SESSION 1

PAPER 2

OPERATIONAL ASPECTS OF INTELLECT

by

DR. D. M. MACKAY

BIOGRAPHICAL NOTE

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

OPERATIONAL ASPECTS OF INTELLECT

by

DR. D. M. MACKAY

SUMMARY

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

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

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

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

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

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

---

1. INTRODUCTION

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

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

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

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

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

2. REPRESENTATION OF INFORMATION

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

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

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

---

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

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

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

3. PROBLEM-SOLVING

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

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

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

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

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

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

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

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

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

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

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

4. THRESHOLD OF EVIDENCE

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

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

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

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

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

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

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

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

and hence the probability or frequency of spontaneous activity.

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

5. ORIENTATION TOWARDS A SOLUTION

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

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

Thus:

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

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

---

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

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

VI. *It must try a range of variations, some on a small scale with high discrimination, others on a larger scale with lower discrimination* to sense more general trends. It may well be that the familiar benefits of practice in problem-solving arise partly from its influence on the range-distribution explored by our own brain-mechanism.

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

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

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

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

6. THE ENVISAGING OF ALTERNATIVES

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

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

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

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

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

- unless and until further evidence changes the picture.

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

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

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

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

---

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

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

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

7. THE IMPORTANCE OF ANALOGUES

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

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

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

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

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

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

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

---

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

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

CONCLUSION

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

REFERENCES

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

10. MACKAY, D. M. On the Logical Indeterminacy of a Free Choice. *Proc. of XIIth Int. Congress of Philosophy, Venice* (1958).

11. POLANYI, M. Problem Solving. *Brit. J. Philosophy of Science*, 1957, 8, 89.

12. POPPER, K. R. Indeterminism in Classical and Quantum Physics. *Brit. J. Philosophy of Science*, 1950, I, 173.

13. POPPER, K. R. The propensity interpretation of the calculus of probability. Observation and Interpretation. *Colston Papers*. (1957).

INTRODUCTION TO APPENDIX

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

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

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

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

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

REFERENCES

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

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

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

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

APPENDIX*

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

by

DR. D. M. MACKAY

1. TECHNIQUES OF COMPUTATION

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

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

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

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

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

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

---

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

† i.e. 1949.

2. THE NATURE OF INFORMATION

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

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

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

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

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

Each logon will be associated with a certain amount of metrical information, analogous to the "quantity of information" defined by R. A. Fisher, *(ref. 3)* and measuring essentially the precision with which the corresponding coefficient or its analogue has been ascertained. The need for precision in scientific statements effectively quantises the action of this information also *(ref. 1)*, but this need not concern us here. The point is that the total information comprised in a scientific statement can be represented by a set of numbers associated with a finite number of degrees of freedom or logons. These numbers may be regarded as coordinates of a point, in a space of dimensionality equal to the number of logons. The *direction* of the "information vector" from the point to the origin defines the structure of the corresponding proposition. The *dimensionality of* the space defines the amount of structural information contained in the proposition (or required for its formulation). The *length* (or rather the square of the length) of the information vector defines the amount of metrical information in the proposition, and hence the degree of confidence with which it may be received.

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

3. DIVISION OF FUNCTION

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

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

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

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

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

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

4. FIELDS OF OPERATION

4.1. The relative importance of analogue and digital techniques in an analytical mechanism will depend on the field from which its information is drawn and with which it interacts. The field of mathematical computation for example, or of mathematical logic, can normally be covered in terms of wholly structural information. Even so there are many processes, such as division, in which much time could be saved by a combination of the two techniques. An analogue mechanism could provide a first approximation to an answer, enabling the digital mechanism to select the required range of operation and reach the precise conclusion more quickly than if it had to seek it systematically.

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

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

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

5. AN AUTONOMOUS ARTEFACT

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

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

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

5.3. A device of this sort can be autonomous, in that it can pursue an active, responsive, and logically - disciplined existence independent of human intervention. In an abstract sense it may be said to show purposeful activity, the purpose being always the attainment of a certain equilibrium configuration, or one of several such configurations, of incoming or present information. The meaning of those purposes in terms of the information field will of course depend on the equilibrium configurations chosen, and the same choice will therefore determine whether or not they appear "sensible" purposes to a human interlocutor.

