mr. C. K. Moore, for their assistance especially in working out lists of questions.

REFERENCES

1. BUSH, VANNEVAR., *Atlantic Monthly*, 1945, **176**, 101.
2. FAIRTHORNE, R. A., Essentials for Document Retrieval. *R.A.E. Library Memorandum No. 23.* (1955).
3. MASTERMAN, M., Potentialities of a Mechanical Thesaurus. *Cambridge Language Research Unit* (1956).
4. MOOERS, C. N., Proceedings of the Third London Symposium on Information Theory. ed. C. Cherry. *Butterworth, London* (1956).

APPENDIX 1

Question	Degree of precision	No. of reports retrieved	No. of relevant reports	No. of irrelevant reports	No. of relevant reports in sample	Ratio of irrelevant to relevant reports
Luminescence of Metallic Salts	Precise	1	1	0	1	0
Post War Structure of Ferromagnetism	Very Vague	3	1	2	1	2
Butterfly Circuits	Vague	1	1	0	1	0
Technical Writing	Vague	4	1	3	1	3
Earth Conductivity	Vague	2	1	1	1	1
High Vacuum Techniques	Vague	3	1	2	1	2
Frequency Converter	Very Vague	35	9	26	9	2.88
Metrology	Very Vague	27	12	15	12	1.25
Welding Techniques	Vague	24	6	18	6	3
Potential Divider	Vague	7	2	5	2	2.5

APPENDIX 2

No.	Head	(Synonyms, Cognate words etc.)
1.	Accoustics	(tune, resonate
2.	Add	(gain, superimpose, sum, application, join, towards
3.	Air	(meteorology, climate, climatology, cloud, wind, storm, gas, sky
4.	Angle	(phase, angular, corner
5.	Anti	(opposed to, prevent, prevention, not, non-
6.	Array	(shape, pattern, matrix, arrangement, structure, lattice
7.	Attenuation	
8.	Auto	(automatic, self
9.	Bend	(reflection, refraction
10.	Calculate	(compute, analog, digital, count, enumerate, numerical, determine, estimate, value, error
11.	Change	(alter, different, alternate
12.	Characteristics	(properties, parameters
13.	Circle	(wheel, rotate, ring, annular, cyclic
14.	Component	(element, part, factor, unit
15.	Constant	(stable, stability, permanent
16.	Control	(servo, feed-back
17.	Defence	(flight, military, gunnery, attack, tactics, tactical, strategy
18.	Electric	(current, conductivity
19.	E.M. above 10 cms	(Above 3000 Mc/s))
20.	E.M. 10 cms. - 1 cm	(3000 - 30,000 Mc/s)) Electromagnetic
21.	E.M. 1 cm - 1 mm	(30,000 - 300,000 Mc/s)) Waves
22.	E.M. below 1 mm.	including IR and beyond)
23.	Equal	(equation, equate
24.	Equipment	(apparatus, instrument, set, machine, mechanical
25.	Explode	(bomb, burst, blow-up, projectile, weapon, missile
26.	Foreign	(various countries
27.	Function	(purpose, use
28.	Generate	(excitation, construct, make, produce, prepare, design
29.	Ground	(land, sea, horizon, earth, cosine
30.	Height	(elevation, sighting, perpendicular, vertical, sine
31.	Infrared	
32.	Integral	(differential
33.	Jam	(interference, RCM
34.	Limit	(finite, test, trial
35.	Line	(linear, direct, length, axis
36.	Locate	(find, position, scan, plot, search
37.	Macro	(large, excessive, increase, amplify, wide
38.	Magnetic	(magnet, magnetostriction, solenoid
39.	Material	(metals, non-metals, mass, resins

