(6)

NATIONAL PHYSICAL LABORATORY

TEDDINGTON, MIDDLESEX, ENGLAND

PAPER 4-13

PHYSICAL ANALOGUES TO THE GROWTH OF A CONCEPT

by

GORDON PASK

To be presented at a Symposium on
The Mechanization of Thought Processes,
which will be held at the National Physical Laboratory, Teddington, Middlesex, from 24th-27th November, 1958. The papers and the discussions are to be published by H.M.S.O. in the Proceedings of the Symposium.

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.

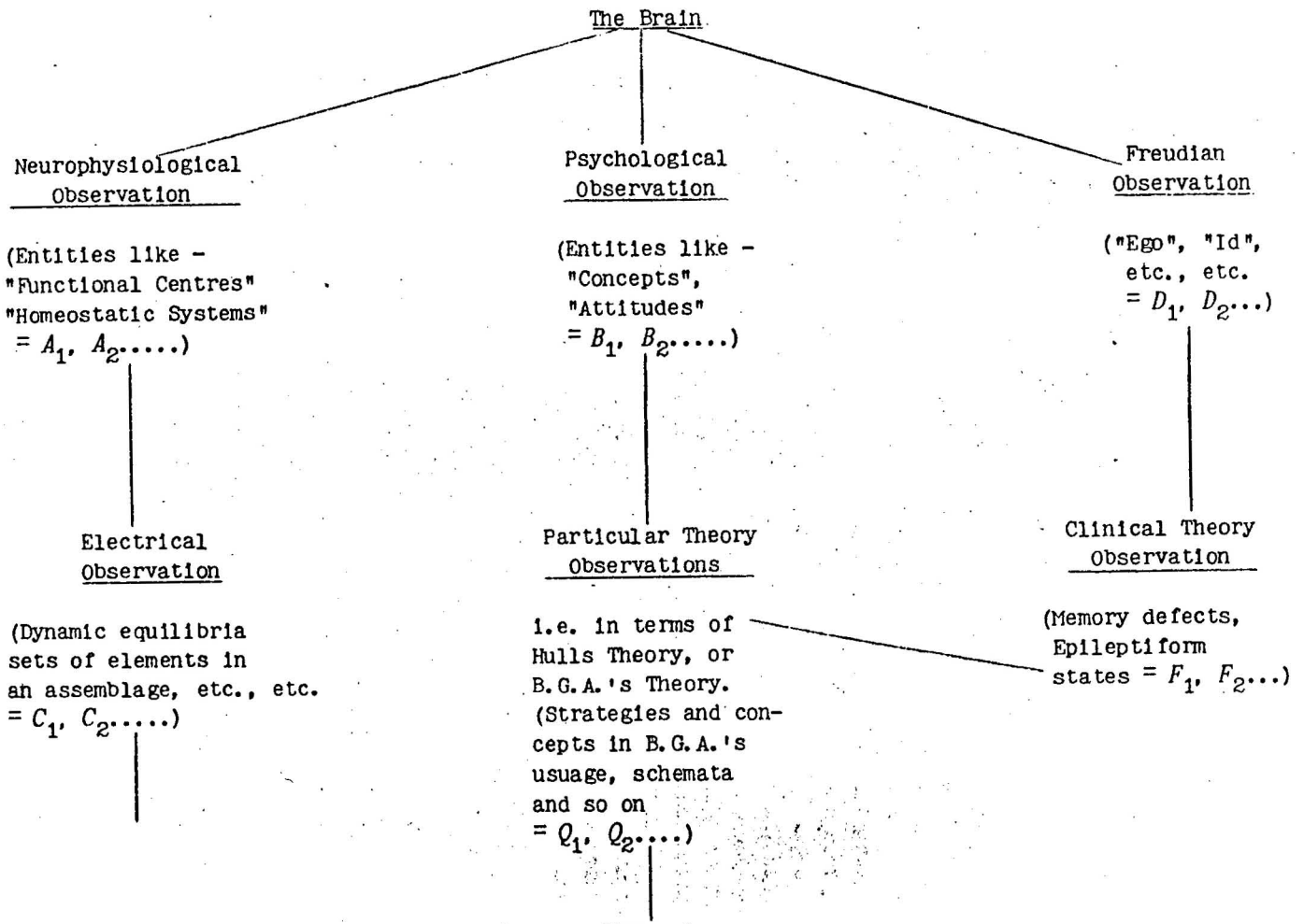
The Brain
Neurophysiological Observation
(Entities like - "Functional Centres" "Homeostatic Systems" = A_1, A_2.....)
Psychological Observation
(Entities like - "Concepts", "Attitudes" = B_1, B_2.....)
Freudian Observation
("Ego", "Id", etc., etc. = D_1, D_2...)
Electrical Observation
(Dynamic equilibria 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_i, 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_i, B_j which have been defined above. The sequence $a_i^* = (a_{u,t}, a_{v,t+1}, \ldots\ldots a_{s,t+\tau})$ at instants $t, t+1, \ldots\ldots t+\tau$. might, for example, be taken to imply A_i.

