THE NATURAL HISTORY OF NETWORKS*

GORDON PASK
Systems Research Ltd., London, England

INTRODUCTORY REMARKS

IN THIS paper I shall put forward a pair of contentions, and examine some of their consequences. The first of these concerns any network which is given in nature or which appears as part of a "Black Box"(1) problem. The contention is that if an observer wishes to use any self-organizing potentialities the network may have, then he must look at the network as though he were a natural historian.

I am using the term "network" in a general sense, to imply any set of interconnected and measurably active physical entities. Naturally occurring networks, of interest because they have a self-organizing character, are, for example, a marsh, a colony of micro-organisms, a research team, and a man.

It is not so easy to say what I mean by a natural historian. Emphatically he is not a meticulous and classifying person. In choosing the name I had the interactive aspects of natural history in mind, the art of knowing about a rabbit run, almost by living the part of a rabbit, the skill of animal training—disciplined enough to permit its discussion—the search for similarities which are cogent within the network itself.

The idea of necessity also needs comment. We can, of course, look at a system in any way we choose, regardless of whether or not it is self-organizing. Thus, we can look at a man from the anatomical point of view and see a creature with two legs, bounded by its skin. Again, we might examine man like the sociologists and see a badly defined game player. The contention is that *in order to use the self-organizing character* of a man we must become natural

* The author wishes to acknowledge his association as Assistant Research Professor with Professor H. von Foerster (Electrical Engineering Research Laboratory) of the University of Illinois, during the writing of his paper. While at the University of Illinois the work was supported by the Information Systems Branch of the Office of Naval Research under Contract Nonr. 1834(21).

historians, which means, for the human system, that we must talk to it. In conversation the system appears to be bounded at one moment by the anatomist's skin, and at the next moment, by its region of influence upon other men in society. Typically a natural historian must change his viewpoint to suit a changeful system.

DIFFERENT METHODS OF OBSERVATION

In order to express these notions precisely we must examine more familiar ways of observing networks and compare these with the natural historian's strategy.

Any pattern of activity in a network, regarded as consistent by some observer, is a system, Certain groups of observers, who share a common body of knowledge, and subscribe to a particular discipline, like "physics" or "biology" (in terms of which they pose hypotheses about the network), will pick out substantially the same systems. On the other hand, observers belonging to different groups will not agree about the activity which is a system.

I shall call a body of knowledge, in which statements are related in a common language, a reference frame.(2) For the observer who adopts it, a reference frame determines the kind of enquiry which is relevant (and thus, the set of physical attributes, of the activity in a network which it is pertinent to observe).

Ultimately, observations and experiments are conducted in order to control the activity in a network. Some observers, a category which will be defined as *specialized observers*, wish to control this activity by discovering the "truth" about how it occurs. They experiment, on the assumption that a sufficiently complicated "truth" is invariant, by trying to identify observable behavior with a hypothetical system. Now any real observer is limited by the number of states he can usefully distinguish in an experiment, and in general he will not be able to identify a sufficiently complicated hypothetical system in any direct fashion with the measurable activity. Rather, he will assimilate results from many experiments, each of which validates only a sub-system of the hypothetical system he assumes to exist. But assimilation is only possible if the experimental observations can be compared and transformed with a well-defined composition rule. It is no accident that the measurable attributes deemed relevant in different reference frames prove incomparable, which implies that results from experiments in different reference frames are not assimilable. Because of this a *specialized observer* necessarily elects to experiment within a single reference frame. Thus, any specialized observer who examines a network will discern only those systems which are manifest:

(1) By changes in a set of relevant variables, or equivalently

(2) Which are composed of the unit elements or components which may be defined if the functional equations of the activity, appropriate for the particular reference frame, are reduced to canonical form.(3)

Systems may also be controlled by interacting with them, and this technique is used by a natural historian. In general, he assumes that the " truth " about the system is not invariant (otherwise he would have examined it like a specialized observer), and his experiments aim either to maximize future interaction or to achieve some more specialized objective (like making a system called an elephant, get up on its hind legs). The natural historian, since he is not seeking the absolute, adopts whatever relevance criteria allow him to achieve interaction.(4) A specialized observer sees him skipping illogically from one frame to another; for example, at different stages in the training process he may " feed " and " entice " the elephant, which are procedures appropriate to strictly incomparable models of the system. As a result of his experiments the natural historian may be able to make assertions about how to interact—like " Give it a bun if its trunk is drooping," or " Pat the creature on the head each day "—but these must not be confused with truths about the system in the previous and rigorous sense. Giving a bun to and patting an elephant, both of which induce it to stand upright, are not procedures comparable with changing the pressure and the temperature of a gas, both of which make it change volume. The laws of gaseous behavior are expressed in a single reference frame. The laws of elephant behavior are not. Thus, a natural historian cannot say anything precise about the way that elephants (or other systems) work. He makes comments only about his interaction.

While admitting this limitation, I believe that a natural historian can answer all of the enquiries it is either legitimate or useful to make about a self-organizing system. The natural historian's language is appropriate for discussing behavioral characteristics some of which are vague, some of which (like the redundancies and stabilities described by McCulloch(5,6) and the habituation described by Ashby(7)), are firmly related to topological parameters of the network, and a few of which like " differentiation " and " memory " will be examined in this paper. The natural historian is unable to say where these behavioral characteristics reside in the network or how they are manifest, but questions of this kind are probably meaningless in the self-organizing context.

Mathematical Representation

It would be inappropriate, in this paper, to discuss the mathematical work which is being done by A. Mullin at the University of Illinois and which will elaborate these ideas. However, I shall present a few descriptive structures with the provision that a more elegant formulation will emerge when the optimum mathematical technique for dealing with the observer and network problem has been determined.

Specialized Observation

A model or hypothetical system $J_j = (U_j G_j)$ is a set U_j^* of points U in a phase space $\mathcal{U}$ together with a finite group G_j of transformations $F \subset G_j$.

Clearly the models J_j with $(j = 1, 2, \ldots)$ are consistent, due to their group character, and are elements of an hierarchy, the coherence of which depends upon the adoption, by a number of specialized observers, of a certain " Composition Law." In the present discussion we shall assume that this " Composition Law " is matrix multiplication and that the groups G_j have thus the usual connotation.

In this hierarchy the lowest hypothetical models will be determined by cyclic groups generated as the powers of a single transformation such as $G_i = F_i F_i^2 \ldots F_i^{c-1}$ with $F_i^c = I$ the identity transformation and $(i = 1, 2, \ldots)$.

The model $J_i = (U_i^* G_i)$ is thus a model of a stable system, for any state $U \subset U_i^*$ will be repeated in c units of activity, in general, in c observational intervals.

In any models J an observer is able to describe, in a way which is unambiguous to other observers adopting his composition law, those features which are kept invariant by $F \subset G$. However, in order to say that this description is the " truth," the model to which it refers, and in the simplest case a model like J_i must be experimentally tested.

Before examining the experimental procedure, let us note that as a result of this testing there will be two disjoint subsets of hypothetical systems in the hierarchy, and thus two disjoint hierarchies; one being built up from experiments which yielded results confirming the existence of hypothetical systems, the description of which is "true," and the other including plausible but experimentally unconfirmed models.

An experiment is an attempt to make an identification Ω_γ between the points U in a hypothetical system and the states of a real network. As a result of this an observer may also relate the abstract transformations F to real mechanisms in a network or alternatively to real stimulus procedures which he can use to modify the state of the network. If identification is possible for any J_i, the system J_i is said to exist in the network.

Because of his limitations an observer is unable to make any experiment he pleases. The restrictions are of two kinds. First of all, if we regard Ω_γ as a mapping between a vector of observed values of measurable attributes of a network and a state of the hypothetical model, the mapping is not isomorphic, but is many to one. However, in order to satisfy the requirements of identification it must preserve the Composition Law of the observer. Thus, Ω_γ is a homomorphism. Secondly, identification implies a qualitative decision to regard only certain attributes of a network as relevant. These chosen relevant attributes determine the observer's reference frame. Thus, in the reference frame α a set of real attributes, say $x_1, x_2 \ldots\ldots x_n$ are examined, so that observations of states of networks are vectors $X_t, X_{t+1} \ldots\ldots$ at instants $t, t+1, \ldots\ldots$, and the real mechanisms are transformations $A_1, A_2 \ldots\ldots$ Similarly, in the reference frame β a set of attributes $y_1, y_2 \ldots\ldots y_m$ is regarded as relevant and observations are the vectors $Y_t, Y_{t+1} \ldots\ldots$ and real mechanisms are $B_1, B_2 \ldots\ldots$

If α and β are distinct it is clear that the set of all experiments performed in α which affirmed the existence of a hypothetical system will determine a sub-hierarchy α^* included within the previously defined "true" hierarchy. We define reference frames α and β so that the variables x_ξ and y_χ are incomparable for $\xi = 1, 2 \ldots\ldots n$ and for $\chi = 1, 2 \ldots\ldots m$, in the domain of the observer's composition law. Because of this α^* and β^* are disjoint.

