Classical

Principles of the self-organizing system

W. Ross Ashby

Originally published as Ashby, W. R. (1962). "Principles of the self-organizing system," in *Principles of Self-Organization: Transactions of the University of Illinois Symposium*, H. Von Foerster and G. W. Zopf, Jr. (eds.), Pergamon Press: London, UK, pp. 255-278. Reproduced with the kind permission of Ross Ashby's daughters, Sally Bannister, Ruth Pettit, and Jill Ashby. We would also like to thank John Ashby for his generous assistance in obtaining their permission.

The brilliant British psychiatrist, neuroscientist, and mathematician Ross Ashby was one of the pioneers in early and mid-phase cybernetics and thereby one of the leading progenitors of modern complexity theory. Not one to take either commonly used terms or popular notions for granted, Ashby probed deeply into the meaning of supposedly self-organizing systems. At the time of the following article, he had been working on a mathematical formalism of his *homeostat*, a hypothetical machine established on an axiomatic, set theoretical foundation that was supposed to offer a sufficient description of a living organism's learning and adaptive intelligence. Ashby's homeostat had a small number of essential variables serving to maintain its operation over a wide range of environmental conditions so that if the latter changed and thereby shifted the variables beyond the range where the homeostat could safely function, a new 'higher' level of the machine was activated in order to randomly reset the lower level's internal connections or organization (see Dupuy, 2000). Like the role of random mutations during evolution, if the new range set at random proved functional, the homeostat survived, otherwise it expired.

One of Ashby's goals was to repudiate that interpretation of the notion of self-organization, one commonly held to this day, which would have it that either a machine or a living organism could by itself change its own organization (or, in his phraseology, the functional mappings). For Ashby, self-organization in this sense was a bit of superfluous metaphysics since he believed not only could his formalism by itself completely delineate the homeostat's lower level organization, the adaptive novelty of his homeostat was purely the result of its upper level randomization that could reorganize the lower level and not some innate propensity for autonomous change. We offer Ashby's careful reasoning here as an enlightening guide for coming to terms with key ideas in complexity theory whose genuine significance lies less with facile bandying about and more with an intensive and extensive examination of the underlying assumptions.

Jeffrey Goldstein

Dupuy, J. (2000). *The Mechanization of the Mind*, Princeton: Princeton University Press.

W. ROSS ASHBY
University of Illinois

PRINCIPLES OF THE SELF-ORGANIZING SYSTEM*

Questions of principle are sometimes regarded as too unpractical to be important, but I suggest that that is certainly not the case in *our* subject. The range of phenomena that we have to deal with is so broad that, were it to be dealt with wholly at the technological or practical level, we would be defeated by the sheer quantity and complexity of it. The total range can be handled only piecemeal; among the pieces are those homomorphisms of the complex whole that we call "abstract theory" or "general principles". They alone give the bird's-eye view that enables us to move about in this vast field without losing our bearings. I propose, then, to attempt such a bird's-eye survey.

WHAT IS "ORGANIZATION"?

At the heart of our work lies the fundamental concept of "organization". What do we mean by it? As it is used in biology it is a somewhat complex concept, built up from several more primitive concepts. Because of this richness it is not readily defined, and it is interesting to notice that while March and Simon (1958) use the word "Organizations" as title for their book, they do not give a formal definition. Here I think they are right, for the word covers a multiplicity of meanings. I think that in future we shall hear the *word* less frequently, though the *operations* to which it corresponds, in the world of computers and brain-like mechanisms, will become of increasing daily importance.

The hard core of the concept is, in my opinion, that of "conditionality". As soon as the relation between two entities *A* and *B*

* The work on which this paper is based was supported by ONR Contract N 049–149.

becomes conditional on C's value or state then a necessary component of "organization" is present. Thus *the theory of organization is partly co-extensive with the theory of functions of more than one variable.*

We can get another angle on the question by asking "what is its converse?" The converse of "conditional on" is "not conditional on", so the converse of "organization" must therefore be, as the mathematical theory shows as clearly, the concept of "reducibility". (It is also called "separability".) This occurs, in mathematical forms, when what looks like a function of several variables (perhaps very many) proves on closer examination to have parts whose actions are *not* conditional on the values of the other parts. It occurs in mechanical forms, in hardware, when what looks like one machine proves to be composed of two (or more) sub-machines, each of which is acting independently of the others.

Questions of "conditionality", and of its converse "reducibility", can, of course, be treated by a number of mathematical and logical methods. I shall say something of such methods later. Here, however, I would like to express the opinion that the method of Uncertainty Analysis, introduced by Garner and McGill (1956), gives us a method for the treatment of conditionality that is not only completely rigorous but is also of extreme generality. Its great generality and suitability for application to complex behavior, lies in the fact that it is applicable to any arbitrarily defined set of states. Its application requires neither linearity, nor continuity, nor a metric, nor even an ordering relation. By this calculus, the *degree* of conditionality can be measured, and analyzed, and apportioned to factors and interactions in a manner exactly parallel to Fisher's method of the analysis of variance; yet it requires no metric in the variables, only the frequencies with which the various combinations of states occur. It seems to me that, just as Fisher's conception of the analysis of variance threw a flood of light on to the complex relations that may exist between variations on a metric, so McGill and Garner's conception of uncertainty analysis may give us an altogether better understanding of how to treat complexities of relation when the variables are non-metric. In psychology and biology such variables occur with great commonness; doubtless they will also occur commonly in the brain-like processes developing in computers. I look forward to the time when the methods of McGill and Garner will become the accepted language in which such matters are to be thought about and treated quantitatively.