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 i.e. 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^*_{II} 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

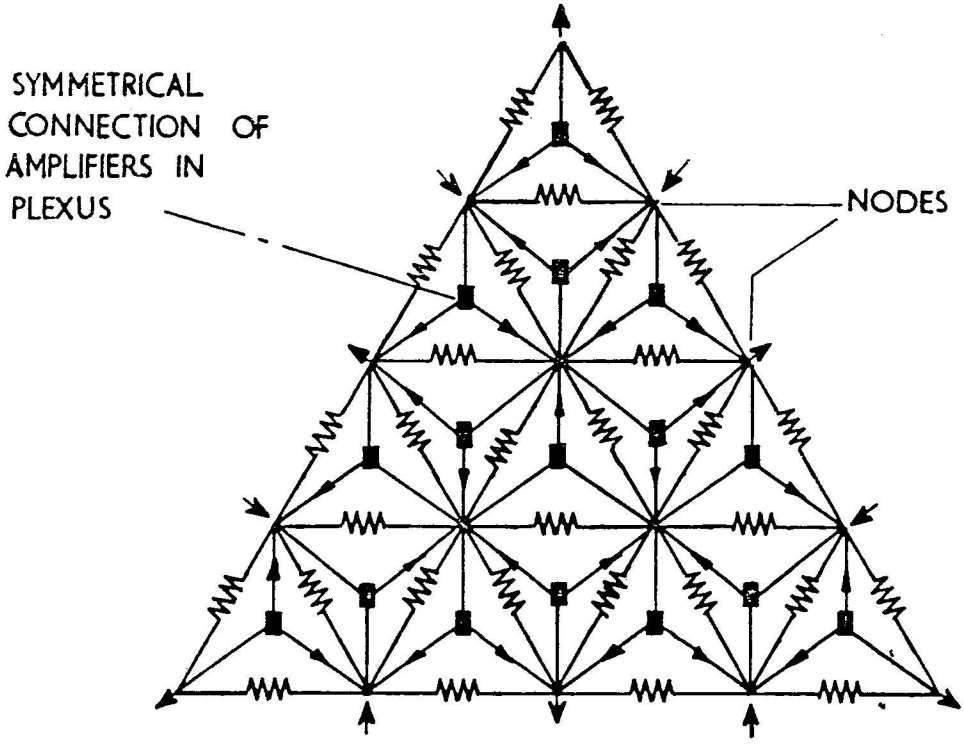

Fig. 2.

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.