Two points must be mentioned. First, it is likely that observers

have selected relevant attributes, merely because affirmative results derived from measuring relevant vectors were self consistent, being in a sub-hierarchy α^*, but were discovered inconsistent with, or unrelated to, those derived from measuring different relevant vectors which would be in a disjoint sub-hierarchy β^*. Secondly, it is clear that reference frames $\alpha, \beta \ldots\ldots$ sub-hierarchies $\alpha^* \beta^* \ldots\ldots$ and even the "true" sub-hierarchy are defined with reference to knowledge at the moment. The concepts are useful, since we are limited observers. However, it is always possible that reference frames will be rendered indistinct, that we shall adopt a new composition law which relates previously incomparable quantities and that the region of "truth" will extend. In particular, the distinction which will be made in a moment between specialized observers and natural historians would disappear if observers were able to know everything about a network.

We can represent the simplest experiments of a specialized observer as shown in Fig. 1. In the particular structure of Fig. 1 a

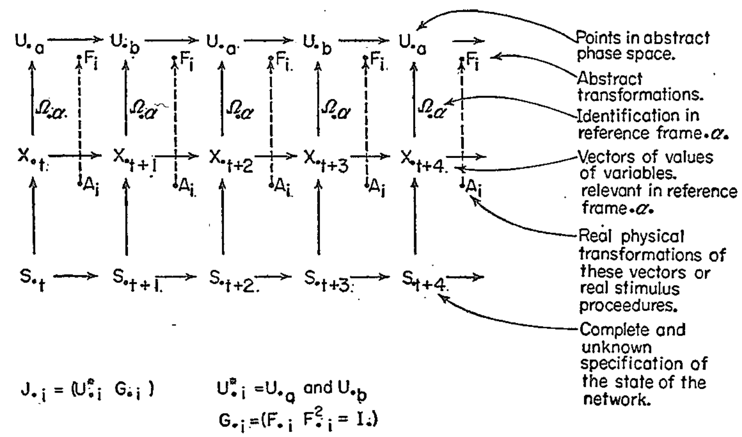


FIG. 1.

hypothetical model is identified in a reference frame α since the abstract state U_a is identified with X_t the abstract state U_b with the X_{t+1} and F_i with the mechanism A_i.

The actual states of the network, namely, $S_t, S_{t+1} \ldots\ldots$ and so on are accessible only to an observer able to know everything, and that is, for the present discussion at least, "unreal." The mapping from any S_t into the relevant variable space of a reference frame is not necessarily unique. The general case of one sequence $S_t, S_{t+1} \ldots\ldots$ and several distinct reference frames $\alpha, \beta, \gamma, \delta$ is illustrated in Fig. 2. Further, as indicated in Fig. 2, it is possible that either there are

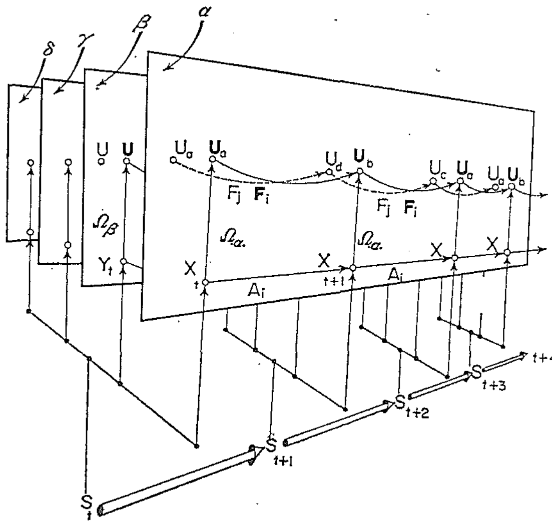


FIG. 2.

two or more observers in two or more different reference frames, who all identify the activity in a network with different hypothetical models and who thus assert that different systems exist in the network, or alternatively that only some of these observers are able to identify the activity with hypothetical models, or finally that none of them can.

It will be possible to identify the activity with a hypothetical model only if the mechanism involved in producing the activity may be discussed in terms of the composition law of the observer. All stationary stochastic processes are, for example, identifiable with stochastic models[(8)] and the derived binary matrices with binary permutation matrices.[(9)] However, there are some non-stationary systems, those which are often encountered in sociology and psychology, such that however long a sequence is observed in a given

reference frame, U_t is not equal to $U_{t+\tau}$ for any finite τ, any identification, and any model J. Ashby[(10)] calls them systems where the " truth " is changeful. It is no longer meaningful to make enquiries, as we tacitly do when adopting a reference frame, on the assumption that some descriptive " truth " remains invariant.

Natural Historians

Non-stationary activity in a network may be quite tractable for a natural historian because he is at liberty to relate entities which are incomparable to a specialized observer and to run back and forth through the dividing planes of the illustration. So far as the natural historian *has* a reference frame, it is simply the context of his *own interaction* with the network. Unlike the specialized observer, the natural historian has few preconceptions about a composition rule or about what entities or situations are equivalent.

Indeed, in the simplest case, when the natural historian is merely trying to " interact " and " make conversation "[(11)] with some system in the network, his strategy is to discover a set of composition rules and equivalence relations such that if he assumes them interaction will be favored.

We thus suppose the existence of composition rules $E_1, E_2 \ldots\ldots$ which the natural historian is able to distinguish and to understand, but is not necessarily able to describe and equivalence relations $R_1, R_2 \ldots\ldots$ of the same calibre. Initially, he chooses some E_i according to his view about the character of the network as a conversation partner—not according to his knowledge of its structure; and he also chooses some R_i, such that two R_i related situations have the same significance with reference to his interaction. He now seeks to modify these assumptions, as a result of interaction, so that interaction is favored.

Let V_μ, V_η, be the names given to any state of the network which the natural historian is able to recognize. Let $P_l, P_n, \ldots\ldots$ be the names given to any procedures which the natural historian can use to modify the activity in the network. According to the analogy of a conversation $P_l(R_i)P_n$ are a pair of equivalent gambits and $V_\mu(R_i)V_\eta$ are a pair of replies with the same meaning, so far as the second partner is concerned. The process of searching, which gives rise to a sequence $R_1 \rightarrow R_2 \ldots\ldots \rightarrow R_p$ and (because alterations in R induce alterations in E) a sequence $E_1 \rightarrow E_2 \ldots\ldots \rightarrow E_p$ is intuitively familiar. The object of the search is to discover R^* and E^* such that

$$V_\mu(E^*)P^c(R^*)V_\mu$$

where

$$V_\mu(R^*)V_\eta \text{ if } (V_\mu(E^*)P_l)R^*(V_\eta(E^*)P_n)$$

when

$$P_l(R^*)P_n$$

Call the above consistency or predictability condition " & " and note that many pairs E^* R^* will permit satisfaction of " & " for different V_μ and P_l. Let b_γ equal the number of names V_μ and P_l for which a particular pair $R^*_\gamma E^*_\gamma$ permit satisfaction of " &." In general the pair

$$E^*_\delta R^*_\delta$$

is preferred to the pair

$$E^*_\gamma R^*_\gamma$$

only if $b_\delta > b_\gamma$.

The searching sequence, which represents the interaction, will ideally approach an E^*_ξ R^*_ξ such that b_ξ is greater than any of the previously obtained values.

If the search process is one-sided so that the natural historian changes his viewpoint a great deal but has little effect upon the network, convergence toward a high valued b_ξ is unreliable and inefficient. We shall later examine conditions (which always apply if the network is self-organizing rather than merely intractable) in which (because the network is modified by a " reward " under the natural historian's control) an appropriate " rewarding strategy " will achieve rapid and efficient convergence.

Second Contention

The second contention refers to networks which are not given in nature, but which are deliberately built, so as to foster any self-organizing systems which appear. Oddly enough, there are conceptual difficulties which force us to look even at these constructed models in the manner of the natural historian. The contention is that these difficulties are not apparent in the abstract formulation but appear when it is embodied in any physical model such as a network.

It will be convenient to discuss the issue by formulating such an abstract model and then constructing one of the many physical realizations.

AN ABSTRACT MODEL OF A SELF-ORGANIZING SYSTEM

We require a number of concepts for building an abstract model.

(1) A space of arbitrary dimension in which a network is defined by asserting connectivity between pairs of points. Let us envisage this space filled with an initially homogeneous but malleable material M.

(2) A currency, which may be identified with energy, which is conserved on the average. The conservation conditions make measurement possible and will be secured if we have a definite rate θ at which currency or energy flows through the space.