The treatment of "conditionality" (whether by functions of many variables, by correlation analysis, by uncertainty analysis, or by other ways) makes us realize that the essential idea is that there is first a product space—that of the *possibilities*—within which some sub-set of points indicates the actualities. This way of looking at "conditionality" makes us realize that it is related to that of "communication"; and it is, of course, quite plausible that we should define parts as being "organized" when "communication" (in some generalized sense) occurs between them. (Again the natural converse is that of independence, which represents non-communication.)

Now "communication" from A to B necessarily implies some constraint, some correlation between what happens at A and what at B. If, for given event at A, all possible events may occur at B, then there is no communication from A to B and no constraint over the possible (A, B)-couples that can occur. Thus the presence of "organization" between variables is equivalent to the existence of a *constraint* in the product-space of the possibilities. I stress this point because while, in the past, biologists have tended to think of organization as something extra, something *added* to the elementary variables, the modern theory, based on the logic of communication, regards organization as a restriction or constraint. The two points of view are thus diametrically opposed; there is no question of either being exclusively right, for each can be appropriate in its context. But with this opposition in existence we must clearly go carefully, especially when we discuss with others, lest we should fall into complete confusion.

This excursion may seem somewhat complex but it is, I am sure, advisable, for we have to recognize that the discussion of organization theory has a peculiarity not found in the more objective sciences of physics and chemistry. The peculiarity comes in with the product space that I have just referred to. Whence comes this product space? Its chief peculiarity is that *it contains more than actually exists in the real physical world*, for it is the latter that gives us the actual, constrained *subset*.

18

The real world gives the subset of what *is*; the product space represents the uncertainty of the *observer*. The product space may therefore change if the observer changes; and two observers may legitimately use different product spaces within which to record the same subset of actual events in some actual thing. The "constraint" is thus a *relation* between observer and thing; the properties of any particular constraint will depend on both the real thing and on *the observer*. It follows that a substantial part of the theory of organization will be concerned with *properties that are not intrinsic to the thing but are relational between observer and thing*. We shall see some striking examples of this fact later.

WHOLE AND PARTS

"If conditionality" is an essential component in the concept of organization, so also is the assumption that we are speaking of a whole composed of parts. This assumption is worth a moment's scrutiny, for research is developing a theory of dynamics that does *not* observe parts and their interactions, but treats the system as an unanalysed whole (Ashby, 1958, a). In physics, of course, we usually start the description of a system by saying "Let the variables be $x_1, x_2, \ldots, x_n$" and thus start by treating the whole as made of n functional parts. The other method, however, deals with unanalysed states, $S_1, S_2, \ldots$ of the whole, without explicit mention of any parts that may be contributing to these states. The dynamics of such a system can then be defined and handled mathematically; I have shown elsewhere (Ashby, 1960, a) how such an approach can be useful. What I wish to point out here is that we can have a sophisticated *dynamics*, of a whole as complex and cross-connected as you please, that makes no reference to any parts and that therefore does *not* use the concept of organization. Thus the concepts of dynamics and of organization are essentially independent, in that all four combinations, of their presence and absence, are possible.

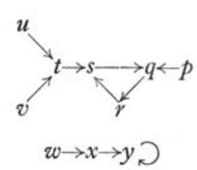
This fact exemplifies what I said, that "organization" is partly in the eye of the beholder. Two observers studying the same real material system, a hive of bees say, may find that one of them, thinking of the hive as an interaction of fifty thousand bee-parts, finds the bees "organized", while the other, observing whole states such as activity, dormancy, swarming, etc., may see *no* organization, only trajectories of these (unanalysed) states.

Another example of the independence of "organization" and "dynamics" is given by the fact that whether or not a real system is organized or reducible depends partly on the point of view taken by the observer. It is well known, for instance, that an organized (i.e. interacting) linear system of n parts, such as a network of pendulums and springs, can be seen from another point of view (that of the so-called "normal" coordinates) in which all the (newly identified) parts are completely separate, so that the whole is reducible. There is therefore nothing perverse about my insistence on the relativity of organization, for advantage of the fact is routinely taken in the study of quite ordinary dynamic systems.

Finally, in order to emphasize how dependent is the organization seen in a system on the observer who sees it, I will state the proposition that: given a whole with arbitrarily given behavior, a great variety of arbitrary "parts" can be seen in it; for all that is necessary, when the arbitrary part is proposed, is that we assume the given part to be coupled to another suitably related part, so that the two together form a whole isomorphic with the whole that was given. For instance, suppose the given whole, W of 10 states, behaves in accordance with the transformation:

$$W \quad \downarrow \begin{array}{cccccccccc} p & q & r & s & t & u & v & w & x & y \\ q & r & s & q & s & t & t & x & y & y \end{array}$$

Its kinematic graph is



and suppose we wish to "see" it as containing the part P, with internal states E and input states A:

$$\left.\begin{array}{cc|cc} & \downarrow & \multicolumn{2}{c}{E} \\ & & 1 & 2 \\ \hline A & 1 & 2 & 1 \\ & 2 & 1 & 1 \end{array}\right\} P$$

