William Ross Ashby
Feedback, Adaptation and Stability
Selected Passages from *Design for a Brain*
(The origin of adaptive behaviour)

(1960)

Note

These are selected passages from Ross Ashby classic text *Design for a Brain* (1960).

They represent a mine of intuitions and data that could be applied to the social environment (e.g. society as a brain in constant search of dynamic balance). This mental exercise would highlight the past and current failings in adapting to the requirements of the environment by any centralised ruler (e.g. the central state, the central bank, the central planner, etc.) intent on ignoring reality and the limits imposed by reality and in pursuit of extravagant power and riches. Being this the actual case, the final result is likely to be a disastrous feedback that amplifies disequilibria and plunges everybody into a protracted depression or even a never ending decadence; unless we understand how the brain-human being works and we are willing to put again nature and human nature as the central focus of our thinking and caring.

Stability

The words 'stability', 'steady state', and 'equilibrium' are used by a variety of authors with a variety of meanings, though there is always the same underlying theme.

The subject may be opened by a presentation of the three standard elementary examples. A cube resting with one face on a horizontal surface typifies 'stable' equilibrium; a sphere resting on a horizontal surface typifies 'neutral' equilibrium; and a balanced cone on its point typifies 'unstable' equilibrium. With neutral and unstable equilibria we shall have little concern but the concept of 'stable equilibrium' will be used repeatedly.

An important feature of stability is that it does not refer to a material body or 'machine' but only to some aspect of it. This statement may be proved most simply by an example showing that a single material body can be in two different equilibrial conditions at the same time. Consider a square card balanced exactly on one edge; to displacements at right angles to this edge the card is unstable; to displacements exactly parallel to this edge it is, theoretically at least, stable.

We conclude tentatively that the concept of 'stability' belongs not to a material body but to a field. It is shown by a field if the lines of behaviour converge.

More important is the underlying theme that in all cases the stable system is characterised by the fact that after a displacement we can assign some *bound* to the subsequent movement of the representative point, whereas in the unstable system such limitation is either impossible or depends on facts outside the subject of discussion. Thus, if a thermostat is set at 37° C. and displaced to 40°, we can predict that in the future it will not go outside specified limits, which might be, in one apparatus, 36° and 40°. On the other hand, if the thermostat has been assembled with a component reversed so that it is unstable and if it is displaced to 40°, then we can give no limits to its subsequent temperatures; unless we introduce such new topics as the melting-point of its solder.

These considerations bring us to the definition. Given the field of a state-determined system and a region in the field, the region is **stable** if the lines of behaviour from all point in the region stay within the region.

Feedback

Such systems whose variables affect one another in one or more circuits possess what the radio-engineer calls 'feedback', they are also sometimes described as 'servo-mechanisms'.

The nature, degree, and polarity of the feedback usually have a decisive effect on the stability or instability of the system.

Instability in such systems is shown by the development of a 'runaway'. The least disturbance is magnified by its passage round the circuit so that it is incessantly built up into a larger and larger deviation from the central state. The phenomenon is identical with that referred to as a 'vicious circle'.

Goal-seeking

Every stable system has the property that if displaced from a state of equilibrium and released, the subsequent movement is so matched to the initial displacement that the system is brought back to the state of equilibrium. *A variety of disturbances will therefore evoke a variety of matched reactions* .

Thus machines with feedback are not subject to the oft-repeated dictum that machines must act blindly and cannot correct their errors. Such a statement is true of machines without feedback, but not of machines in general.

Once it is appreciated that feedback can be used to correct any deviation we like, it is easy to understand that there is no limit to the complexity of goal-seeking behaviour which may occur in machines quite devoid of any 'vital' factor.

A system with feedback may be both wholly automatic and yet actively and complexly goal-seeking. There is no incompatibility.

It will be noticed that stability, as defined, in no way implies fixity or rigidity. It is true that the stable system usually has a state of equilibrium at which it shows no change; but the lack of change is deceptive if it suggests rigidity: if displaced from the state of equilibrium it will show

active, perhaps extensive and complex movements. The stable system is restricted only in that it does not show the unrestricted divergencies of instability.

Stability and the whole

An important feature of a system's stability (or instability) is that *it is a property of the whole system and can be assigned to no part of it*

.

The stability belongs only to the combination; it cannot be related to the parts considered separately.

The fact that the stability of the system is a property of the system as a whole is related to the fact that the presence of stability always implies some co-ordination of the actions between the parts.

As the system and the feedbacks become more complex, so does the achievement of stability become more difficult and the likelihood of instability greater.

Adaptation as Stability : Homeostasis

The concept of 'adaptation' has so far been used without definition; this vagueness must be corrected.