(3) A set of currency seeking servomechanisms. We can identify these elements with von Foerster's Maxwell Demon servomechanisms,[12] or equally with the catalysts which Prigogine and others[13,14] describe as inducing open reaction systems in a stationary state network. They are, thus, non-linear amplifiers or oscillators with a local energy or currency store. Let us assume these servomechanisms exist at uniformly distributed points in the space. The only sense in which any one of these servomechanisms can increase the currency it has available is by influencing the activity of the others. This it may do by transmitting a trial or signal, which other elements sense, and the servomechanism in question is informed of the state of the network by receiving the effect exerted by the trials of other elements at its own input location. However, in making a trial or sending a signal, each servomechanism loses currency—in other words—there is a definite cost per trial.

(4) A set of rules determining the change in signal and currency connectivity induced by activity in the space. These are conveniently expressed by the signal impedance characteristics of the malleable material M.

(*a*) Consider a pair of points ij in M and a path m_{ij} connecting ij in M. Suppose a signal traverses the path m_{ij} at t, the signal impedance of m_{ij} at $t+1$, say $\rho_{(ij)\ t+1}$ is less than the previous signal impedance $\rho_{(ij)\ t}$, the decrease being due to passage of a signal at t. In the absence of further signals along this pathway, the signal impedance will increase and reach its original value in some finite interval τ, then $\rho_{(ij)\,t} > \rho_{(ij)\,t+1}$ only if a signal passed along m_{ij} in some preceding interval less than τ. A pathway may thus be defined simply as an m_{ij} connective in M where the signal impedance is greater than the average impedance in the ij neighborhood. Clearly, pathways, structures of signal connectivity, or networks arise necessarily if the servomechanisms are active and continue to exist only if they are used.

(*b*) When a servomechanism is active it uses currency. Let some servomechanism at a point i be active so that a more than average flow of currency must occur in M along pathway m_{ij}. When such a flow of currency occurs the currency impedance of $m_{(ij)}$, say $\varkappa_{(ij)}$ will increase. If the flow of currency is greatly decreased, due to a suppression in the activity of the servomechanism at i, the currency impedance $\varkappa_{(ij)}$ decays over a finite interval.

It may be argued:

(i) that activity must occur, for the currency available must be used.

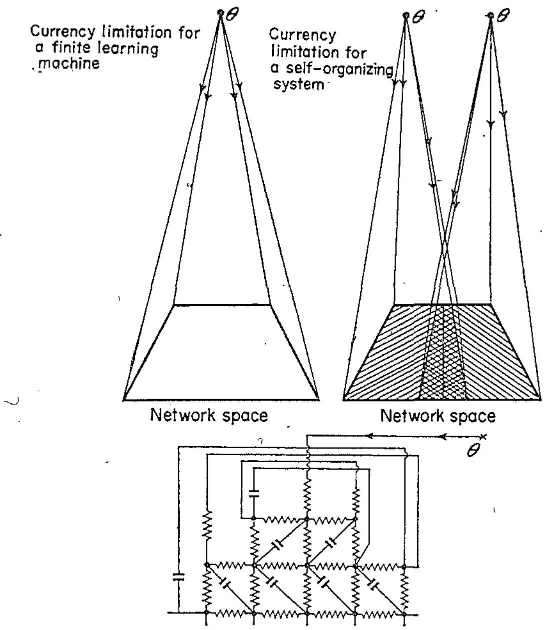
(ii) that this activity will give rise to some kind of connectivity in M.

(iii) that an active set of connections imply local activity.

Since local activity engenders local currency depletion, a new currency distribution and thus a new activity distribution is induced by the existence of the original connectivity.

It is clear that uniform connectivity in M together with uniform activity of the servomechanism elements is, in general, an unstable equilibrium, because in these conditions the system is searching for and is maximally sensitive to, any disturbance which will interrupt the uniformity. However, such a uniform state, which I shall call a λ state, may be shown to be the only stable state if the system is closed (except with reference to currency), so that no disturbance can occur. We may obtain this result by applying von Foerster's[15] Multiservo Convergence Theorem to the *servomechanism elements*, when, assuming uniform connectivity, the representative points of *the ensemble of servomechanisms* must converge to a fixed point. But, if this tendency should occur, then whatever the initial connectivity of a closed system of this kind, the terminal connectivity will approach a uniform pattern. In such a variable connectivity

system there are indeed many different connectivities equivalent in that they are all uniform; thus the system will approach one of a set of λ states. The limit case of a system, such as Dr. McKay's trial making servomechanism,[16] which has fixed connectivity and approaches a unique λ state is, however, included by the formulation.



It is assumed that the solid angles shown above are filled up with a material which has a use dependent currency impedance. If currency is electrical current a structure of this kind is an acceptable analogue

FIG. 3.

Clearly, if the state of the system is coupled to parameters of an environment and the state of the environment is made to modify parameters of the system, a learning process will occur. Such an arrangement will be called a *Finite Learning Machine*, since it has a definite capacity. It is, of course, an active learning mechanism which trades with its surroundings. Indeed it is the limit case of a self-organizing system which will appear in the network if the currency supply is generalized.

Suppose that the network space is indefinitely extensive and that instead of limiting currency flow to θ per network space we restrict the flow to θ per unit volume of the network space. In this case there is an advantage to be gained in terms of the competition for currency between the servomechanism elements, if these elements co-operate. In other words, a set of servomechanisms is at an advantage if its activity extends the connected region in the network space. The extension will only be limited by the gain of the servomechanisms and the currency available. In realizable systems an active connected region moves around the network space capturing uninvolved servomechanisms. Such a system will be called *an abstract self-organizing system.*[17] Since we cannot satisfactorily demark the active system, the inactive region in the network, and the environment; closure cannot be applied. However, if it could, there would be an indefinitely large number of λ states and these are approached as closure is approximated.

When related to a specific environment, this is a learning machine but not a finite learning machine, since the extent of the active system depends upon the external conditions. The relation of such a self-organizing system to the finite learning machine is indicated in Fig. 3.

Physical Construction of the Model

When any physical model is constructed, its maker has to accept certain essential constraints inherent in the medium. The existence of non-linearity in any real amplifier, the thermal coefficients of any real resistor are, for example, essential constraints. However, in building most models it is possible to select only one set of restrictions as being relevant to the action of the physical artefacts. Thus, when an electrical analogue computer is used to embody some abstract mechanism we say that an electrical model has been constructed, meaning that in realizing this abstraction we take account of the electrical model, but that we disregard, as not being relevant, the mechanical and thermal constraints inherent in the computer. Indeed, the computer is designed with this object in view, and a different computer might have been designed to embody abstractions in a mechanical model, electrical effects being discounted.

In terms of our convention, a computer analogue is a physical model so designed that its activity is explicable in a single reference frame, in the case I have cited, an electrical reference frame. In terms of engineering, such a mechanism is designed with components, like valves and resistances, which have a well specified function. A valve, for example, accepts only a electrical input and provides an amplified electrical output. If it also responds to temperature or vibration, it is to this extent a bad valve. The logical simplicity of the computer model is a consequence of being able to put one's finger upon a component which performs a known function and to reject the imperfections as irrelevant.

When trying to construct the physical model of an abstract self-organizing system we are beset with a peculiar difficulty. Not only are there many possible mechanisms which embody the abstract concept, but any mechanism we choose will embody it in an ambiguous manner. The logical requirements force us to use media such that, when a physical model is constructed, we cannot specify components which have a well defined function, and we cannot separate inputs and outputs into a set which are relevant and a set which may be discounted.

It is inherent in the logical character of the abstract self-organizing system that all available methods of organization are used, and that it cannot be realized in a single reference frame. Thus, any of the tricks which the physical model can perform, such as learning and remembering, may be performed by one or all of a variety of mechanisms, chemical or electrical or mechanical.

Thus, however much we try, we cannot achieve an electrical model or a mechanical model or a chemical model of a self-organizing system. Any physical model necessarily includes them all in varying degrees, and to a specialized observer they will appear distinct and incomparable, although a natural historian will be able to see them as equivalent.

To emphasize this point let us consider the process of differentiation. Suppose that at an early stage in its development the system has learned the advantage of having " individuality " in the sense that it has developed a primitive mechanism using specific substances, for example—proteins—to tag each " individual entity." Later the system learns about the existing primitive mechanism and evolves a more efficient device whereby " individual entities " are spatially separated regions connected, distinctively, with fibers as in a nervous system. The primitive and the efficient mechanisms are functionally equivalent to a natural historian who regards " individuality " as a " behavioral characteristic " but incomparable to a specialized observer for whom " individuality " is unmeasurable. (In the network I shall describe, it happens that regional connectivity is given, but it is possible to distinguish at least two equivalent mechanisms which are developed for mutual inhibition, one acting by energy depletion and one which involves a specific connectivity.)