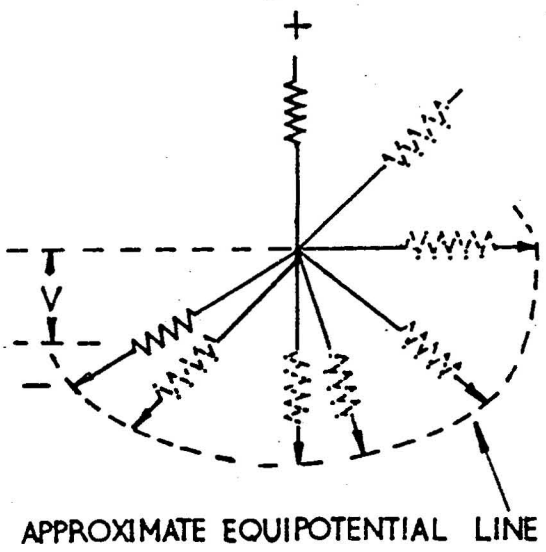

Fig.3

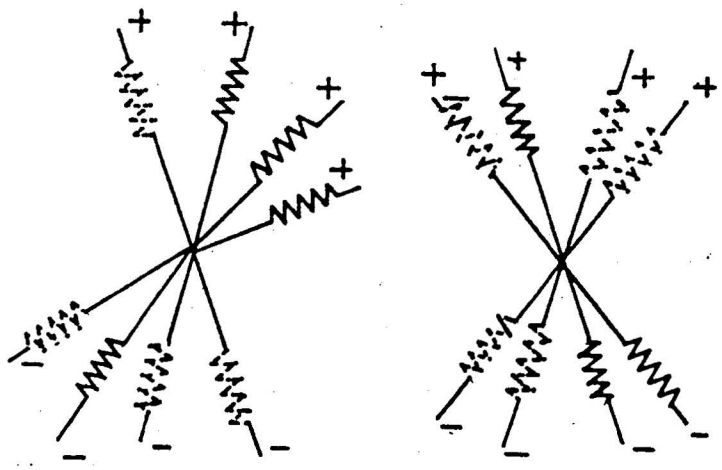

Fig.4

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).