I propose the definition that *a form of behaviour is adaptive if it maintains the essential variables within physiological limits* .
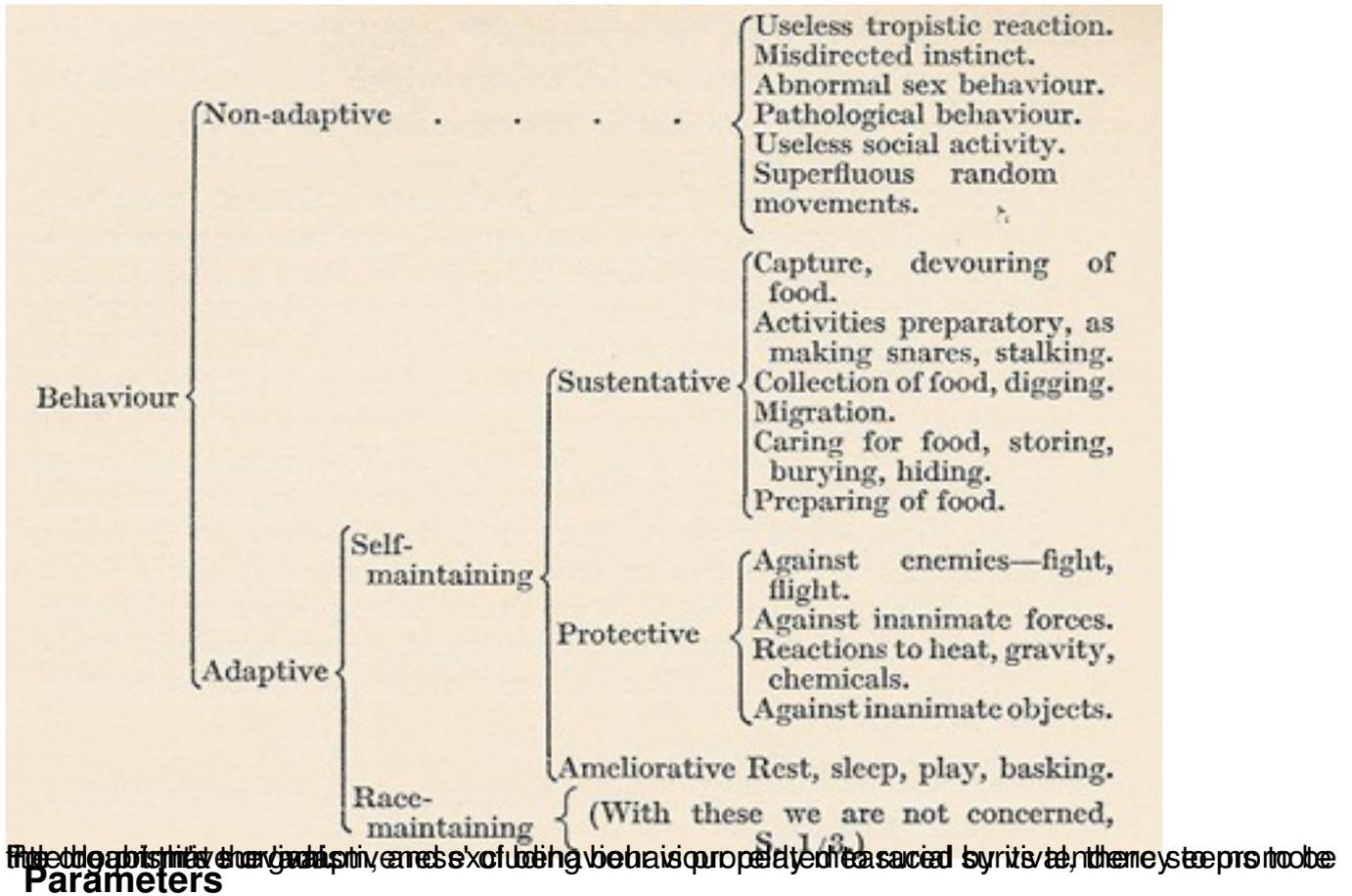
(1) Each mechanism is 'adapted' to its end.

(2) Its end is the maintenance of the values of some essential variables within physiological limits.

(3) Almost all the behaviour of an animal's vegetative system is due to such mechanism.

We can now recognise that 'adaptive' behaviour is equivalent to the behaviour of a stable system, the region of the stability being the region of the phase-space in which all the essential variables lie within their normal limits.

Survival

In order to survey rapidly the types of behaviour of the more primitive animals, we may examine the classification of Holmes (S. J. Holmes, A tentative classification of the forms of animal behaviour, *Journal of comparative Psychology*, 2, 173, 1922) who intends his list to be exhaustive but constructed it with no reference to the concept of stability. The reader will be able to judge how far our formulations is consistent with his scheme.