These ideas can be placed on a firm theoretical foundation by considering the system as it approaches a λ state. In this limiting condition a specialized observer sees a meaningless activity from which he can only infer the existence of a chance machine. A natural historian, on the other hand, sees a system which is maximally sensitive to any disturbance and liable to develop any one of many equivalent structures according to the disturbance it happens to appreciate. However, I shall not pursue this theoretical argument. Having made the point that we must view constructed networks as though we were natural historians, just as we have to view the self-organizing networks given in nature, it will be more instructive to examine the behavior of a real mechanism.

A Particular Physical Model

I shall describe a model[18] in which the " currency " of the abstract system becomes electrical energy and signals may be thought of, initially, as electrical impulses. It will be convenient to describe the physical representation of servomechanism elements and of malleable material separately.

(1) In the model a " Maxwell Demon " servomechanism is an energy dependent trial making amplifier. Due to a mechanism which involves a refractory interval, it may distinguish its own output from the output of other elements, or the delayed effect of its own output acting as an input.* The element produces electrical output impulses called trials, because in the first place sufficient electrical energy is dissipated to modify the state and thus the connectivity of the surrounding material, and secondly, because

* In this respect the model is similar to the scheme discussed in ref. 19.

the impulse, transmitted through any existing connections may affect the trial making activity of other elements.

Each element has an electrical reservoir in which trial making energy is accumulated. Occasionally, the stored energy is dissipated by autonomous trial. In general, however, the sequence of trials is modified by inputs received, at a much lower energy level, from other elements in the sense that an input stimulates the occurrence of a trial. The gain of the element, as an amplifier, is a function of the average difference between input and trial energy, and in practice, we may look upon any element as a servomechanism which is seeking to maximize interaction, subject both to energetic constraints and those imposed by the connectivity built up as a result of the previous activity.

(2) To clarify the presentation I have separated the energy supply network from the connectivity or signal network.

As shown in Fig. 4 the trial making amplifiers receive energy—in this case electrical current—from a resistance–capacitance network into which a current θ passes at one or more points of symmetry.

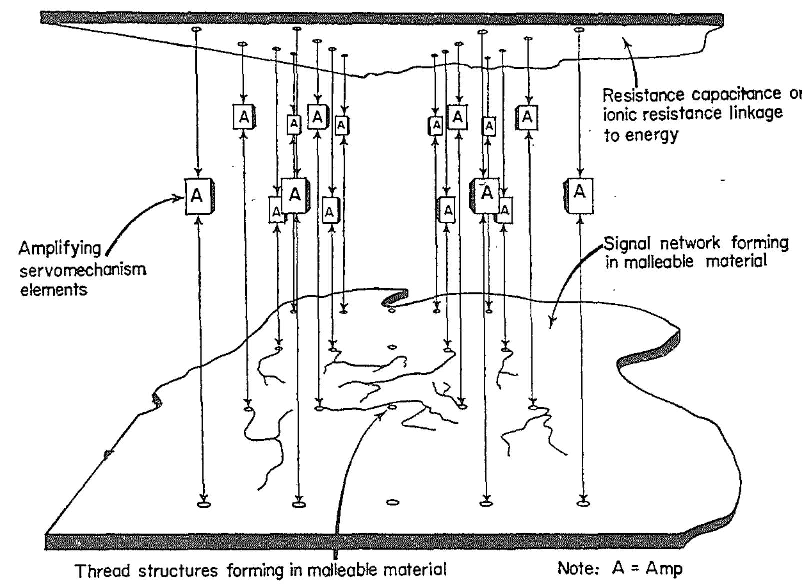


FIG. 4.

The resistance–capacitance network may be loaded as a result of excessive trial making activity on the part of a particular amplifier. If so, a local depletion will occur thus reducing the effective source-potential and the effective gain of the amplifier concerned.

The signal network is built up as a result of current passed by these amplifiers through a solution of ferrous sulphate, which is the malleable material *M*. As shown in Fig. 4, the current path is completed via a set of electrodes.

The electrode associated with each amplifier may act either as a source or a sink of d.c. current, according to whether a trial is or is not being made. If a trial is being made the amplifier also produces an a.c. signal at its electrode, which may be received by any electrode which is acting as a sink for d.c. current.

The solution itself is moderately conductive to the a.c. and the d.c. signals. However, if a d.c. current is passed between a source and a sink, a very low-resistance metallic thread develops from the sink along the line of maximum current, and gradually an entire network of threads is built up. The line of maximum current, where a particular thread develops, will depend upon the electrodes which are energized and also upon the existing network of threads since these, being of low resistance, act as extensions of the point electrodes. Thus, the network of threads not only distributes the a.c. signals which deliver inputs to the elements, but determines the further development of the network itself.

Once a thread is formed, there is a tendency for it to dissolve, due to a local acidity. A stable thread is thus in a dynamic equilibrium determined by the competition of a building up and a dissolving back process. A thread exists as a stable entity only if it is passing sufficient current to keep it intact and if the current is appropriately distributed. Clearly the distribution which *is* appropriate depends upon the entire network and its activity. In general, the network of threads determines the environmental parameters in which any particular thread develops and any particular thread determines the environmental parameters in which a small segment develops. Thus, the natural history of this network presents an over-all appearance akin to the natural history of a developing embryo or that of certain ecological systems.

Some of the mechanisms used in development are illustrated in Fig. 5.

(1) Shows a thread developing between electrodes *d* and *e*. It develops by a process of successive trial, nearly all the terminal trial threads being abortive.

(2) Shows the introduction of a further electrode *f*, as a result of which the thread may bifurcate.

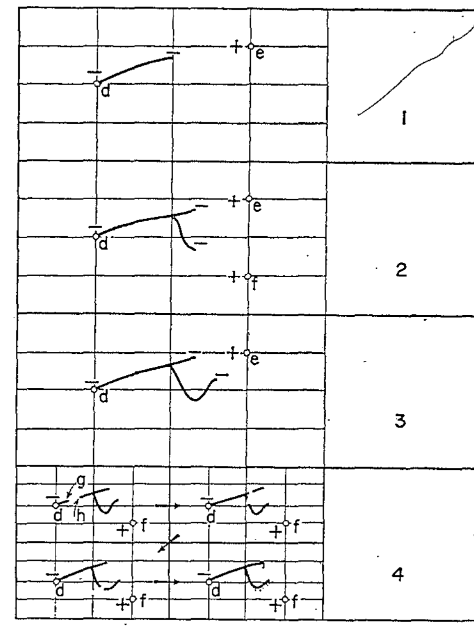


FIG. 5.

(3) Shows the development of the thread after bifurcation, but with only *d* and *e* energized. The effect of having previously energized *f* is apparent.

(4) Shows what will happen if, either due to instability or mechanical injury, the thread is split. The point *g*, being relatively negative, builds up new thread whilst the point *h*, being relatively positive, suffers dissolution. The process gives rise to regeneration of the thread as a whole which occurs up the branches *de* and *df*, even though only *d* and *f* are switched on. The existence of the thread has transformed the field which would have induced regenerative

development along *df* only into a field which induces development of *de* and *df*.

If a subset of the electrodes are associated with the output connections of sensory devices which in turn receive an input from an environment, and if a further subset of these electrodes are associated with devices able to effect the environment the network will interact and change state, seeking dynamic equilibrium with reference to the environment. The extent of the active region, produced in the network as a result of this search, will depend upon the informational variety of the environment and the value of θ (Fig. 6).

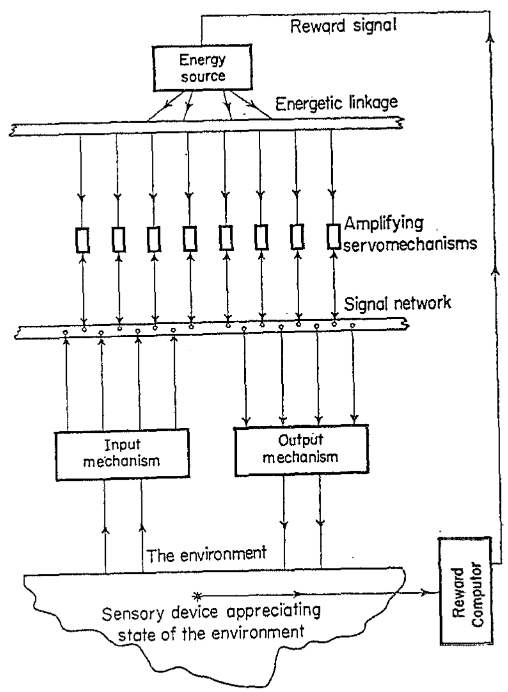


FIG. 6.