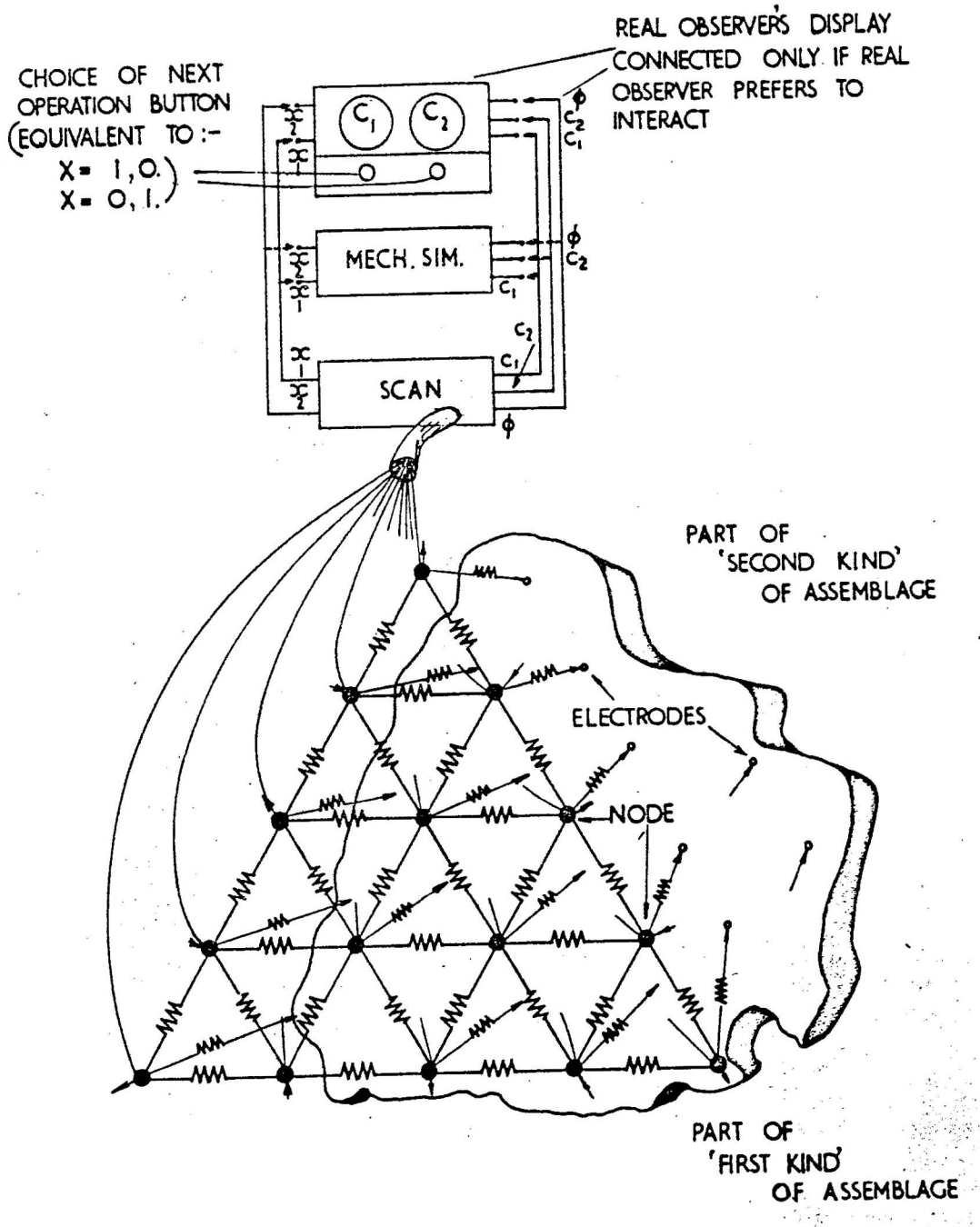
REAL OBSERVER'S DISPLAY CONNECTED ONLY IF REAL OBSERVER PREFERS TO INTERACT
CHOICE OF NEXT OPERATION BUTTON (EQUIVALENT TO :- X= 1,0. X= 0,1.)
C1
C2
MECH. SIM.
SCAN
PART OF 'SECOND KIND' OF ASSEMBLAGE
ELECTRODES
NODE
PART OF 'FIRST KIND' OF ASSEMBLAGE