Non-adaptive . . . . {
Useless tropistic reaction.
Misdirected instinct.
Abnormal sex behaviour.
Pathological behaviour.
Useless social activity.
Superfluous random movements.

Behaviour {

Adaptive {

Self-maintaining {

Sustentative {
Capture, devouring of food.
Activities preparatory, as making snares, stalking.
Collection of food, digging.
Migration.
Caring for food, storing, burying, hiding.
Preparing of food.

Protective {
Against enemies—fight, flight.
Against inanimate forces.
Reactions to heat, gravity, chemicals.
Against inanimate objects.

Ameliorative Rest, sleep, play, basking.

Race-maintaining {
(With these we are not concerned,

Parameters

Given a system, a variable not included in it is a **parameter**. The word *variable* will, from now on, be reserved for one within the system.

A change in the value of an effective parameter changes the line of behaviour from each state. From this follows at once: *a change in the value of an effective parameter changes the field*.

From this follows the important quantitative relation: a system can, in general, show as many fields as its parameters can show combinations of values.

The importance of distinguishing between change of a variable and change of a parameter, that is between change of state and change of field, can hardly be over-estimated. It is this distinction that will enable us to avoid the confusion between those changes that *are* behaviour and changes that occur from one behaviour to another.

Parameters and stability

We now reach the main point. Because a change of parameter-value changes the field, and because a system's stability depends on its field, *a change of parameter-value will in general change a system's stability in some way*   .

Here we need only the relationship which is reciprocal: in a state-determined system, *a change of stability can only be due to change of value of a parameter*
, and
change of value of a parameter causes a change in stability
.

The ultrastable system

To be adapted, the organism, guided by information from the environment, must control its essential variables, forcing them to go within the proper limits, by so manipulating the environment (through its motor control of it) that the environment then acts on them appropriately. Thus the diagram of immediate effects of this process is
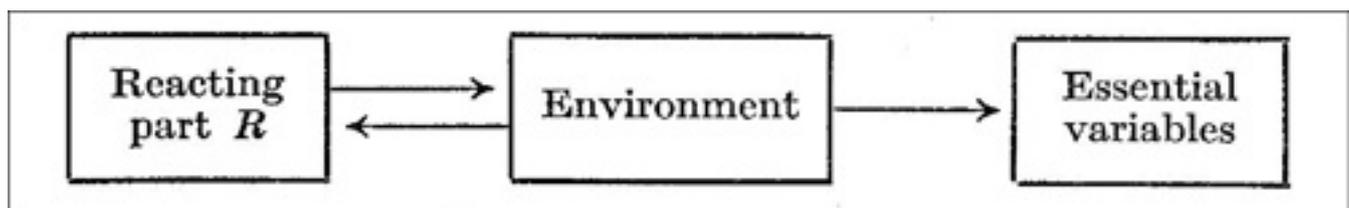
In the case we are considering, the reacting part R is not specially related or adjusted to what is in the environment and how it is joined to the essential variables. Thus the reacting part R can be thought of as an organism trying to control the output of a Black Box (the environment), the content of which is unknown to it.

It is axiomatic (for *any* Black Box when the range of its inputs is *given*) that the only way in which the nature of its contents can be elicited is by the transmission of actions through it. This means that input-values must be given, output-values observed, and the relationships in the paired values noticed. In the kitten's case this means that the kitten must
do
various things to the environment and must later act in accordance with how these actions affected the essential variables. In other words, it must proceed by trial and error.

Trail and error

Adaptation by trial and error is sometimes treated in psychological writings as if it were merely one way of adaptation, and an inferior way at that. The argument given above shows that the method of trial and error holds a much more fundamental place in the methods of adaptation. The argument shows, in fact, that when the organism has to adapt (to get its essential variables within physiological limits, by working through an environment that is of the nature of the Black Box, then the process of trial and error is *necessary*, for only such a process can elicit the required information.

The process of trial and error can thus be viewed from two very different points of view. On the one hand it can be regarded as simply an attempt at success; so that when it fails we give zero marks for success. From this point of view it is merely a second-rate way of getting to success. There is, however, the other point of view that gives it an altogether higher status, for the process may be playing the invaluable part of *gathering information*, information that is absolutely necessary if adaptation is to be successfully achieved.