No.	Head	(Synonyms, Cognate words etc.)
40.	Measure (correct, adjust, assess, calibrate, verification	
41.	Micro (miniature, small, narrow	
42.	Movement (transport, transportability, mobile, shift, dynamics, displacement	
43.	Multi (combine, synthetic, many, cascade, mixture	
44.	Name (proper names - e.g., Geiger-Muller	
45.	Network (circuit	
46.	Operator (operation, human	
47.	Particle (nuclear, atomic, electrons, molecules	
48.	Point	
49.	Positive (improvement	
50.	Power (energy, strength, voltage, force, tension	
51.	Practical (experimental, performance, method, maintenance	
52.	Probability (statistics, random, chance, noise	
53.	Propagate (mode	
54.	Receive	
55.	Record (photography, report, display, data, diagram, table, list	
56.	Reverse (loop, response	
57.	Sense (detect, detector, discriminate, sensitivity, indicate	
58.	Sequence (series, iterate, repeat, group	
59.	Solid (cube, crystal	
60.	Square (area, surface, mean, field, plane	
61.	Store	
62.	Subtract (filter, loss, corrode, negative	
63.	Tele (distance, range	
64.	Temperature (heat, cold, hot, freeze, melt, boil, icing	
65.	Theoretical (study, analysis	
66.	Time, (clocks, interval	
67.	Transmit (tell, inform, radiate, radiation, feed, warm, emit	
68.	Valve (anode, cathode, C.R. tube	
69.	Vehicle	
70.	Vision (see, visual, optics, light, observation	
71.	Wave (wavelength, frequency, spectrum, ripple, band, waveband	
72.	Liquid	
73.	Star (solar flares, prominences, eclipse, meteors, sun	
74.	Ratio (relation)	

APPENDICES

APPENDIX I

Attendance List

*An asterisk indicates either that the person was a part-time deputy for a delegate, or that their attendance was restricted to particular Sessions by accomodation difficulties.

*ALDRIDGE-COX, Mr. D.
Tube Investments Technological Centre,
The Airport,
Walsall, Staffs.

ALEKSANDROV, DR. M. S.
Academy of Sciences,
Moscow, USSR.

ALLANSON, MR. J. T.
Dept. of Electrical Engineering,
The University,
Birmingham, 15.

ALLWRIGHT, MR. E. A.
British Thomson-Houston Co. Ltd.,
New Parks Boulevard,
Leicester.

ANGYAN, DR. A. J.
University Medical School,
Budapest, Hungary.

*APPLETON, MR. W. J.
H.M. Treasury,
Gt. George Street,
London, S.W.1.

ASHBY, DR. W. ROSS.
Barnwood House Hospital,
Gloucester.

ASTON, MR. B. R.
I.C.T. Ltd.,
36, Upper Brook Street,
London, W.1.

BACKUS, MR. J. W.
I.B.M. Corporation,
590, Madison Avenue,
New York 22, U.S.A.

BAILEY, MR. C. E. G.
The Solartron Electronic Group Ltd.,
Goodwyns Place,
Tower Hill, Dorking, Surrey.

BANE, MR. W. T.
SHAPE Air Defense Technical Centre,
P.O. Box 174,
The Hague, Holland.

BAR-HILLEL, PROF. Y.
The Hebrew University,
Jerusalem.

BARLOW, DR. H. B.
Physiology Dept.,
University of Cambridge,
Cambridge.

*BARRON, DR. D. W.
University Mathematical Laboratory,
Corn Exchange Street,
Cambridge.

BARTLETT, SIR FREDERICK
Applied Psychology
Research Unit,
15, Chaucer Road,
Cambridge.

BATES, DR. J. A. V.
Medical Research Council,
National Hospital,
Queen Square,
London, W.C.1.

BEER, MR. S.
United Steel Cos. Ltd.,
Cybor House,
1, Tapton House Road,
Sheffield, 10.

BENJAMIN, MR. R.
Admiralty Signal &
Radar Establishment,
Portsdown, Portsmouth.

BEURLE, DR. R. L.
English Electric Valve Co. Ltd.,
Waterhouse Lane,
Chelmsford, Essex.