With a little ingenuity we find that if part P is coupled to part Q (with states (F, G) and input B) with transformation Q:

$$
Q: \quad
\begin{array}{cc|cccccc}
 & & & & (F, G) & & & \\
 & \downarrow & 1,1 & 1,2 & 1,3 & 2,1 & 2,2 & 2,3 \\
\hline
\multirow{2}{*}{B} & 1 & 2,1 & 1,2 & 1,2 & 2,1 & 1,2 & 1,2 \\
 & 2 & \cdot & 2,3 & \cdot & 2,1 & 2,2 & 2,2
\end{array}
$$

by putting $A = F$ and $B = E$, then the new whole W' has transformation

$$
W': \quad \downarrow \begin{array}{ccccc} 1,1,1 & 1,1,2 & 1,1,3 & 1,2,1, & \text{etc.} \\ 2,2,1 & 2,1,2 & 2,1,2 & 1,2,1, & \text{etc.} \end{array}
$$

which is *isomorphic with W* under the one–one correspondence

$$
\downarrow \begin{array}{ccccc} 1,1,1 & 1,1,2 & 1,1,3 & 1,2,1, & \text{etc.} \\ w & s & p & y & \text{, etc.} \end{array}
$$

Thus, subject only to certain requirements (e.g. that equilibria map into equilibria) *any dynamic system can be made to display a variety of arbitrarily assigned "parts"*, simply by a change in the *observer's* view point.

MACHINES IN GENERAL

I have just used a way of representing two "parts", "coupled" to form a "whole", that anticipates the question: what do we mean by a "machine" in general?

Here we are obviously encroaching on what has been called "general system theory", but this last discipline always seemed to me to be uncertain whether it was dealing with *physical* systems, and therefore tied to whatever the real world provides, or with mathematical systems, in which the sole demand is that the work shall be free from internal contradictions. It is, I think, one of the substantial advances of the last decade that we have at last identified the *essentials* of the "machine in general".

Before the essentials could be seen, we had to realize that two factors must be *excluded as irrelevant*. The first is "materiality"—the idea that a machine must be made of actual matter, of the hundred or so existent elements. This is wrong, for examples can readily be given (e.g. Ashby, 1958, a) showing that what is essential is whether the system, of angels and ectoplasm if you please, *behaves* in a law-abiding and machine-like way. Also to be excluded as irrelevant is any reference to energy, for any calculating machine shows that what matters is the *regularity* of the behavior—whether energy is gained or lost, or even created, is simply irrelevant.

The fundamental concept of "machine" proves to have a form that was formulated at least a century ago, but this concept has not, so far as I am aware, ever been used and exploited vigorously. A "machine" is that which behaves in a machine-like way, namely, that its internal state, and the state of its surroundings, defines uniquely the next state it will go to.

This definition, formally proposed fifteen years ago (Ashby, 1945) has withstood the passage of time and is now becoming generally accepted (e.g. Jeffrey, 1959). It appears in many forms. When the variables are continuous it corresponds to the description of a dynamic system by giving a set of ordinary differential equations with time as the independent variable. The *fundamental* nature of such a representation (as contrasted with a merely convenient one) has been recognized by many earlier workers such as Poincaré, Lotka (1925), and von Bertalanffy (1950 and earlier).

Such a representation by differential equations is, however, too restricted for the needs of a science that includes biological systems and calculating machines, in which discontinuity is ubiquitous. So arises the modern definition, able to include both the continuous and the discontinuous and even the discrete, without the slightest loss of rigor. The "machine with input" (Ashby, 1958, a) or the "finite automaton" (Jeffrey, 1959) is today defined by a set S of internal states, a set I of input or surrounding states, and a mapping, f say, of the product set $I \times S$ into S. Here, in my opinion, we have the very essence of the "machine"; all known types of machine are to be found here; and all interesting deviations from the concept are to be found by the corresponding deviation from the definition.

We are now in a position to say without ambiguity or evasion what we mean by a machine's "organization". First we specify which system we are talking about by specifying its states S and its conditions I. If S is a product set, so that $S = \Pi_i T_i$ say, then the parts i are each specified by its set of states T_i. *The "organization" between these parts is then specified by the mapping f.* Change f and the organization changes. In other words, the possible organizations between the parts can be set into one–one correspondence with the set of possible mappings of $I \times S$ into S. Thus "organization" and "mapping" are two ways of looking at the same thing—the organization being noticed by the observer of the actual system, and the mapping being recorded by the person who represents the behavior in mathematical or other symbolism.

"GOOD" ORGANIZATION

At this point some of you, especially the biologists, may be feeling uneasy; for this definition of organization makes no reference to any *usefulness* of the organization. It demands only that there be conditionality between the parts and regularity in behavior. In this I believe the definition to be right, for the question whether a given organization is "good" or "bad" is quite independent of the prior test of whether it is or is not an organization.

I feel inclined to stress this point, for here the engineers and the biologists are likely to think along widely differing lines. The engineer, having put together some electronic hardware and having found the assembled network to be roaring with parasitic oscillations, is quite accustomed to the idea of a "bad" organization; and he knows that the "good" organization has to be searched for. The biologist, however, studies mostly animal species that have survived the long process of natural selection; so almost all the organizations he sees have already been selected to be good ones, and he is apt to think of "organizations" as *necessarily* good. This point of view may often be true in the biological world but it is most emphatically not true in the world in which we people here are working. We *must* accept that

(1) most organizations are bad ones;

(2) the good ones have to be sought for; and

(3) what is meant by "good" must be clearly defined, explicitly if necessary, *in every case*.