The adaptive process will lead to some system which can interact, in a stable fashion, with the environment. However, suppose that an observer tags a subset of the possible stable relationships as desirable; in particular, suppose he wishes the system to act as a control mechanism, which achieves some state of affairs in the environment. He can train the system to adapt specifically in this way by controlling θ. Even a procedure such as increasing the value of θ only if the desired state is achieved will lead by natural selection to a structure which is a control mechanism aiming to achieve this state.

The over-all energy, or over-all currency, variable θ is here being used as a reward variable. This use of the term " reward " is, perhaps, unusual, and is certainly distinct from " rewards " which imply that a certain rewarded action becomes more probable. In the present case, when the network is rewarded, we mean that it is given permission to develop, that more of the constructional material may be used for making threads, and that more amplifiers may be included in the signal network and as part of the system. However, no restriction is placed upon the kind of development, which depends upon the existing structure.

A specialized observer would find this an unsatisfactory learning machine, because although it will learn what he wishes, he cannot tell how it learns, how to reward it, or how large it is. Before considering how a natural historian might administer a reward (according to the present contentions, with greater success), I should like to exhibit a few more characteristics of this model.

(*a*) We have already seen that given appropriate surroundings (namely a world of ferrous sulphate liberally bespattered with amplifiers) the system could extend wherever there is energy. Such a world is unlikely, so we enquire what will happen when a developing system reaches a boundary such that there is no more ferrous sulphate. This is the most primitive possible demarcation of an environment. Equally, of course, a boundary can be imposed as shown in Fig. 6.

(*b*) Keeping to the primitive case, the answer is that the system will endeavour to trade with the environment in the sense that some way of effecting the environment or some change of state in response to changes in the environment will elicit the reward of more energy.

(*c*) It will simplify the discussion to suppose that the system has its state changes coupled in some determinate manner to parameters of the environment and that the problem of getting a reward is thus a problem of sensing those changes in the environment which require a particular response in order to achieve a reward. How, then, does the system appreciate its surroundings?

(*d*) It does so by developing specific sensory receptors. Note, first of all, that a thread structure is slightly sensitive to many disturbances, mechanical, chemical and electrical. Such disturbances, which will be encountered at the boundary, elicit some change of state. If this state change is unrewarded the disturbance in question will, however, be taken as irrelevant and will have little effect upon the system. But suppose that a disturbance, for example, a vibration or a change of acidity induces a state change which is rewarded (in other words suppose the environment is such that when part of the boundary is acid or vibrates some particular modification of the environment parameters makes more energy available to the system) then the system will adapt so that the boundary region becomes specifically sensitive to acidity or vibration. No teleological arguments are required to describe this process of building a sensory receptor for those variables which are sensed with advantage. Reward means permission to build more structures out of basic material. Thus a sensory receptor (which appears because the particular region of the boundary which *did* minimally respond to the environmental stimulus is duplicated and enlarged) will form as a logical consequence of specifically rewarding a system in which the elementary units have no well specified function and may not be regarded as components. Thread structures are just as good parts for pH meters and microphones as they are parts of memory registers and connections.

(*e*) From this it appears that the system can act like a natural historian and develop its own criteria of relevance and bring about any relation with respect to its environment.

(*f*) Although the input of such a system is badly specified, so far as a specialized observer is concerned this does not mean that the system is unable to distinguish input variables precisely. On the contrary, a system may discriminate variables by elaborating its receptor mechanisms to an arbitrary extent, if reward is contingent upon sensing the variables in question. However, it is possible to show by a recursive argument, that a specialized observer will nearly always be ignorant of the variables which are, at any moment, being sensed by the system.

(*g*) Even if the world of ferrous sulphate is finite the system can trade in an indefinite number of ways with its surroundings and in this sense the environment boundary is a fiction. Using this assertion we can mark imaginary boundaries at arbitrary points in the system and examine the trade which takes place across them. This technique was tacitly adopted when discussing the different " mechanisms " for achieving " individuality." In other words different kinds of trading—different sensory receptors—are a special and dramatic case of the different mechanisms which exist as a commonplace feature of the system's activity.

(*h*) Again, these mechanisms evolve one from another. When we were discussing " individuality " the system " learned " about a primitive mechanism in order to evolve a less primitive and (to the natural historian) equivalent mechanism. At any stage in its development (when the system is observed over a short interval) there will exist an heirarchy of mechanisms, corresponding to different stages in its evolution.* Most of these will be vestigial, but the stability of the system derives from their existence and the possibility of their reactivation in adverse conditions.†

Rewarding Strategies

Returning to the strategy of a natural historian, it will be convenient to use some descriptive mathematics, again with the provision that the approach is tentative and will probably be replaced by more elegant and tractable techniques. In the first place let us recall:

(*a*) That a " behavioral characteristic " is, for example, possession of " memory " or " docility " or " habituation."

(*b*) The predictability condition " & " required by a natural historian.

(*c*) The idea of a reference frame.

(*d*) The assertion that a reference frame determines either:

(i) a set of relevant variables, or

(ii) a set of components which have some function—such as being neurones or something even more specific.

---

* This evolutionary structure is typical of biological systems, as pointed out by Professor Bishop.(20)

† If, in the course of interaction with a system, we define certain regions as unit elements [Professor McCulloch's (Ref. 21)] reliability calculus is immediately applicable.

It will now be more convenient to adopt the latter restriction and to suppose that an observer in α sees components (the things which he calls neurones) as entities, $a_1, a_2, \ldots\ldots$ in a network. Similarly, an observer in β sees $b_1, b_2, \ldots\ldots$ (as the different entities he calls neurones). On the other hand, a natural historian takes different entities as being his basic elements on different occasions. He chooses the entities which allow him to make sense of and interact with the system.

We may now argue:

(1) If $\psi_1 \psi_2 \ldots\ldots \psi_\epsilon$ are the behavioral characteristics discussed by a natural historian in terms of the names $V_\mu\, V_\zeta$ of recognized states and the names $P_l\, P_m$ of stimulus procedures, there will be sets of names, some empty, such that all V_ζ and all P_1 included in ξ_p refer to the pth-characteristic.

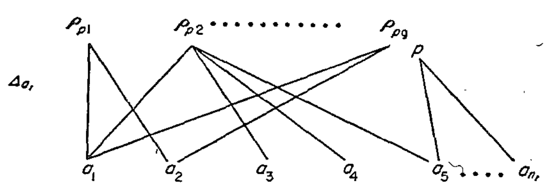
(2) A Behavioral Characteristic may, at an instant t imply several distinct mechanisms. Thus, the pth-characteristic may imply any of g_p mechanisms.

$$\rho_1\, \rho_2 \ldots\ldots \rho_{g_p}$$

(3) In a Reference Frame α we may assert, at t the existence of n_{α_t} active elements a and in β assert n_{β_t} active elements b. In general:

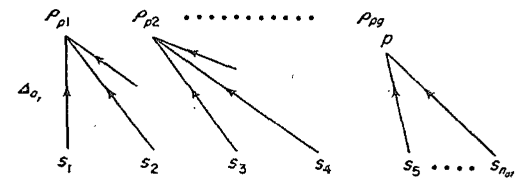
$$n_{\alpha_t} \neq n_{\beta_t}$$

(4) There is a relation Δ which maps the $m_t < \sum_p g$ mechanisms active at t into the set of elements defined in α. Considering only the pth-characteristic this relation Δ_α may be:



If the mechanism p is in a self-organizing system then Δ is, as shown, a many to many relation. Further Δ is different for all $\alpha\beta$.

(5) Sufficiently microscopic observation would discern elements $s_1\, s_2 \ldots\ldots s_{n_{0t}}$ with $n_{0t} > n_{\alpha t}$, $n_{0t} > n_{\beta t}$, for all α and β such that mapping Δ_0 is many to one like:



However, if the system is self-organizing, the elements s are inaccessible to a real observer.

(6) Let $\Phi_{(pd)t}$ equal the number of mapping arrows which converge upon $\rho_{(pd)}$ in the mapping Δ_0 specified at t.

(7) Let $\sigma_{(pd)t}$ be a variable which is equal to 1 if and only if ρ_{pd} is active at t namely if and only if at least one mapping arrow converges upon $\rho_{(pd)}$ in the mapping Δ_0 specified at t.

(8) When a natural historian rewards a system he increases the value of a variable θ so that if a system is rewarded:

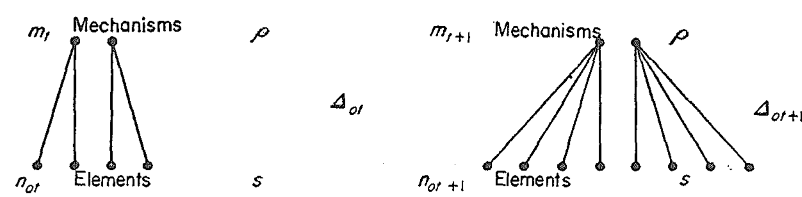
$$\theta_{t+1} > \theta_t$$

(9) We have argued that the effect of increasing θ is to allow those mechanisms active at t to develop.
Thus $\Phi_{(pd)t+1} > \Phi_{(pd)t}$ if and only if $\theta_{t+1} > \theta_t$ and $\sigma_{(pd)t} = 1$.
In general this implies:

$$n_{0t+1} > n_{0t}$$

The effect of a reward upon Δ_α and Δ_β is, or course, unspecified.