BIELECHI, MR. C.
Interpreter

BIERS, MR. J. A.
Olivetti Co.,
Via Porpora 12, Rome.

BIRAM, MISS M.
Atomic Energy Authority,
Risley, Warrington,
Lancs.

BLACHMAN, MR. N. M.
Office of Naval Research,
Keysign House,
429, Oxford Street,
London, W.1.

BLACK, DR. G.
U.K. Atomic Energy
Authority,
Risley, Warrington,
Lancs.

BLAIR, MR. C. R.
536, Beacon Road,
Silver Spring,
Maryland, U.S.A.

BOURICIUS, DR. W. G.
I.B.M. Research Center,
Yorktown Heights,
New York, U.S.A.

BRANDWOOD, DR. L.
Dept. of Classics,
The University,
Manchester, 13.

BRAY, DR. J. W.
I.C.I. Ltd.,
Wilton Works,
Middlesborough.

BREWER, MR. R. C.
Imperial College,
South Kensington,
London, S.W.7.

BRIX, MR. V. H.
Interpreter

BROCKBANK, MR. A. J.
Glaxo Laboratories Ltd.,
Greenford Road,
Greenford, Middx.

BROOKER, MR. R. A.
Computing Laboratory,
The University,
Manchester, 13.

BROWN, DR. J.
Birkbeck College,
Malet Street,
London, W.C.1.

BRUCE, MR. D. J.
Dept., of Psychology,
The University,
Reading, Berks.

*BUCKINGHAM, DR. R. A.
University of London
Computing Laboratory,
44, Gordon Square,
London, W.C.1.

BUCKINGHAM, DR. M. J.
Duke University,
Durham, N.C., U.S.A.

CAMPBELL, DR. F. W.
Physiology Department,
Cambridge University,
Downing Place, Cambridge.

CARPENTER, MR. H. G.
I.B.M. British Laboratories,
101, Wigmore Street,
London, W.1.

CHAPMAN, MR. B. L. M.
Department of Psychology,
University of Bristol,
27, Belgrave Road,
Bristol, 8.

CHERRY, PROF. C.
Imperial College of Science,
London, S.W.7.

CLARKE, MR. S. L. H.
Elliott Bros. (London) Ltd.,
Elstree Way, Borehamwood,
Herts.

CLEAVE DR. J. P.
Computation Laboratory,
The University,
Southampton.

COALES, MR. J. F.
Engineering Laboratory,
Trumpington Street,
Cambridge.

COMET, MR. S.
Matematikmaskinnamunden,
Box 6131,
Stockholm, Sweden.

COOPER, DR. F. S.
Haskins Laboratories,
305, East 43rd Street,
New York 17, N.Y., U.S.A.

COPPOCK, MR. S. W.
Ministry of Supply,
Shell Mex House,
Strand, London, W.C.2.

CRAWLEY, MR. H. J.
National Research Development Corporation,
1, Tilney Street,
London, W.1.

CROSSLAND, DR. J.
St. Andrew's University,
St. Andrews, Fife.

CROSSMAN, DR. E. R. F.
Department of Psychology,
The University,
Reading, Berks.

CUSHING, MR. G. W.
National Cash Register
Co. Ltd.,
36, Upper Brook Street,
London, W.1.

DAVID, DR. A.
6, rue du Casino,
Aix-les-Bains, France.

DAVIS, MR. G. M.
English Electric Co. Ltd.,
Marconi House,
Strand, London, W.C.2.

DENES, MR. P.
Department of Phonetics,
University College,
Gower Street,
London, W.C.1.

DENISON, MR. S. J. M.
English Electric Co. Ltd.,
Nelson Research Laboratories,
Blackheath Lane, Stafford.

DENMARK, MR. E. J.
Ministry of Supply,
St. Giles Court,
1-13 St. Giles High Street,
London, W.C.2.