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

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

6. THE INTRODUCTION OF INDETERMINACY

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

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

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

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

7. AUTONOMY AND INDETERMINACY

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

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

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

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

8. GENERAL CHARACTERISTICS

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

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

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

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

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

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

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

9. CONCLUSION

It might be well to conclude with a question which will certainly be asked: why should anyone want to make such a machine? It is probably not sufficient in these days to suggest that the study of these devices and their potentialities in a worthwhile pursuit of knowledge for its own sake. In the author's view however the questions raised by these possibilities are of profound importance, and are properly the province of a borderline study, - the science of knowledge itself. As a tool for such work, a device capable of the most general possible transformations of information appears to be of the first importance. Reference has already been made to the possible usefulness of such studies conducted in parallel with the study of cerebral mechanisms, both normal and disordered. In such cooperation between specialists, as Wiener's group has shown,* the benefits to be derived are assuredly not unilateral. The new study indeed would seem to require the full collaboration at least of engineers, logicians, psychologists, physiologists and physicists for its adequate development.

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

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

REFERENCE

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

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

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

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

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

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

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

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

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

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

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

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

For example, let us assume we have a computer working at a number of levels. At the lowest level we have basic inputs - n of these say, at the next level we have units which count coincidences between the various possible pairs of inputs - there will be about $n^2/2$ of these. At the next level we have counter units for the various possible triples - about $n^3/6$ of these - and so on. All told there will be 2^n counter units. By comparing these counts we can calculate the probability of having any combination of inputs given any other combination. The number of units required however is very great. If however one restricts the computer, in the first place to the two-at-a-time layer one might well find that only a few of the possible coincidences in fact occur. Indeed there might be enough low count units available that these could be extracted and used to allow for every combination of useful pair count units and inputs (which would give all likely three units). Similarly there might well prove to be enough three input coincidence counters left little used to allow for all likely four-units, and so on.

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

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

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

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

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

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

REFERENCES

1. BARTLETT, F. "Thinking".

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

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

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

1. There has been much talk about "logical jumps". Programmes can and do make "logical jumps" through the use of heuristics which abandon the certainty of algorithms. There need be nothing mysterious about the appearance of logical jumps in programs whose basis is, to begin with, pragmatic rather than logical.

2. The problem of exponential growth that arises in the model being considered could be regarded as a central problem of artificial intelligence--- something important will be contributed when we are shown a model where the structural tree does not grow at such a rate with the input complexity.

3. While I am fully in accord with what Dr. MacKay has said about analogues, I think we should all be reminded that analogues need not be "analogue" in their physical structure! One can, after all, have digital maps and the like. The concept of ability to interpolate is clearly not limited to measurement of physical quantities. In connection with the relation between a problem and an analogue I heard the term "probabilistic topology". I suggest that some new term, e.g., "heuristic connection", be used to describe the relation so that you do not have to worry whether or not there is a better mathematician in the room.

4. In regard to mutations---there is a big logical jump in assuming that the small variations necessarily have to be on a random basis. If you look at your theories carefully you will often find that the probabilistic element has been inserted only for simplicity and convenience of description. Clearly the more 'creative' people choose their ideas from better ensembles (they are too consistent to be merely lucky); it is the establishment of a good trial space and a priori weights that is important— it matters little whether the further search is random or systematic.

DR. D. M. MACKAY (in reply): Mr Williams is right about the importance of 'criticism', and he will find a good deal on the lines he suggests in some of the earlier references in my paper, where I described mechanisms which seek to have their (internally generated) working hypotheses dis-confirmed and modified by evidence. The evidence-bearing signals, however, would still function by controlling gain or bias, so that this 'critical feedback' does not represent a separate case.

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

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

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

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

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

REFERENCE

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

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

NOTE

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