What then is meant by "good", in our context of brain-like mechanisms and computers? We must proceed cautiously, for the word suggests some evaluation whose origin has not yet been considered.

In some cases the distinction between the "good" organization and the "bad" is obvious, in the sense that as everyone in these cases would tend to use the same criterion, it would not need explicit mention. The brain of a living organism, for instance, is usually judged as having a "good" organization if the organization (whether inborn or learned) acts so as to further the organism's survival. This consideration readily generalizes to all those cases in which the organization (whether of a cat or an automatic pilot or an oil refinery) is judged "good" if and only if it acts so as to keep an assigned set of variables, the "essential" variables, within assigned limits. Here are all the mechanisms for homeostasis, both in the original sense of Cannon and in the generalized sense. From this criterion comes the related one that an organization is "good" if it makes the system stable around an assigned equilibrium. Sommerhoff (1950) in particular has given a wealth of examples, drawn from a great range of biological and mechanical phenomena, showing how in all cases the idea of a "good organization" has as its essence the idea of a number of parts so interacting as to achieve some given "focal condition". I would like to say here that I do not consider that Sommerhoff's contribution to our subject has yet been adequately recognized. His identification of *exactly* what is meant by coordination and integration is, in my opinion, on a par with Cauchy's identification of exactly what was meant by convergence. Cauchy's discovery was a real discovery, and was an enormous help to later workers by providing them with a concept, rigorously defined, that could be used again and again, in a vast range of contexts, and always with exactly the same meaning. Sommerhoff's discovery of how to represent *exactly* what is meant by coordination and integration and good organization will, I am sure, eventually play a similarly fundamental part in our work.

His work illustrates, and emphasizes, what I want to say here—*there is no such thing as "good organization" in any absolute sense.* Always it is relative; and an organization that is good in one context or under one criterion may be bad under another.

Sometimes this statement is so obvious as to arouse no opposition. If we have half a dozen lenses, for instance, that can be assembled this way to make a telescope or that way to make a microscope, the goodness of an assembly obviously depends on whether one wants to look at the moon or a cheese mite.

But the subject is more contentious than that! The thesis implies that there is no such thing as a brain (natural or artificial) that is good in any absolute sense—it all depends on the circumstances and on what is wanted. Every faculty that a brain can show is "good" only conditionally, for there exists at least one environment against which the brain is handicapped by the possession of this faculty. Sommerhoff's formulation enables us to show this at once: whatever the faculty or organization achieves, let that be *not* in the "focal conditions".

We know, of course, lots of examples where the thesis is true in a somewhat trivial way. Curiosity tends to be good, but many an antelope has lost its life by stopping to see what the hunter's hat is. Whether the organization of the antelope's brain should be of the type that does, or does not, lead to temporary immobility clearly depends on whether hunters with rifles are or are not plentiful in its world.

From a different angle we can notice Pribram's results (1957), who found that brain-operated monkeys scored higher in a certain test than the normals. (The operated were plodding and patient while the normals were restless and distractible.) Be that as it may, one cannot say which brain (normal or operated) had the "good" organization until one has decided which sort of temperament is wanted.

Do you still find this non-contentious? Then I am prepared to assert that there is not a single mental faculty ascribed to Man that is good in the absolute sense. If any particular faculty is *usually* good, this is solely because our terrestrial environment is so lacking in variety that its usual form makes that faculty usually good. But change the environment, go to really different conditions, and possession of that faculty may be harmful. And "bad", by implication, is the brain organization that produces it.

I believe that there is not a single faculty or property of the brain, usually regarded as desirable, that does not become *un*desirable in some type of environment. Here are some examples in illustration.

The first is Memory. Is it not good that a brain should have memory? Not at all, I reply—only when the environment is of a type in which the future often *copies* the past; should the future often be the *inverse* of the past, memory is actually disadvantageous. A well known example is given when the sewer rat faces the environmental system known as "pre-baiting". The naïve rat is very suspicious, and takes strange food only in small quantities. If, however, wholesome food appears at some place for three days in succession, the sewer rat will learn, and on the fourth day will eat to repletion, and die. The rat without memory, however, is as suspicious on the fourth day as on the first, and lives. Thus, in *this* environment, memory is positively disadvantageous. Prolonged contact with this environment will lead, other things being equal, to evolution in the direction of diminished memory-capacity.

As a second example, consider organization itself in the sense of connectedness. Is it not good that a brain should have its parts in rich functional connection? I say, No—not *in general*; only when the environment is itself richly connected. When the environment's parts are *not* richly connected (when it is highly reducible, in other words), adaptation will go on faster if the brain is also highly reducible, i.e. if its connectivity is small (Ashby, 1960, d). Thus the *degree* of organization can be too high as well as too low; the degree we humans possess is probably adjusted to be somewhere near the optimum for the usual terrestrial environment. It does not in any way follow that this degree will be optimal or good if the brain is a mechanical one, working against some grossly non-terrestrial environment—one existing only inside a big computer, say.

As another example, what of the "organization" that the biologist always points to with pride—the development in evolution of specialized organs such as brain, intestines, heart and blood vessels. Is not this good? Good or not, it is certainly a specialization made possible only because the earth has an atmosphere; without it, we would be incessantly bombarded by tiny meteorites, any one of which, passing through our chest, might strike a large blood vessel and kill us. Under such conditions a better form for survival would be the slime mould, which specializes in being able to flow through a tangle of twigs without loss of function. Thus the development of organs is not good unconditionally, but is a specialization to a world free from flying particles.