DONALDSON, MR. P. E. K.
Physiology Dept.,
The University,
Cambridge.

DOUGLAS, DR. A. S.
Electronic Computing
Laboratory,
Eldon Hall,
Woodhouse Lane,
Leeds, 2.

DUFF, MR. W.
Afra Industrial
Export Corpn.
10, Amsterdam House,
Quartz Street,
Johannesburg, S. Africa.

*DUNCAN, MR. F. G.
English Electric Co. Ltd.,
Nelson Research
Laboratories, Stafford.

EADES, MR. J. R.
Production Engineering Ltd.,
30, Waterloo Street,
Birmingham, 2.

EFRON, DR. R.
National Hospital,
Queen Square,
London, W.C.1.

ELBOURNE, MR. K. B.
I.C.T. Ltd.,
36, Upper Brook Street,
London, W.1.

ELLIOTT, MR. W. S.
I.B.M. United Kingdom Ltd.,
101, Wigmore Street,
London, W.1.

ERSHOV, DR. A. P.
Academy of Sciences,
Moscow, USSR.

ERSKINE, DR. G.
CERN European Organisation
for Nuclear Research,
Geneve 23,
Switzerland.

FOURCIN, MR. A. J.
Signals Research &
Development Establishment,
Christchurch, Hants.

FREEBODY, MR. J. W.
Automatic Data Processing
Technical Support Unit,
General Post Office,
2-12 Gresham Street,
London, E.C.2.

FREEDMAN, MR. A. L.
I.C.T. Ltd.,
12, Jowitt House,
Sish Lane,
Stevenage, Herts.

FRY, PROF. D. B.
Department of Phonetics,
University College,
Gower Street,
London, W.C.1.

GABOR, DR. D.
Imperial College,
London, S.W.7.

GARWICK, DR. J. V.
SHAPE Air Defense
Technical Center,
P.O. Box 174,
The Hague, Holland.

*GEARING, MR. M. W.
Metal Box Co. Ltd.,
37, Baker Street,
London, W.1.

GERHARD, MR. D. J.
D.S.I.R. Charles House,
5-11, Regent Street,
London, S.W.1.

GILL, DR. S.
Ferranti Ltd.,
21, Portland Place,
London, W.1.

GLADMAN, MR. J. C.
Metropolitan Vickers Co. Ltd.,
Trafford Park,
Manchester, 17.

GLENNIE, MR. A. E.
U.K. Atomic Energy
Authority,
Aldermaston, Berks.

Goblick, DR. T. J.
London House,
Guilford Street,
London, W.C.1.

*GOLDMAN-EISLER, DR. F.
Department of Phonetics,
University College,
Gower Street,
London, W.C.1.

GOLDSCHMIDT-CLERMONT, DR. Y.
CERN European Organisation
for Nuclear Research,
Geneva 23,
Switzerland.

*GOODWIN, MR. B.
Institute of Animal
Genetics,
West Mains Road,
Edinburgh, 9.

GOSDEN, MR. J. A.
Leo Computors Ltd.,
Cadby Hall,
Blythe Road,
London, W.14.

GREENALL, MR. P. D.
Department of Scientific
& Industrial Research,
Charles House,
5-11 Regent Street,
London, S.W.1.

GREGORY, MR. R. L.
Psychological Laboratory,
Downing Street,
Cambridge.

GRIMSDALE, DR. R. L.
Electrical Engineering
Laboratories,
The University,
Manchester 13.

GUREWITSCH, MR. A. M.
General Electric
Company Research
Laboratory,
Pelikanstrasse 37,
Zurich, Switzerland.

GUTTRIDGE, MAJOR E. J.
I.C.T. Ltd.,
Research Division,
Godstone Road,
Whyteleaf, Surrey.

HALLIDAY, DR. A. M.
Neurological Research Unit,
National Hospital,
Queen Square,
London, W.C.1.

HALSBURY, THE RT. HON. EARL OF
National Research
Development Corporation,
1, Tilney Street,
London, W.1.