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 -

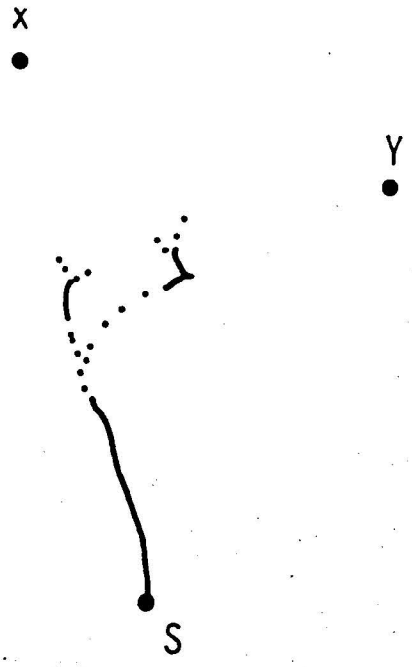

Fig.6

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

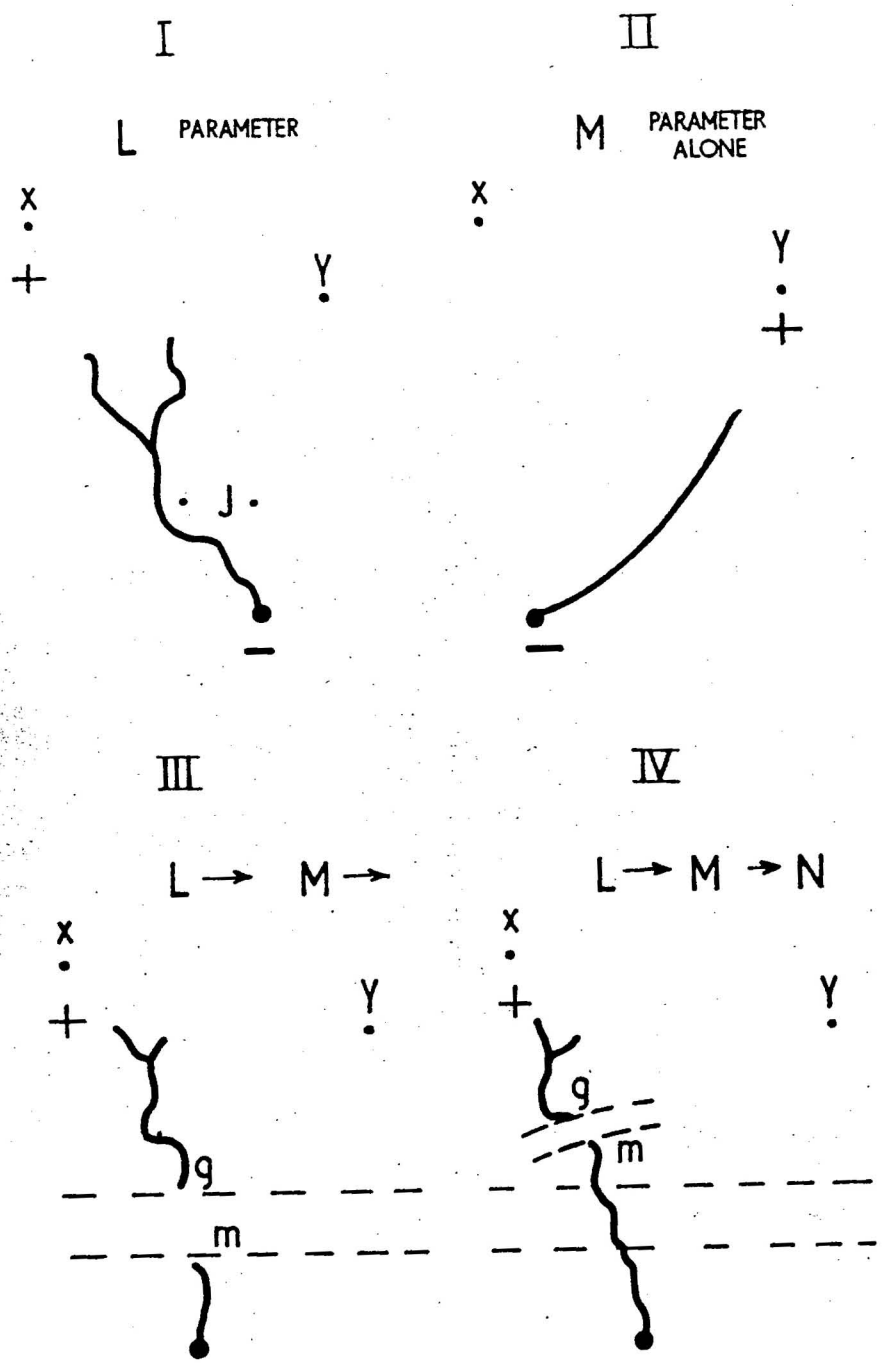

Fig.8

(94009)

4-13.p31

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 $c_j^* = [c_{u,t}, c_{v,t+1}, \ldots\ldots c_{s,t+T}]$.

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.