After these actual instances, we can return to theory. It is here that Sommerhoff's formulation gives such helpful clarification. He shows that in all cases there must be given, and specified, first a *set of disturbances* (values of his "coenetic variable") and secondly a goal (his "focal condition"); the disturbances threaten to drive the outcome outside the focal condition. The "good" organization is then of the nature of a *relation* between the set of disturbances and the goal. Change the set of disturbances, and the organization, without itself changing, is evaluated "bad" instead of "good". As I said, there is no property of an organization that is good in any absolute sense; all are relative to some given environment, or to some given set of threats and disturbances, or to some given set of problems.

SELF-ORGANIZING SYSTEMS

I hope I have not wearied you by belaboring this relativity too much, but it is fundamental, and is only too readily forgotten when one comes to deal with organizations that are either biological in origin or are in imitation of such systems. With this in mind, we can now start to consider the so-called "self-organizing" system. We must proceed with some caution here if we are not to land in confusion, for the adjective is, if used loosely, ambiguous, and, if used precisely, self-contradictory.

To say a system is "self-organizing" leaves open two quite different meanings.

There is a first meaning that is simple and unobjectionable. This refers to the system that starts with its parts separate (so that the behavior of each is independent of the others' states) and whose parts then act so that they change towards forming connections of some type. Such a system is "self-organizing" in the sense that it changes from "parts separated" to "parts joined". An example is the embryo nervous system, which starts with cells having little or no effect on one another, and changes, by the growth of dendrites and formation of synapses, to one in which each part's behavior is very much affected by the other parts. Another example is Pask's system of electrolytic centers, in which the growth of a filament from one electrode is at first little affected by growths at the other electrodes; then the growths become more and more affected by one another as filaments approach the other electrodes. In general such systems can be more simply characterized as "self-*connecting*", for the change from independence between the parts to conditionality can always be seen as some form of "connection", even if it is as purely functional as that from a radio transmitter to a receiver.

Here, then, is a perfectly straightforward form of self-organizing system; but I must emphasize that there can be no assumption at this point that the organization developed will be a good one. If we wish it to be a "good" one, we must first provide a criterion for distinguishing between the bad and the good, and then we must ensure that the appropriate selection is made.

We are here approaching the second meaning of "self-organizing" (Ashby, 1947). "Organizing" may have the first meaning, just discussed, of "changing from unorganized to organized". But it may also mean "changing from a bad organization to a good one", and this is the case I wish to discuss now, and more fully. This is the case of peculiar interest to *us*, for this is the case of the system that changes itself from a bad way of behaving to a good. A well known example is the child that starts with a brain organization that makes it fire-seeking; then a change occurs, and a new brain organization appears that makes the child fire-avoiding. Another example would occur if an automatic pilot and a plane were so coupled, by mistake, that positive feedback made the whole error-aggravating rather than error-correcting. Here the organization is bad. The system would be "self-organizing" if a change were *automatically* made to the feedback, changing it from positive to negative; then the whole would have changed from a bad organization to a good. Clearly, *this* type of "self-organization" is of peculiar interest to us. What is implied by it?

Before the question is answered we must notice, if we are not to be in perpetual danger of confusion, that *no machine can be self-organizing in this sense*. The reasoning is simple. Define the set S of states so as to specify which machine we are talking about. The "organization" must then, as I said above, be identified with f, the mapping of S into S that the basic drive of the machine (whatever force it may be) imposes. Now the logical relation here is that f determines the changes of S:—f is *defined* as the set of couples (s_i, s_j) such that the internal drive of the system will force state s_i to change to s_j. To allow f to be a function of the state is to make nonsense of the whole concept.

Since the argument is fundamental in the theory of self-organizing systems, I may help explanation by a parallel example. Newton's law of gravitation says that $F = M_1M_2/d^2$, in particular, that the force varies inversely as the distance to power 2. To power 3 would be a different law. But suppose it were suggested that, not the force F but the *law* changed with the distance, so that the power was not 2 but some function of the distance, $\phi(d)$. This suggestion is illogical; for we now have that $F = M_1M_2/d^{\phi(d)}$, and this represents not a law that varies with the distance but *one* law covering all distances; that is, were this the case we would *re-define* the law. Analogously, were f in the machine to be some function of the state S, we would have to re-define our machine. Let me be quite explicit with an example. Suppose S had three states: a, b, c. If f depended on S there would be three f's: f_a, f_b, f_c say. Then if they are

$\downarrow$	a	b	c
f_a	**b**	a	b
f_b	c	**a**	a
f_c	b	b	**a**

then the transform of a must be under f_a, and is therefore b, so the whole set of f's would amount to the *single* transformation:

$$\downarrow \begin{matrix} a & b & c \\ b & a & a \end{matrix}$$

It is clearly illogical to talk of f as being a function of S, for such talk would refer to operations, such as $f_a(b)$, which cannot in fact occur.

If, then, no machine can properly be said to be self-organizing, how do we regard, say, the Homeostat, that rearranges its own wiring; or the computer that writes out its own program?