HAMMOND, MR. P. H.
Royal Radar Establishment,
Malvern, Worcs.

HAMMOND, DR. W. H.
Criminal Research Unit,
Home Office,
Whitehall,
London, S.W.1.

*HANCOCK, MR. C.
H.M. Treasury,
Gt. George Street,
London, S.W.1.

HASKINS, DR. C. P.
Carnegie Institution,
1530 "P" Street NW.
Washington, D.C., U.S.A.

*HAWKINS, MR. E. N.
English Electric Co. Ltd.,
Nelson Research Laboratories,
Stafford.

HEGGIE, MR. W.
Establishment and
Organisation Division,
Home Office,
Whitehall,
London, S.W.1.

HOFFMAN, MR. W.
International Business
Machines Corpn.
Zurichstrasse 108,
Zurich, Switzerland.

HOLLAND-MARTIN, MR. G. G.
I.C.T. Ltd.,
Gunnels Wood Road,
Stevenage, Herts.

HOLLINGDALE, DR. S. H.
Royal Aircraft Establishment,
Farnborough, Hants.

HOPPER, DR. GRACE, M.
Remington Road Univac,
19th Street and
W. Allegheny Avenue,
Philadelphia, 29.
Pa., U.S.A.

HOWELL, MR. T. C.
The Plessey Co. Ltd.,
Ilford, Essex.

HUGGINS, MR. P.
T.I. Technological Dept.,
The Airport,
Walsall, Staffs.

*HUNTER, MR. D. G. N.
Standard Telecommunications
Laboratories Ltd.,
Progress Way,
Great Cambridge Road,
Enfield, Middx.

IL'IN, PROF. V. A.
Academy of Sciences,
Moscow, USSR.

INGHAM, MR. W. E.
E.M.I. Electronics Ltd.,
Hayes, Middx.

JEEVES, DR. M. A.
Department of Psychology,
The University,
Leeds, 2.

JEFFERY, CAPT. S.
Rome Air Development Centre,
Air Research & Development
Command,
U.S. Air Force,
New York, U.S.A.

JENKINS, DR. D. P.
Royal Radar Establishment,
Malvern, Worcs.

*JENKINS, MR. L. E.
H.M. Treasury,
Gt. George Street,
London, S.W.1.

JONES, DR. F. E.
Mullards Ltd.,
Mullard House,
Torrington Place,
London, W.C.1.

*JORDAN, MR. G. H. S.
H.M. Treasury,
Gt. George Street,
London, S.W.1.

KERR, MR. D.
Ultra Electric Ltd.,
Western Avenue,
Acton,
London, W.3.

KITZ, MR. N.
Bell Punch Co. Ltd.,
The Island, Uxbridge,
Middx.

KOMARITSKY, DR. N. R.
Academy of Sciences,
Moscow, USSR.

KOWARSKI, DR. L.
Director, STS Division,
CERN European Organisation
for Nuclear Research,
Geneva 23, Switzerland.

KUN, Mr. E. R.
Laan van Clingendael 13,
The Hague, Holland.

LADEFOGED, MR. P.
Department of Phonetics,
The University, Edinburgh.

LAWRENCE, MR. W.
Signals Research and
Development Establishment,
Christchurch, Hants.

LUBBOCK, MR. J. K.
Engineering Laboratory,
Trumpington Street,
Cambridge.

McCARTHY, DR. J.
Room 26-261, Massachusetts
Institute of Technology,
Cambridge 39,
Mass., U.S.A.

McCULLOCH, DR. W. S.
Room 26-027, R.L.E.,
Massachusetts Institute of
Technology, Cambridge, 39,
Mass., U.S.A.

McDERMOTT, MR. T. J.
Neurophsychiatric Research
Unit,
Whitchurch Hospital,
Cardiff.

McDONNELL, MR. D.
Vickers Group Research
Establishment,
Weybridge, Surrey.