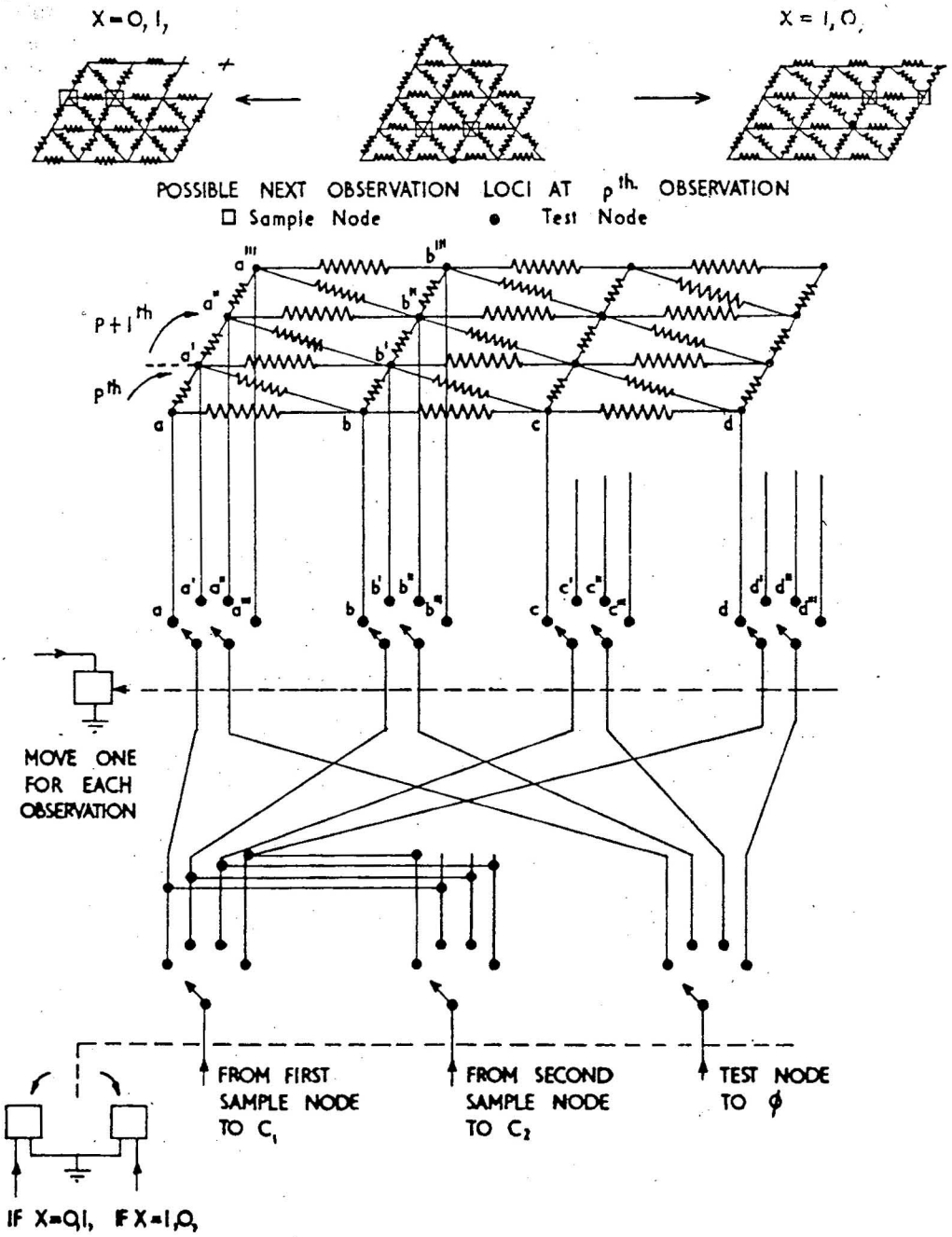
X = 0, 1,
X = 1, 0,
POSSIBLE NEXT OBSERVATION LOCI AT p th. OBSERVATION
□ Sample Node
● Test Node
P+1 th
p th
MOVE ONE FOR EACH OBSERVATION
FROM FIRST SAMPLE NODE TO C1
FROM SECOND SAMPLE NODE TO C2
TEST NODE TO ϕ
IF X=0,1, IF X=1,0,

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_i 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 show in 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 $X = 1,0$ 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.