The new logic of mechanism enables us to treat the question rigorously. We start with the set S of states, and assume that f changes, to g say. So we really have a *variable*, $\alpha(t)$ say, a function of the time that had at first the value f and later the value g. This change, as we have just seen, cannot be ascribed to any cause in the set S; so it must have come from some outside agent, acting on the system S as input. If the system is to be in some sense "*self*-organizing", the "self" must be enlarged to include this variable α, and, to keep the whole bounded, the cause of α's change must be in S (or α).

Thus the appearance of being "self-organizing" can be given only by the machine S being coupled to another machine (of one part):

$$\boxed{S} \rightleftarrows \boxed{\alpha}$$

Then the part S can be "self-organizing" within the whole $S+\alpha$.

Only in this partial and strictly qualified sense can we understand that a system is "*self*-organizing" without being self-contradictory.

Since no system can correctly be said to be self-organizing, and since use of the phrase "self-organizing" tends to perpetuate a fundamentally confused and inconsistent way of looking at the subject, the phrase is probably better allowed to die out.

THE SPONTANEOUS GENERATION OF ORGANIZATION

When I say that no system can properly be said to be self-organizing, the listener may not be satisfied. What, he may ask, of those changes that occurred a billion years ago, that led lots of carbon atoms, scattered in little molecules of carbon dioxide, methane, carbonate, etc., to get together until they formed proteins, and then went on to form those large active lumps that today we call "animals"? Was not this process, on an isolated planet, one of "self-organization"? And if it occurred on a planetary surface can it not be made to occur in a computer? I am, of course, now discussing the origin of life. Has modern system theory anything to say on this topic?

It has a great deal to say, and some of it flatly contradictory to what has been said ever since the idea of evolution was first considered. In the past, when a writer discussed the topic, he usually assumed that the generation of life was rare and peculiar, and he then tried to display some way that would enable this rare and peculiar event to occur. So he tried to display that there is *some* route from, say, carbon dioxide to the amino acid, and thence to the protein, and so, through natural selection and evolution, to intelligent beings. I say that this looking for special conditions is quite wrong. The truth is the opposite—*every* dynamic system generates its own form of intelligent life, is self-organizing in this sense. (I will demonstrate the fact in a moment.) Why we have failed to recognize this fact is that until recently we have had no experience of systems of medium complexity; either they have been like the watch and the pendulum, and we have found their properties few and trivial, or they have been like the dog and the human being, and we have found their properties so rich and remarkable that we have thought them supernatural. Only in the last few years has the general-purpose computer given us a system rich enough to be interesting yet still simple enough to be understandable. With this machine as tutor we can now begin to think about systems that are simple enough to be comprehensible in detail yet also rich enough to be suggestive. With their aid we can see the truth of the statement that *every isolated determinate dynamic system obeying unchanging laws will develop "organisms" that are adapted to their "environments"*.

The argument is simple enough in principle. We start with the fact that systems in general go to equilibrium. Now most of a system's states are non-equilibrial (if we exclude the extreme case of the system in neutral equilibrium). So in going from *any* state to one of the equilibria, the system is going from a larger number of states to a smaller. In this way it is performing a selection, in the purely objective sense that it rejects some states, by leaving them, and retains some other state, by sticking to it. Thus, as every determinate system goes to equilibrium, so does it select. We have heard *ad nauseam* the dictum that a machine cannot select; the truth is just the opposite: every machine, as it goes to equilibrium, performs the corresponding act of selection.

Now, equilibrium in simple systems is usually trivial and uninteresting; it is the pendulum hanging vertically; it is the watch with its main-spring run down; the cube resting flat on one face. Today, however, we know that when the system is more complex and dynamic, equilibrium, and the stability around it, can be much more interesting. Here we have the automatic pilot successfully combating an eddy; the person redistributing his blood flow after a severe haemorrhage; the business firm restocking after a sudden increase in consumption; the economic system restoring a distribution of supplies after a sudden destruction of a food crop; and it is a man successfully getting at least one meal a day during a lifetime of hardship and unemployment.

What makes the change, from trivial to interesting, is simply the *scale* of the events. "Going to equilibrium" *is* trivial in the simple pendulum, for the equilibrium is no more than a single point. But when the system is more complex; when, say, a country's economy goes back from wartime to normal methods then the stable region is vast, and much interesting activity can occur within it. The computer is heaven-sent in this context, for it enables us to bridge the enormous conceptual gap from the simple and understandable to the complex and interesting. Thus we can gain a considerable insight into the so-called spontaneous generation of life by just seeing how a somewhat simpler version will appear in a computer.

COMPETITION

Here is an example of a simpler version. The competition between species is often treated as if it were essentially biological; it is in fact an expression of a process of far greater generality. Suppose we have a computer, for instance, whose stores are filled at random with the digits 0 to 9. Suppose its dynamic law is that the digits are continuously being multiplied in pairs, and the right-hand digit of the product going to replace the first digit taken. Start the machine, and let it "evolve"; what will happen? Now under the laws of this particular world, even times even gives even, and odd times odd gives odd. But even times odd gives even; so after a mixed encounter *the even has the better chance of survival.* So as this system evolves, we shall see the evens favored in the struggle, steadily replacing the odds in the stores and eventually exterminating them.

But the evens are not homogeneous, and among them the zeros are best suited to survive in this particular world; and, as we watch, we shall see the zeros exterminating their fellow-evens, until eventually they inherit this particular earth.