MacKAY, DR. D. M.
King's College,
Strand,
London, W.C.2.

MacMILLAN, PROF. R. H.
Department of Engineering,
University College,
Singleton Park,
Swansea.

MALING, MR. K.
Massachusetts Institute
of Technology,
Cambridge 39, Mass., U.S.A.

MARKS, MR. C. P.
Ministry of Supply,
Leatherhead Road,
Chessington, Surrey.

MARILL, DR. T.
Bolt, Beranek & Newman Inc.,
50, Moulton Street,
Cambridge 38,
Mass., U.S.A.

MARSHALL, CAPT. P. R.
Government Communications
Headquarters,
Cheltenham, Gloucs.

MASSEY, MR. R. G.
British Iron & Steel
Research Association,
11, Park Lane,
London, W.1.

MEHL, DR. M. L.
Ecole Nationale
D'Administration,
56, Rue des St. Peres,
Paris, France.

MERRIDITH, DR. J. F.
Smiths Aircraft
Instruments Ltd.,
Bishops Cleave,
Nr. Cheltenham, Glos.

MERRIMAN, MR. J. H. H.
H.M. Treasury,
Great George Street,
London, S.W.1.

MERRY, MR. I. W.
I.B.M. British Laboratories,
101, Wigmore Street,
London, W.1.

MINSKY, DR. M. L.
Massachusetts Institute of
Technology,
Lexington 73,
Mass., U.S.A.

MORGANTI, MR. I.
Olivetti Co.,
Via Porpora 12,
Rome.

MORTON, MR. J.
Department of Psychology,
University of Reading,
Reading, Berks.

MUNTZ, MR. W.R.A.
Department of Anatomy,
University College,
Gower Street,
London, W.C.1.

*NASH, DR. F. A.
Western Hospital,
Seagrave Road,
Fulham,
London, S.W.6.

NEWMAN, MR. G. B.
Royal Marsden Hospital,
Radiotherapy Dept.,
Fulham Road,
London, S.W.3.

NICOLL, MR. G. R.
Manchester College of
Science & Technology,
Manchester, 1.

*PANNELL, MR. A.
Atomic Energy Authority,
Risley,
Warrington, Lancs.

PASK, MR. A. G. S.
The Solartron Electronic
Group Ltd.,
Goodwyns Place,
Tower Hill,
Dorking, Surrey.

PATRY, DR. J.
Reaktor AG,
c/o Escher Wyss AG,
Hardstrasse 319,
Postfach, Zurich 23.

PAULA, MR. F. C. de
Robson, Morrow & Co. Ltd.,
59, New Cavendish Street,
London, W.1.

PAYCHA, DR. F.
114, Rue Caulaincourt,
Paris 18, France.

PAYNE, DR. L. C.
Decca Radar Ltd.,
Corner House,
Albert Road,
New Malden, Surrey.

*PEARCEY DR. T.
Royal Radar Establishment,
Malvern, Worcs.

REAM, MR. N.
Battersea College of
Technology,
Battersea Park Road,
London, S.W.11.

REDFERN, MR. P.
Central Statistical Office,
Gt. George Street,
London, S.W.1.

REES, MR. N. W.
Dept., of Engineering,
University College,
Singleton Park, Swansea.

REMOND, DR. A.
Hopital Saltpetriese,
131, Boulevard Malesherbes,
Paris, 17 eme, France.

RICHARD, MR. T.
Dufourstr. 153,
Zurich 8,
Switzerland.

RICHARDS, MR. J. H.
Mullard Research Laboratories,
Cross Oak Lane,
Salfords,
Nr. Redhill, Surrey.

RICHENS, MR. R. H.
School of Agriculture,
Downing Street,
Cambridge.

RINGROSE, MR. J. W.
The Plessey Co. Ltd.,
Vicarage Lane,
Ilford, Essex.