(10) Thus if a natural historian knows, or is able to determine that, the variable $\sigma_{pd} = 1$ he can reward the system, so that ρ_{pd} becomes a dominant mechanism, for mediating the behavioral characteristic ψ_p. This will be the case only if the condition "&" is satisfied. Visually represented:



(11) Consider only the behavioral characteristic ψ_p and thus assume that all V_μ and P_l are included in ξ_p.

(12) Assume that each mechanism ρ_{pj} embodies a Composition Rule E_j in the sense that operations $V_\mu(E_j)P_l$ correspond to the operations of this mechanism.

(13) If so the pair E_jR_j will permit satisfaction of " & " and consequently, we write $E_j = E_j^*$. Equally, all Composition Rules E_i^* are embodied in some mechanism.

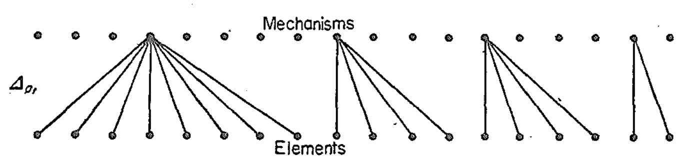
(14) From (9) and (10) if the natural historian rewards the system only if the condition " & " is satisfied for his choice of E_i^* the mechanism ρ_{pi} will become dominant, the Composition Rule E_i^* will become more widely applicable, and b_i will increase.

(15) The process is symmetrical for we can regard the natural historian as rewarded by achieving " &."

(16) More specific strategies are needed if the system is being trained as a specific control mechanism. A single case will be suggested.

(17) Let us apply these arguments to each p.

(18) In this case it is possible to conceive a convergence of interaction by reward such that Δ_0 is so modified that for at least one α the mapping Δ_α is a many to one projection.



In this case, in its terminal condition, the system may be described in a single reference frame α which is not, however, determined at the outset of interaction. Thus the system will have been trained not to be a self-organizing system. It is a finite learning machine.

At the moment it is impossible to provide a general description of the way in which any specified objective should modify the process of interaction. One might avoid the problem by thinking of the natural historian as always a real person and the objective as

something he keeps in mind. This would, however, seriously restrict the admissible training procedures.

Moreover, a theoretical solution is almost certainly possible, since any arbitrary partitioning of a self-organizing system should produce two sub-systems, one of which is being trained by the other. As an Empirical Confirmation of this, it is possible to train a self-organizing system using an adaptive teaching machine[22,23] (which is itself a finite learning machine) as a natural historian in place of the human being.

THE CHARACTER AND UTILITY OF SELF-ORGANIZING SYSTEMS

In conclusion let us review a number of self-organizing systems and consider how our knowledge of their natural history can be used.

By restricting the energy supply of an initially undifferentiated system (according to an appropriate rewarding strategy) it can be trained to act as a Control Mechanism. Clearly, however, this Control Mechanism has little in common with a programmed computer connected to a process by well defined input devices and output devices.

Take, for example, a chemical process. The control mechanism, in the present sense, is something which exists, perhaps on a catalytic surface, within the reaction vessel. It is sold by the cubic foot. Its inputs are sensory receptors, developed in the same manner as the rest of the network, and thus, although a variable such as pH may be sensed, we should not be able to indicate what part of the network sensed it, or to say that any part sensed it exclusively. The outputs of the control mechanisms might, in this case, be regional activations of the catalytic surface supporting the network.

Although metallic thread structures are conceptually useful, the particular model has a limited field of practical application. However, many building materials are available, and it seems likely that the optimum choice of materials will be different for different applications. Thus, in chemical control mechanism, the self-organizing system can sometimes be built up from the actual reactants.

There are two lines of thought, one leading us closer to conventional control mechanisms and one leading us further away. Pursuing the first, it is always possible to replace the process which creates elements in a self-organizing system by an equivalent process which activates elements that already exist in large numbers. Similarly a network in which connectivity is built up is always equivalent to some network which has been derived from a large and fully connected plexus by removing connection.

Applying one or both of those equivalences we arrive at self-organizing systems either specially constructed, like some of the networks of the Illinois project,(24) or, when feasible, programmed on to a computer, like Selfridge's Pandemonium.(25) In the latter case, currency appears as an abstract cost function, but it is important to notice that reward must still mean permission to develop; for example, permission to take over more storage capacity in the computer or to replicate a demon. This connotation of reward is typical of a self-organizing system and distinguishes it from a superficially comparable structure.

Programmed systems of this kind are useful as research tools, as functional models of the brain, or when associated with a process by conventional input and output devices as control mechanisms which can deal with a non-stationary process.(26)

Pursuing the second line of thought it occurs that nature has provided us with excellent physical models of self-organizing systems made out of protein. Beer(27) has discussed the possibility of using unicellulars, the colonial behavior of which is taken as organization, as the amplifying elements in a machine. He and I have examined models, where currency is food supply and unicellulars like paramecium, are active elements, sufficiently to show that such colonial organization may be coupled to a real process.

Self-organizing systems lie around us. There are quagmires, the fish in the sea, or intractable systems like clouds. Surely we can make these work things out for us, act as our control mechanisms, or perhaps most important of all, we can couple these seemingly uncontrollable entities together so that they control each other. Why not, for example, couple the traffic chaos in Chicago to the traffic chaos of New York in order to obtain an acceptably self-organizing whole? Why not associate individual brains to achieve a group intelligence?(28)

These will remain intriguing ideas and no more until definite procedures are specified. There is a great deal of work to be done, but even at the present stage it is possible to envisage the form these

procedures must take. According to the present contentions there exist at least two sets of rules. The first set are rules, of a somewhat ephemeral character, which help a natural historian (or an adaptive machine) to interact efficiently with a self-organizing system (or in some cases determine the interaction of two self-organizing systems). As a result of the interaction some continually changing descriptive model is built up. Knowing this model, and in particular knowing the entities which are regarded at any moment as " elements " of the self-organizing system a second set of rules (which refer, for example, to reliability and habituation) come into play. The second set of rules are, in themselves determinate but they are applied to a " model " (the natural historian's model) which continually changes its relation to the real world.

REFERENCES

1. R. ASHBY, *Introduction to Cybernetics*, Chapman & Hall (1956).
2. G. PASK, Physical analogues for the growth of a concept, *Proc. Symposium, Mechanisation Thinking Process*, N.P.L. Teddington, 1958.
3. M. KOCHEN, Organised systems with discrete information transfer, *General Systems Society Yearbook*, 1957.
4. G. PASK, Organic control and the cybernetic method, *Cybernetica* 3 (1958).
5. W. MCCULLOCH, Agathe Tyche, *Proc. Symposium, Mechanisation Thinking Process*, N.P.L. Teddington, 1958.
6. W. MCCULLOCH, Reliability of biological systems. This volume, p. 264.
7. R. ASHBY, The mechanism of habituation, *Proc. Symposium, Mechanisation Thinking Process*, N.P.L. Teddington, 1958.
8. BUSH, MOSTELLER and THOMPSON, *A Formal Structure for Multiple Choice Situations, Decision Processes*, John Wiley, New York, 1954.
9. D. ROSENBLATT, On linear models and the graphs of Minkowski Leontief matrices, *Econometrica*, No. 2, 25 (1957).
10. R. ASHBY, *Cybernetica*, No. 2, Namur (1958).
11. G. PASK, Teaching machines, *Proc. 2nd Cong. Intern. Assoc. Cybernetics*, Namur, 1958.
12. H. von FOERSTER, Aspects in the design of biological computers, *2nd Cong. Intern. Assoc. Cybernetics*, Namur, 1958.
13. (For a recent discussion and full references to Prigogine's work.) C. FOSTER, A. RAPPORT and E. TRUCO, Some unsolved problems in the theory of non-isolated systems, *General Systems Society Yearbook*, 1957.
14. (For cellular catalysis.) J. POLONSKY, Quelques relexions sur l'aspect physique de la vie, *2nd Cong. Intern. Assoc. Cybernetics*, Namur, 1958.
15. H. von FOERSTER, Basic concepts of homeostasis, homeostatic mechanisms, *Brookhaven Symposia in Biology*, No. 10 (1957).
16. D. M. MCKAY, The epistemological problem for automata, *Automata Studies*, Princetown (1956).
17. Quarterly Report, Contract Nonr 1834(21), Electrical Engineering Research Laboratory, University of Illinois, Urbana (1959).