What we have here is an example of a thesis of extreme generality. From one point of view we have simply a well defined operator (the multiplication and replacement law) which drives on towards equilibrium. In doing so it *automatically* selects those operands that are *specially resistant* to its change-making tendency (for the zeros are uniquely resistant to change by multiplication). This process, of progression towards the specially resistant form, is of extreme generality, demanding only that the operator (or the physical laws of any physical system) be determinate and unchanging. This is the general or abstract point of view. The biologist sees a special case of it when he observes the march of evolution, survival of the fittest, and the inevitable emergence of the highest biological functions and intelligence. Thus, when we ask: What was necessary that life and intelligence should appear? the answer is not carbon, or amino acids or any other special feature but only that the dynamic laws of the process should be *unchanging*, i.e. that the system should be *isolated. In any isolated system, life and intelligence inevitably develop* (they may, in degenerate cases, develop to only zero degree).

So the answer to the question: How can we generate intelligence synthetically? is as follows. Take a dynamic system whose laws are unchanging and single-valued, and whose size is so large that after it has gone to an equilibrium that involves only a small fraction of its total states, this small fraction is still large enough to allow room for a good deal of change and behavior. Let it go on for a long enough time to get to such an equilibrium. Then examine the equilibrium in detail. You will find that the states or forms now in being are peculiarly able to survive against the changes induced by the laws. Split the equilibrium in two, call one part "organism" and the other part "environment": you will find that this "organism" is peculiarly able to survive against the disturbances from this "environment". The *degree* of adaptation and complexity that this organism can develop is bounded only by the size of the whole dynamic system and by the time over which it is allowed to progress towards equilibrium. Thus, as I said, every isolated determinate dynamic system will develop organisms that are adapted to their environments. There is thus no difficulty in principle, in developing synthetic organisms as complex or as intelligent as we please.

In *this* sense, then, *every* machine can be thought of as "self-organizing", for it will develop, to such degree as its size and complexity allow, some functional structure homologous with an "adapted organism". But does this give us what we at this Conference are looking for? Only partly; for nothing said so far has any implication about the organization being good or bad; the criterion that would make the distinction has not yet been introduced. It is true, of course, that the developed organism, being stable, will have its own essential variables, and it will show its stability by vigorous reactions that tend to preserve its own existence. To *itself*, its own organization will *always*, by definition, be good. The wasp finds the stinging reflex a good thing, and the leech finds the blood-sucking reflex a good thing. But these criteria come *after* the organization for survival; having seen *what* survives we then see what is "good" for that form. What emerges depends simply on what are the system's laws and from what state it started; there is no implication that the organization developed will be "good" in any absolute sense, or according to the criterion of any outside body such as ourselves.

To summarize briefly: there is no difficulty, in principle, in developing *synthetic organisms as complex, and as intelligent as we please*. But we must notice two fundamental qualifications; first, their intelligence will be an adaptation to, and a specialization towards, their particular environment, with no implication of validity for any other environment such as ours; and secondly, their intelligence will be directed towards keeping their own essential variables within limits. They will be fundamentally selfish. So we now have to ask: In view of these qualifications, can we yet turn these processes to our advantage?

REQUISITE VARIETY

In this matter I do not think enough attention has yet been paid to Shannon's Tenth Theorem (1949) or to the simpler "law of requisite variety" in which I have expressed the same basic idea (Ashby, 1958, a). Shannon's theorem says that if a correction-channel has capacity H, then equivocation of amount H can be removed, *but no more*. Shannon stated his theorem in the context of telephone or similar communication, but the formulation is just as true of a biological regulatory channel trying to exert some sort of corrective control. He thought of the case with a lot of message and a little error; the biologist faces the case where the "message" is small but the disturbing errors are many and large. The theorem can then be applied to the brain (or any other regulatory and selective device), when it says that the amount of regulatory or selective action that the brain can achieve is absolutely bounded by its capacity as a channel (Ashby, 1958, b). Another way of expressing the same idea is to say that any quantity K of appropriate selection demands the transmission or processing of quantity K of information (Ashby, 1960, b.) *There is no getting of selection for nothing*.

I think that here we have a principle that we shall hear much of in the future, for it dominates all work with complex systems. It enters the subject somewhat as the law of conservation of energy enters power engineering. When that law first came in, about a hundred years ago, many engineers thought of it as a disappointment, for it stopped all hopes of perpetual motion. Nevertheless, it did in fact lead to the great practical engineering triumphs of the nineteenth century, because it made power engineering more realistic.

I suggest that when the full implications of Shannon's Tenth Theorem are grasped we shall be, first sobered, and then helped, for we shall then be able to focus our activities on the problems that are properly realistic, and actually solvable.

THE FUTURE

Here I have completed this bird's-eye survey of the principles that govern the self-organizing system. I hope I have given justification for my belief that these principles, based on the logic of mechanism and on information theory, are now essentially *complete*, in the sense that there is now no area that is grossly mysterious.

Before I end, however, I would like to indicate, very briefly, the directions in which future research seems to me to be most likely to be profitable.

One direction in which I believe a great deal to be readily discoverable, is in the discovery of new types of dynamic process. Most of the machine-processes that we know today are very specialized, depending on exactly what parts are used and how they are joined together. But there are systems of more net-like construction in which what happens can only be treated statistically. There are processes here like, for instance, the spread of epidemics, the fluctuations of animal populations over a territory, the spread of wave-like phenomena over a nerve-net. These processes are, in themselves, neither good nor bad, but they exist, with all their curious properties, and doubtless the brain will use them should they be of advantage. What I want to emphasize here is that they often show very surprising and peculiar properties; such as the tendency, in epidemics, for the outbreaks to occur in waves. Such peculiar new properties may be just what some machine designer wants, and that he might otherwise not know how to achieve.

