PRINCIPLES OF THE SELF-ORGANIZING SYSTEM

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 "conditional". As soon as the relation between two entities A and B 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.
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.
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 $x_1, x_2, ..., x_n$" and thus start by treating the whole as made of $n$ functional parts. The other method, however, deals with unanalysed states, $S_1, S_2, ...$ 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:

\[
W \downarrow p q r s t u v w x y
\]

\[
q r s q s t t x y y
\]

Its kinematic graph is

\[
\begin{array}{c}
u \\
\begin{array}{c}
\text{t} \\
\text{v}
\end{array}
\begin{array}{c}
\rightarrow q \\
\leftarrow p
\end{array}
\end{array}
\]

\[
w \rightarrow x \rightarrow y
\]

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

\[
\begin{array}{c|cc}
E & 1 & 2 \\
\hline
1 & & \\
2 & & \\
\hline
A & 1 & 2 \\
2 & 1 & 1
\end{array}
\]

\[
\begin{array}{c|ccc}
P & & & \\
\hline
1 & & & \\
2 & & & \\
1 & & & \\
1 & & & \\
\end{array}
\]
With a little ingenuity we find that if part $P$ is coupled to part $Q$ (with states $(F, G)$ and input $B$) with transformation $Q$:

$$(F, G)$$

\[
\begin{array}{c|ccccccc}
\downarrow & 1,1 & 1,2 & 1,3 & 2,1 & 2,2 & 2,3 \\
B & 1 & 2,1 & 1,2 & 1,2 & 2,1 & 1,2 & 1,2 \\
2 & \cdot & 2,3 & \cdot & 2,1 & 2,2 & 2,2 \\
\end{array}
\]

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

$$W': \quad \downarrow \quad 1,1,1 \quad 1,1,2 \quad 1,1,3 \quad 1,2,1, \text{ etc.}$$
\[
2,2,1 \quad 2,1,2 \quad 2,1,2 \quad 1,2,1, \text{ etc.}
\]

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

$$\quad \downarrow \quad 1,1,1 \quad 1,1,2 \quad 1,1,3 \quad 1,2,1, \text{ etc.}$$
\[
\quad w \quad s \quad p \quad y \quad \text{ etc.}
\]

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 viewpoint.*

**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.

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.

**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 $(s_t, s_f)$ such that the internal drive of the system will
force state $s_t$ to change to $s_f$. To allow $f$ to be a function of the
state is to make nonsense of the whole concept.

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, $z(t)$ 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 $z$, 
and, to keep the whole bounded, the cause of $z$'s change must be
in $S$ (or $z$).
Thus the appearance of being "self-organizing" can be given only by the machine $S$ being coupled to another machine (of one part):

$$\begin{array}{c}
S \\
\rightarrow \\
\leftarrow \\
\alpha
\end{array}$$

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

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.

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.
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.

Today, the principles of the self-organizing system are known with some completeness, in the sense that no major part of the subject is wholly mysterious.

We have a secure base. Today we know exactly what we mean by "machine", by "organization", by "integration", and by "self-organization". We understand these concepts as thoroughly and as rigorously as the mathematician understands "continuity" or "convergence".

In these terms we can see today that the artificial generation of dynamic systems with "life" and "intelligence" is not merely simple—it is unavoidable if only the basic requirements are met. These are not carbon, water, or any other material entities but the persistence, over a long time, of the action of any operator that is both unchanging and single-valued. Every such operator forces the development of its own form of life and intelligence.
REFERENCES