18. G. PASK, The growth process, *Proc. 2nd Cong. Intern. Assoc. Cybernetics*, Namur, 1958.
19. J. PLATT, Amplification aspects of biological response and mental activity, *Amer. Scientist*, Vol. 44, No. 2, April 1956.
20. G. H. BISHOP, The importance of feedback through the environment in the determination of motor activity. This volume, p. 122.
21. W. MCCULLOCH, Agathe Tyche, *Proc. Symposium, Mechanisation Thinking Process*, N.P.L. Teddington, 1958.
22. G. PASK, Electronic teaching techniques, *British Communications and Electronics*, Vol. 4, No. 4, April 1957.
23. G. PASK, The teaching machine, *Overseas Engineer*, February 1955.
24. Reports of Contract Nonr 1834(21), Electrical Engineering Research Laboratory, University of Illinois, Urbana.
25. O. G. SELFRIDGE, Pandemonium, a paradigm of learning, *Proc. Symposium, Mechanisation Thinking Process*, N.P.L. Teddington, 1958.
26. S. BEER, The irrelevance of automation, *Proc. 2nd Cong. Intern. Assoc. Cybernetics*, Namur, 1958.
27. S. BEER, *Cybernetics and Industry*, English Universities Press (in print).
28. O. D. WELLS, Artorg Publications.

DISCUSSION

NEWELL: I am not quite sure that I understand this notion of the point of view on the natural historian. Let me ask you a question by proposing an example and seeing how this fits with your notion.

PASK: Very good.

NEWELL: If we wish to describe a human behaving, solving problems in logic, and we wish to build a program that describes him, then empirically it turns out about the only way any of us ever found out how to do it is to start from the beginning and proceed to build the program forward. We thus proceed for the first little bit of his behavior until that goes foul and then go back and proceed to reconstruct the program and again start out from the beginning until we run into the next little bit of behavior where it doesn't work, and so on, until in some sense we proceed through the entire structure of behavior.

Whenever you tell this to people they always ask why you don't in some sense first characterize what he does roughly and then specify this a little more and the empirical answer is, every time we try it this way we can't do it and the only time we ever seem to make progress is in some sense where we proceed almost at full detail at least the full details we can tolerate—and simply cruise in and follow, if I may now impose the word, the natural history, and follow in the footsteps of the person himself building the program as we go along.

Can you juxtapose this with your notion of a natural historian in some way?

PASK: If you will permit me, I will invert the example because the distinction I am making is only made for a finite observer. In other words, this distinction of a natural historian from a specialized observer is one that exists simply because observers are not almighty. If indeed they were, of course, they could get at a rock bottom, unambiguous, state description of the system. Equally well, they could split the system up into parts, each of which would have a definite function. However, we are imperfect observers but still find scattered around us in the real world, systems like dogs and elephants and such like things, which it would be too difficult, perhaps, too costly, for us to split down analytically in this way.

The question arises whether we can usefully observe these without trying to describe them in any " absolute " and " true " sense. Because although we are simply incapable of reaching the whole truth we can still usefully employ these admirable self-organizing systems.

The approach of a natural historian is a compromise, but I believe a very valid compromise. It is a way of dealing with systems that are not accessible to us and it is a method of doing it in much the manner which we use when we are trying to control another human being.

For example, you and I at the moment are having a conversation and you are putting an idea over and I am putting an idea over and we are using a procedure which relies in the first place upon inferring similarity between ourselves. We agree we are similar human beings. We understand each other's language and we suppose that certain similarity relations exist within the system as a whole.

I submit we infer all this on the basis of properties such as habituation, or such as having the kind of redundancy which Dr. McCulloch, I think, will describe this afternoon (that is the redundancy of potential command). Some of these properties, like the particular ones I have cited, may be expressed in rather exact terms. You can give a very good topological interpretation of the network structures that will evidence these properties.

On the other hand, some of them are vague. There are properties like " attitudes " and " mental set," and in such cases, we simply can't operate in terms that will allow us to express the concepts with which we infer similarity in mathematical language. The natural historian's approach is an endeavor to get a calculus which allows us to deal with situations like this and to interact with systems of this sort. A Natural Historian, as defined, has nothing to do with the *best* way of finding out *how* a system works. He is concerned actively to trade with the system. Given that he can, I further suggest it is possible to state, with precision, how we may interact with a system, even though we are unable to say how the system works. Does this answer you, sir, or have I missed your point?

NEWELL: That is a partial answer. I would like elucidation on this point. Is it fair to say that the human reacts by having a functional description rather than a human? Is it the kind of terms they use that work?

PASK: It is a functional description, sir, but in a changing language. You change your sample space and relevance criteria as you go along. At one stage of the conversation you talk about things which are logically speaking incomparable with other things introduced later. Yet you, yourself, and those in conversation with you, are able to tie these logically incomparable entities together.

CHAIRMAN MCCARTHY: Have you tried yet any of these other forms of stimuli such as pH?

PASK: We have made an ear and we have made a magnetic receptor. The ear can discriminate two frequencies, one of the order of fifty cycles per second and the other of the order of one hundred cycles per second. The "training" procedure takes approximately half a day and once having got the ability to recognize sound at all, the ability to recognize and discriminate two sounds comes more rapidly. I can't give anything more detailed than this qualitative assertion. The ear, incidentally, looks rather like an ear. It is a gap in the thread structure in which you have fibrils which resonate with the excitation frequency.

BASTIN (*King's College, Cambridge*): Concerning the production of a sense organ in a mechanical assembly, is this an illustrative model? If it is more can you tell us how it can help us discover things that we have not known before?

PASK: Yes, I think it is more than an illustrative model. The development of a sense organ is a dramatic way of showing what we mean by relevance criteria in connection with a machine.

Let us distinguish between that sort of machine that is made out of known bits and pieces, such as a computer (in terms of which we can make models of most physical assemblages, those normally studied in physics, and perhaps some of those we normally study in biology), and a machine which consists of a possibly unlimited number of components such that the function of these components is not defined beforehand. In other words, these " components " are simply " building material " which can be assembled in a variety of ways to make different entities. In particular the designer need not specify the set of possible entities.

Now, to say that this sort of machine will make sense organs is illustrative, because a sense organ is a rather special component construction in the sense that it specifies the boundary between the machine and its environment, and if the machine constructs its own sense organs out of the building material this boundary is apt to change continually. Further, it helps us to understand why we find it is difficult to observe these self-organizing systems, namely, because, looking at them in a single reference frame, in the capacity of scientific observers we cannot lay down definitions which allow us to compartmentalize the functional components of the machine or even the machine itself.

To see that this illustrative model is non-trivial you must recall that we reward the system without specifying, for example, that it will be rewarded for assembling component A and component B. We simply say it will get more reward—(meaning that it will be allowed to use more components) if a structure that acts as a sense organ is constructed. Clearly, unless you are prepared to take into account a variety of different ways of describing this machine (in different reference frames) you can't make sense of it, and that is what I meant by saying we must look at a self-organizing system as many coexisting models, mechanical, electrical and so on—rather than a single model—such as we have in a computer simulation.

Of course you can always transform the physical mechanism into an equivalent " single " model if you are landed with components such as those in a computer which have well defined functions, for example, a collection of storage devices. But you can do this only if you are prepared to have an indefinitely large number of these well defined components because you are led to construct a growing model which, if left on its own, will expand indefinitely. Of course, in the case of a computer model the idea of rewarding by a supply of energy or food is inappropriate. You have instead a value function so defined that it costs something to take over more bits of store.

This is certainly equivalent, but in fact, when we are thinking of real live systems which we are going to use, for example, in controlling chemical processes, we are really much more interested in the finite systems which are rendered non-bounded by the interesting condition that they can alter their own relevance criteria, and in particular, by the expedient of building sense organs, can alter their relationship to the environment according to whether or not a trial relationship is rewarded.

If I can have a minute to comment here, it is interesting that in any such system you will find a hierarchy of " vestigial " mechanisms of just the sort Professor Bishop commented upon yesterday. In other words, a primitive mechanism will evolve for sensing a variable and the machine will then learn about this mechanism and it will evolve a more sophisticated one. At any stage

of the system's development we should be able to observe both mechanisms more or less active. One is inclined to say it is crazy, that this machine has two ways of doing the same thing. But on second thought it is natural for a system to learn from its past attempts and make improvements.