The study of such systems must be essentially statistical, but this does not mean that each system must be individually stochastic. On the contrary, it has recently been shown (Ashby, 1960, c) that no system can have greater efficiency than the determinate when acting as a regulator; so, as regulation is the one function that counts biologically, we can expect that natural selection will have made the brain as determinate as possible. It follows that we can confine our interest to the lesser range in which the sample space is over a set of mechanisms each of which is individually determinate.

As a particular case, a type of system that deserves much more thorough investigation is the large system that is built of parts that have many states of equilibrium. Such systems are extremely common in the terrestrial world; they exist all around us, and in fact, intelligence as we know it would be almost impossible otherwise (Ashby, 1960, d). This is another way of referring to the system whose variables behave largely as part-functions. I have shown elsewhere (Ashby, 1960, a) that such systems tend to show habituation (extinction) and to be able to adapt progressively (Ashby, 1960, d). There is reason to believe that some of the well-known but obscure biological phenomena such as conditioning, association, and Jennings' (1906) law of the resolution of physiological states may be more or less simple and direct expressions of the multiplicity of equilibrial states. At the moment I am investigating the possibility that the transfer of "structure", such as that of three-dimensional space, into a dynamic system—the sort of learning that Piaget has specially considered—may be an *automatic* process when the input comes to a system with many equilibria. Be that as it may, there can be little doubt that the study of such systems is likely to reveal a variety of new dynamic processes, giving us dynamic resources not at present available.

A particular type of system with many equilibria is the system whose parts have a high "threshold"—those that tend to stay at some "basic" state unless some function of the input exceeds some value. The general properties of such systems is still largely unknown, although Beurle (1956) has made a most interesting start. They deserve extensive investigation; for, with their basic tendency to develop avalanche-like waves of activity, their dynamic properties are likely to prove exciting and even dramatic. The fact that the mammalian brain uses the property extensively suggests that it may have some peculiar, and useful, property not readily obtainable in any other way.

Reference to the system with many equilibria brings me to the second line of investigation that seems to me to be in the highest degree promising—I refer to the discovery of *the living organism's memory store*: the identification of its physical nature.

At the moment, our knowledge of the living brain is grossly out of balance. With regard to what happens from one millisecond to the next we know a great deal, and many laboratories are working to add yet more detail. But when we ask what happens in the brain from one hour to the next, or from one year to the next, practically nothing is known. Yet it is these longer-term changes that are the really significant ones in human behavior.

It seems to me, therefore, that if there is one thing that is crying out to be investigated it is the physical basis of the brain's memory-stores. There was a time when "memory" was a very vague and metaphysical subject; but those days are gone. "Memory", as a *constraint* holding over events of the past and the present, and a *relation* between them, is today firmly grasped by the logic of mechanism. We know exactly what we mean by it behavioristically and operationally. What we need now is the provision of adequate resources for its investigation. Surely the time has come for the world to be able to find resources for *one* team to go into the

REFERENCES

1. W. Ross Ashby, The physical origin of adaptation by trial and error, *J. Gen. Psychol.* **32**, pp. 13–25 (1945).
2. W. Ross Ashby, Principles of the self-organizing dynamic system. *J. Gen. Psychol.* **37**, pp. 125–8 (1947).
3. W. Ross Ashby, *An Introduction to Cybernetics*, Wiley, New York, 3rd imp. (1958, a).
4. W. Ross Ashby, Requisite variety and its implications for the control of complex systems, *Cybernetica*, **1**, pp. 83–99 (1958, b).
5. W. Ross Ashby, The mechanism of habituation. In: *The Mechanization of thought Processes*. (Natl. Phys. Lab. Symposium No. 10) H.M.S.O., London (1960).
6. W. Ross Ashby, Computers and decision-making, *New Scientist*, **7**, p. 746 1960, b).
7. W. Ross Ashby, The brain as regulator, *Nature*, *Lond.* **186**, p. 413 (1960, c).
8. W. Ross Ashby, *Design for a Brain; the Origin of Adaptive Behavior*, Wiley, New York, 2nd ed. (1960, d).
9. L. von Bertalanffy, An outline of general system theory, *Brit. J. Phil. Sci.* **1**, pp. 134–65 (1950).
10. R. L. Beurle, Properties of a mass of cells capable of regenerating pulses, *Proc. Roy. Soc.* B**240**, pp. 55–94 (1956).
11. W. R. Garner and W. J. McGill, The relation between information and variance analyses, *Psychometrika* **21**, pp. 219–28 (1956).
12. R. C. Jeffrey, Some recent simplifications of the theory of finite automata. Technical Report 219, Research Laboratory of Electronics, Massachusetts Institute of Technology (27 May 1959).
13. H. S. Jennings, *Behavior of the Lower Organisms*, New York (1906).
14. A. J. Lotka, *Elements of Physical Biology*, Williams & Wilkins, Baltimore (1925).
15. J. G. March and J. A. Simon, *Organizations*, Wiley, New York (1958).
16. K. H. Pribram, Fifteenth International Congress of Psychology, Brussels (1957).
17. C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication*, University of Illinois Press, Urbana (1949).
18. G. Sommerhoff, *Analytical Biology*, Oxford University Press, London (1950).