*ROBINSON, MR. C.
English Electric Co. Ltd.,
Nelson Research Laboratories,
Stafford.

ROSENBLATT, DR. F.
Cornell University,
4455, Genesee Street,
Buffalo 21,
New York, U.S.A.

RUSSELL, MR. G.
Royal Radar Establishment,
Malvern, Worcs.

SACERDOTI, MR. I.
Olivetti Co.,
Via Porpora 12,
Rome.

SCHLAG, DR. J.
Universite de Liège,
25, Quai de Rome,
Liège, Belgium.

*SCHUHMANN, MR. W. S.
Standard Telephone
Laboratories Ltd.,
Progress Way,
Great Cambridge Road,
Enfield, Middx.

SCOTT, MR. B.
Solartron Electronic
Group Ltd.,
Solartron Works,
Thames Ditton, Surrey.

SELFRIDGE, DR. O. G.
Massachusetts Institute
of Technology,
Lexington 73,
Mass., U.S.A.

SHACKEL, MR. B.
E.M.I. Electronics Ltd.,
Victoria Road,
Feltham, Middlesex.

SHACKLETON MR. P.
Elliott Bros. (London)Ltd.,
Borehamwood, Herts.

SHAW, MR. G. L.
Air Trainers Link Ltd.,
Bicester Road,
Aylesbury, Bucks.

SHELLEY, MR. J. H.
Smiths Aircraft
Instruments Ltd.,
Bishops Cleeve,
Nr. Cheltenham.

SHERWOOD, DR. S. L.
Medical Research Council,
38, Addison Gardens,
London, W.14.

SHUEY, DR. R. L.
General Electric Co.,
Research Laboratory,
P.O. Box 1088,
Schenectady,
N.Y., U.S.A.

SMITH, MR. J. L.
Data Processing
Systems Division,
National Bureau
of Standards,
Washington 25.
D.C., U.S.A.

SOEST, PROF. VAN.
Technishe Hochschule,
Delft, Holland.

SOTSKOV, PROF. B.S.
Academy of Sciences,
Moscow, USSR.

SPETNER, DR. L. M.
John Hopkins University,
Silver Springs,
Maryland, U.S.A.

STANWORTH, DR. J. E.
British Thomson-Houston
Co. Ltd.,
Rugby.

STEVENS, MISS M.
National Bureau of
Standards,
Washington 25,
D.C., U.S.A.

STIEBER, MR. A.
Cornell University,
4455, Genesse Street,
Buffalo, 21,
New York, U.S.A.

STOCKBRIDGE, MR. H.C.W.
Ministry of Supply,
Clothing and Stores Experi-
mental Establishment,
c/o R.A.E. Farnborough,
Hants.

STRACHEY, MR. C.
National Research
Development Corporation,
1, Tilney Street,
London, W.1.

STUMPERS, DR. F. L.
Philips Research
Laboratories,
Eindhoven, Holland.

SUMNER, DR. F. H.
Electrical Engineering
Laboratories,
The University, Manchester 13.

SUTHERLAND, DR. N. S.
Institute of Experimental
Psychology,
1, South Parks Road,
Oxford.

*SUTTON, MR. G. G.
Royal Radar Establishment,
Malvern, Worcs.

SWIFT, MR. P.
Mullard Research
Laboratories,
Cross Oak Lane,
Salfords, Surrey.

*TAYLOR, MR. P.
Royal Radar Establishment,
Malvern, Worcs.

TAYLOR, DR. W. K.
Anatomy Dept.
University College,
Gower Street,
London, W.C.1.

TIMMS, DR. G.
Government Communications
Headquarters,
Cheltenham, Gloucs.

TIZARD, MR. R. H.
London School of Economics,
Houghton Street,
Aldwych, London, W.C.2.

TOCHER, DR. K. D.
United Steel Companies Ltd.,
Cybor House,
1, Tapton House Road,
Sheffield, 10.