Editors
YOVITS
CAMERON

SELF-ORGANIZING SYSTEMS
PROCEEDINGS OF AN INTERDISCIPLINARY CONFERENCE

PERGAMON
PRESS

International Tracts in Computer Science and Technology and their application
VOLUME 2

SELF-ORGANIZING SYSTEMS

PROCEEDINGS OF AN INTERDISCIPLINARY CONFERENCE

Co-sponsored by
Information Systems Branch
of Office of Naval Research
and
Armour Research Foundation
of Illinois Institute of Technology

Editors
MARSHALL C. YOVITS
SCOTT CAMERON

Pergamon Press
OXFORD · LONDON · NEW YORK · PARIS

DURING the last decade or so there have been many great and important advances in the capabilities of information processing equipments and techniques. However, certain types of problems, mostly those involving inherently non-numerical types of information, can be solved efficiently only with the use of machines exhibiting a high degree of learning or self-organizing capability. Examples of problems of this type include automatic print reading, speech recognition, pattern recognition, automatic language translation, information retrieval, and control of large and complex systems.

SELF-ORGANIZING SYSTEMS contains the proceedings of a conference co-sponsored by the Information Systems Branch of the U.S. Office of Naval Research and Armour Research Foundation, and includes both the prepared papers and the associated discussion.

Fourteen formal papers were presented, the authors of which provided almost equal representation in the fields of Biology, Engineering, Mathematics, and Psychology. These papers fall naturally into four different interdisciplinary groups:

I. Perception of the Environment
II. Effects of Environmental Feedback
III. Learning in Finite Automata
IV. Structure of Self-Organizing Systems

This extremely topical and important volume, containing as it does a complete presentation of the existing research activities relating to the theory of self-organization, will be invaluable to the scientific community; it will encourage them in their efforts and provide a basis for additional research.

Already published in this series:

A PRACTICAL MANUAL ON THE MONTE CARLO METHOD FOR RANDOM WALK PROBLEMS. By E. D. Cashwell and C. J. Everett, 1959.

SELF-ORGANIZING SYSTEMS

PROCEEDINGS OF AN
INTERDISCIPLINARY CONFERENCE
5 and 6 May, 1959

Co-sponsored by

Information Systems Branch
of Office of Naval Research

and

Armour Research Foundation
of Illinois Institute of Technology

Editors

MARSHALL C. YOVITS
Office of Naval Research
CONFERENCE CHAIRMAN

SCOTT CAMERON
Armour Research Foundation
CONFERENCE SECRETARY

SYMPOSIUM PUBLICATIONS DIVISION
PERGAMON PRESS
OXFORD · LONDON · NEW YORK · PARIS
1960

PERGAMON PRESS LTD.
Headington Hill Hall, Oxford
4 & 5 Fitzroy Square, London W.1.

PERGAMON PRESS INC.
122 East 55th Street, New York 22, N.Y.
1404 New York Avenue N.W., Washington, D.C.

PERGAMON PRESS S.A.R.L.
24 Rue des Écoles, Paris Ve

PERGAMON PRESS G.m.b.H.
Kaiserstrasse 75, Frankfurt-am-Main

LIBRARY OF CONGRESS CARD NUMBER: 60-12574

Printed in Great Britain by Bell & Bain, Ltd., Glasgow

PREFACE

DURING the last decade or so there have been many great and important advances in the capabilities of information processing equipments and techniques. For the most part these have been based on the concept of the fixed stored program computer. This type of computer has shown itself able to solve large classes of problems which can first be put into some sort of numerical form and for which a program in machine language can be written.

Improvements in the speed, size, and sophistication of existing machines by factors of ten or a hundred appear to be near at hand, and a certain amount of research is presently directed toward the millimicrosecond or " ultimate " stored program computer. These advances will certainly open the way to making many more classes of problems conveniently tractable by digital computers.

However, it appears that certain types of problems, mostly those involving inherently non-numerical types of information, can be solved efficiently only with the use of machines exhibiting a high degree of learning or self-organizing capability. Examples of problems of this type include automatic print reading, speech recognition, pattern recognition, automatic language translation, information retrieval, and control of large and complex systems. Efficient solutions to problems of these types will probably require some combination of a fixed stored program computer and a self-organizing machine.

In recent months, it had become evident that interest in this field of cognitive systems was growing quite rapidly. A number of new groups had been formed to do research which would lead toward an understanding of these self-organizing systems. On the one hand the psychologist, the embryologist, the neurophysiologist and others involved in the life sciences were attempting to understand the self-organizing properties of biological systems, while mathematicians, engineers, and physical scientists were attempting to design artificial systems which could exhibit self-organizing properties.

Accordingly, the Information Systems Branch of the Office of Naval Research together with Armour Research Foundation, decided to sponsor a conference enabling the workers in the many disciplines involved to meet together to discuss their research activities and to explore common problems, mutual interests, and similar directions of research. This volume comprises the Proceedings of that Conference, including the prepared papers and the associated discussion. The name, Interdisciplinary Conference on Self-Organizing Systems, was chosen in order to emphasize deliberately the broad nature of the research necessary to pursue these endeavors.

The papers for the program were chosen entirely by invitation of the Conference Committee and were intended to be representative of the appropriate research activities in progress by workers in the various disciplines. Most of the speakers invited had already achieved reputations within their respective fields and were known rather widely within the scientific community. Fourteen formal papers were presented, the authors of which provided almost equal representation of the fields of Biology, Engineering, Mathematics, and Psychology. In addition, an after-dinner address was presented at the Conference banquet by Dr. A. M. Uttley of the National Physical Laboratory, Teddington, England. The welcoming address was delivered by Dr. H. A. Leedy, Director of the Armour Research Foundation, and the opening address of the Conference was delivered by Dr. F. J. Weyl, Research Director of the Office of Naval Research. Almost 400 people, divided among the various scientific, engineering, and social-scientific disciplines, attended the Conference.

There are, of course, in addition to those who participated in the Conference, many other research workers of equally great reputation who are contributing significantly to an understanding of self-organizing systems. The Committee deeply regrets the inability to include these people on the program. To the many people who kindly offered to present papers at the Conference, the Committee gives its sincere thanks as well as its apologies for the inability to include these contributions. The Conference could easily have been extended several more days with contributed papers, most of which would have been of very high quality. Perhaps it would have been

wiser to plan the Conference for more than the two days which were actually used. It was the Committee's feeling, however, that significantly more of the important interested people would be able to attend a two-day meeting than one of longer duration and that the fundamental purposes of the Conference would be best accomplished by the shorter meeting.

Those paper that were presented appeared to fall naturally into four different interdisciplinary Groups. These were, with the authors of the papers in the order presented:

I. *Perception of the Environment*
Farley; von Foerster; Estes; Rosenblatt.

II. *Effects of Environmental Feedback*
Auerbach; Goldman; Bishop.

III. *Learning in Finite Automata*
Newell, Shaw, Simon; Milner; Minsky; Campbell.

IV. *Structure of Self-Organizing Systems*
Pask; McCulloch; Burks.

The Committee wishes to thank each of these authors and co-authors, as well as Drs. Leedy, Weyl and Uttley, for participating in the Conference and helping to make it a success. Their assistance is greatly appreciated.

The papers in Groups I and II were presented during the first day under the Chairmanship of Dr. Otto M. Schmitt, Departments of Zoology and Physics, University of Minnesota, Minneapolis, Minnesota. The papers in Groups III and IV were presented during the second day under the Chairmanship of Dr. John McCarthy, Department of Mathematics, Massachusetts Institute of Technology, Cambridge, Massachusetts. The Committee wishes to express its thanks to Dr. Schmitt and Dr. McCarthy for the excellent way in which they handled the presentations and the subsequent discussions. They were instrumental in bringing out many interesting points that would not otherwise have arisen. The success of the Conference was due in large part to these Chairmen.

Each of the papers was scheduled for thirty minutes, with ten minutes allotted for questions. At the end of each day a panel, made up of the authors who spoke that day and the Chairman for the day, held an open discussion, with questions and comments being welcomed from the audience. It was apparent in the editing that many of the comments in the open discussions were directed at specific papers which had been presented that day and were therefore accordingly placed after the appropriate papers in the written proceedings in the interest of clarity for the reader.

All of the papers presented at the Conference are included in these Proceedings in the order presented, each followed by the appropriate discussion as explained above, with the exception of the paper entitled " Progress on the Advice Taker " by Dr. Marvin Minsky, Department of Mathematics, Massachusetts Institute of Technology, Cambridge, Massachusetts. Because of other comittments it was not possible for Dr. Minsky to submit his manuscript prior to publication of the Proceedings. It is expected that the paper will eventually be submitted to one of the scientific journals for publication.

The Committee which planned this Conference was made up of Marshall C. Yovits, Office of Naval Research, Chairman; Scott Cameron, Armour Research Foundation, Secretary; Albert R. Dawe, Office of Naval Research; Gordon D. Goldstein, Office of Naval Research; Harold Kantner, Armour Research Foundation; and Maynard Shelly, Office of Naval Research.

For the Committee,

MARSHALL C. YOVITS, Chairman
Head, Information Systems Branch
Office of Naval Research

CONTENTS