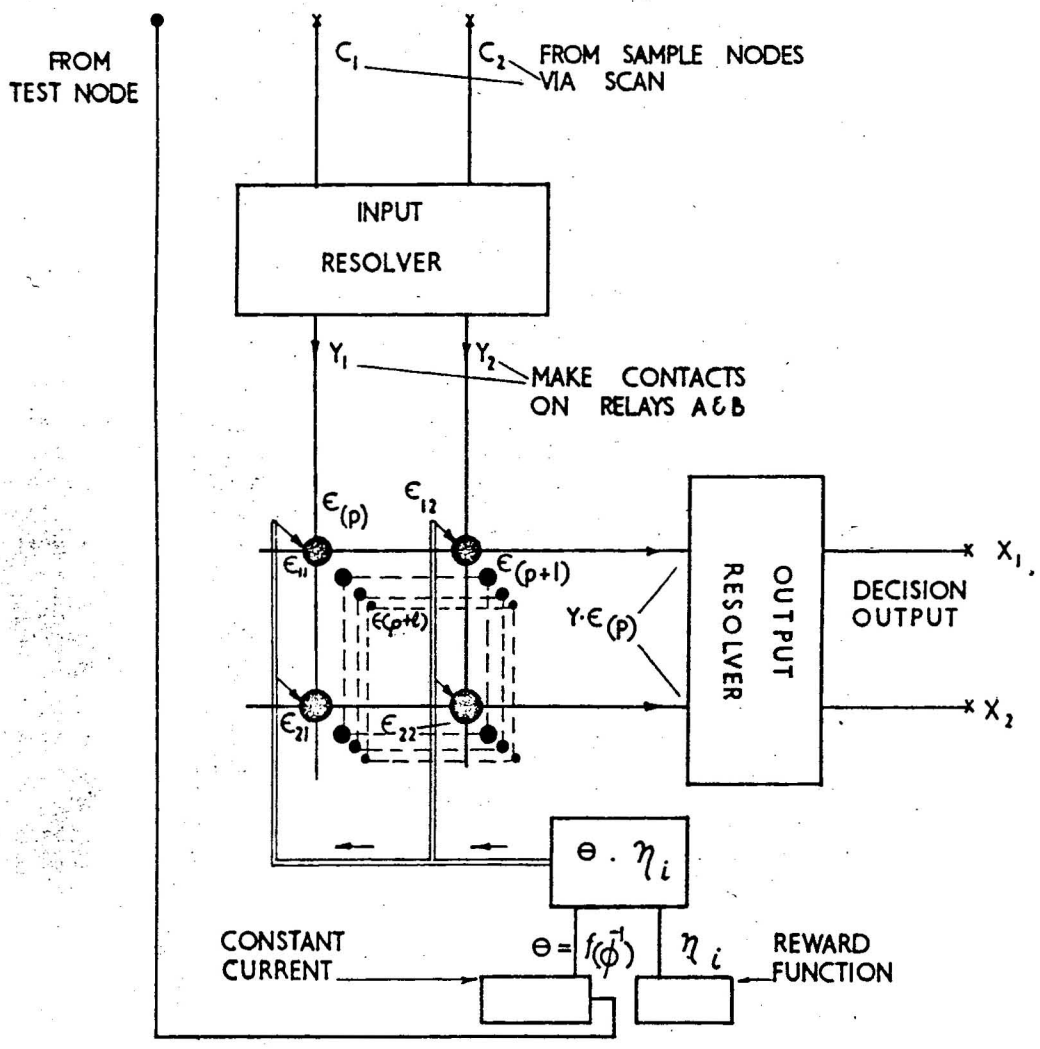
FROM
TEST NODE
C1
C2
FROM SAMPLE NODES
VIA SCAN
INPUT
RESOLVER
Y1
Y2
MAKE CONTACTS
ON RELAYS A & B
ε(p)
ε12
ε11
ε(p+1)
ε(p+ℓ)
Y·ε(P)
ε21
ε22
OUTPUT
RESOLVER
X1
DECISION
OUTPUT
X2
Θ . ηi
CONSTANT
CURRENT
Θ = f(φ⃗)
ηi
REWARD
FUNCTION

Fig. 10

p 38
missing

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. C. (Mullard 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).