TOOTHILL, MR. G. C.
Royal Aircraft
Establishment,
Farnborough, Hants.

TREWEEK, MR. K. H.
Royal Aircraft
Establishment,
Farnborough, Hants.

TRIER, MR. P. E.
Mullard Research
Laboratories,
Cross Oak Lane,
Salfords, Nr. Redhill,
Surrey.

TSYPKIN, PROF. Ya. Z.
USSR Academy of Sciences,
Moscow.

VAJDA, DR. S.
Admiralty Research
Laboratory,
Queens Road,
Teddington, Middx.

VARJU, DR. D.
Max-Planck-Institut
für Biologie
Spemanstrasse 34,
Tübingen, Germany.

VERBEEK, MR. L. A. M.
Technishe Hochschule,
Delft, Holland.

VOWLES, DR. D. M.
Dept., of Psychology,
The University,
Reading, Berks.

WALTER, DR. W. GREY.
Burden Neurological
Institute,
Stoke Lane,
Stapleton, Bristol.

WASON, DR. P. C.
Medical Research Council,
University College,
17, Gordon Square,
London, W.C.1.

WATSON, MR. A. J.
Psychological Laboratory,
Downing Place,
Cambridge.

*WELLS, MR. O. D.
Artorg,
Beaulieu, Hants.

WHELAN, MR. S.
Royal Radar Establishment,
Malvern, Worcs.

*WHITEHEAD, MR. E. A. N.
Tube Investments
Technological Centre,
The Airport,
Walsall, Staffs.

WHITFIELD, DR. I. C.
National Institute of
Mental Health,
St. Elizabeth's Hospital,
Washington 20,
D.C., U.S.A.

WILLIAMS, MR. G. M. E.
Park House,
Wick Road,
Egham, Surrey.

WILLIAMS, MISS T. M.
Itek Corportion,
1605, Trapelo Road,
Waltham 54,
Mass., U.S.A.

WILLIS, MR. D. W.
Decca Radar Ltd.,
Molesey Road,
Walton-on-Thames,
Surrey.

WOODWARD, MR. P. M.
Royal Radar Establishment,
Malvern,
Worcs.

WOOLNER, MISS A. D.
I.C.T. Ltd.,
Gunnels Wood Road,
Stevenage,
Herts.

YOUNG, MR. A. J.
I.C.I. Ltd.,
Bozedown House,
Whitchurch Hill,
Reading,
Berks.

*YOUNG, MR. D. A.
De Havilland Propellers Ltd.,
Hatfield,
Herts.

YOUNG, PROF. J. Z.
Anatomy Dept.,
University College,
Gower Street,
London, W.C.1.

YOVITS, DR. M. C.
Office of Naval Research,
Washington 25,
D.C., U.S.A.

Delegates from the National Physical Laboratory, Teddington, Middlesex.

SUTHERLAND, DR. G. B. B. M., Director.
ANDREW, DR. A. M.
BARBER, MR. D. L. A.
BARRELL, DR. H.
BLAKE, MR. D. V.
CLAYDEN, MR. D. O.
DAVIES, MR. D. W.
FROOME, DR. K. D.
GOODWIN, DR. E. T.
McDANIEL, MR. J.
NEWMAN, MR. E. A.
OSBORNE, MR. C. F.
PAGE, MR. L. J.
ROBERTSON, DR. H. H.
SINNOTT, MR. C. S.
STUART, DR. P. R.
UTTLEY, DR. A. M.
VICKERS, MR. T.
WILSON, MR. W.
WOODGER, MR. M.
WRIGHT, MR. M. A.

APPENDIX 2

Index to authors and contributors

References to papers are in bold type: others are contributions to discussions.

* I.C.I. Ltd. Whitchurch Hill, Reading, Berks.
† H.M. Treasury, Gt. George Street, London, S.W.1.
ø CME Division, NPL.

DS 94009/1/Wt.4203 K.10 3/59 DL 980

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*