The organism that can adapt thus has a motor output to the environment and *two* feedback loops. The first loop consists of the ordinary sensory input from eye, ear, joints, etc. giving the organism non-affective information about the world around it. The second feedback goes through the essential variables; it carries information about whether the essential variables are or are not driven outside the normal limits, and it acts on the parameters S. The first feedback plays its part   *w
ithin*
each reaction; the second determines which reaction shall occur.
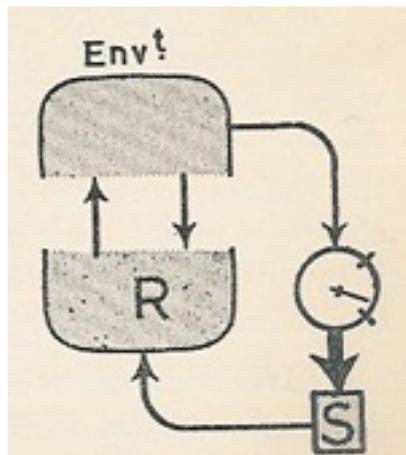
Envt = Environment ; R = Reacting Part ; S = Parameters

The basic rule for adaptation by trial and error is: If the trial is unsuccessful change the way of behaving; when and only when it is successful, retain the way of behaving.

Training

The process of 'training' will now be shown in its relation to ultrastability.

All training involves some use of 'punishment' or 'reward'.

Some apparent faults

Even if the ultrastable system is suitable arranged - if the critical states are encountered before

the essential variables reach their limits - it usually cannot adapt to an environment that behaves with sudden discontinuities.

Another weakness shown by the ultrastable system's method is that success is dependent on the system's using a suitable period of delay between each trial.

Both extremes of delay may be fatal: too hurried a change from trial to trial may not allow time for 'success' to declare itself; and too prolonged a testing of a wrong trial may allow serious damage to occur. The optimal duration of a trial is clearly the time taken by information to travel from the step-mechanisms that initiate the trial, through the environment, to the essential variable that show the outcome. If the ultrastable system requires the duration to be adjusted, so does the living organism; for there can be little doubt that on many occasions living organisms have missed success either by abandoning a trial too quickly, or by persisting too long with a trial that was actually useless.

If we grade an ultrastable system's environment according to the difficulty they present, we shall find that at the 'easy' end are those that consist of a few variables, independent of each other, and that at the 'difficult' end are those that contain many variables richly cross-linked to form a complex whole.

Adaptation-time of the fully-joined system

Suppose N events each have a probability p of success, and the probabilities are independent. An example would occur if N wheels bore letters A and B on the rim, with A's occupying the fraction p of the circumference and B's the remainder. All are spun and allowed to come to rest; those that stop at an A count as successes Let us compare three ways of compounding these minor successes to a Grand Success, which, we assume, occurs only when every wheel is stopped at an A.

Case 1: All N wheels are spun; if all show an A, Success is recorded and the trials ended; otherwise all are spun again, and so on till 'all A's' comes up at one spin.

Case 2: The first wheel is spun; if it stops at an A, it is left there; otherwise it is spun again. When it eventually stops at an A the second wheel is spun similarly; and so on down the line of N wheels, one at the time, till all show A's.

Case 3: All N wheels are spun; those that show an A are left to continue showing it, and those that show B are spun again. When further A's occur they also are left alone. So the number spun gets fewer and fewer, until all are at A's.

We now ask, how many trials, on the average, will the three cases require?

Suppose, for instance, that *p* is 1/2, that spins occurs at one a second, and that N is 1,000. Then if T_1, T_2, and T_3 are the average time to reach Success in Cases 1, 2, and 3 respectively, … T_1 is about 10^{293} years, T_2 is about 8 minutes, and T_3 is a few seconds.

Thus, while getting Success under the rule of Case 1 (all simultaneously) is practically impossible, getting it under Case 2 and 3 is easy.

We may draw, then, the following conclusion. *A compound event that is impossible if the components have to occur simultaneously may be readily achievable if they can occur in sequence or independently* .

The organism does not reach its full adult adaptation by making trial after trial, all of which count for nothing until suddenly everything come right!

On the contrary, it conforms more to rules of Cases 2 or 3 achieving partial successes and then *retaining them*
while improving what is still unsatisfactory.

Cumulative adaptation

For adaptation to accumulate, there must *not* be channels from some step-mechanisms to some variables not from some variables to others. Thus, for the accumulation of adaptations to be possible the system *must not be fully joined*. The idea so often implicit in physiological writings, that all will be well if only sufficient cross-connexions are available, is, in this context, quite wrong.

This is the point. If the method of ultrastability is to succeed within a reasonable short time, then partial successes must be retained. For this to be possible it is necessary that certain parts should *not* communicate to, or have an effect on, certain other parts.

The views held about the amount of internal connexion in the nervous system - its degree of 'wholeness' - have tended to range from an extreme to the other. The 'reflexologists' from Bell onwards recognised that in some of its activities the nervous system could be treated as a collection of independent parts.

On the other hand, the Gestalt school recognised that many activities of the nervous system were characterised by wholeness, so that what happened at one point was conditional on what was happening at other points. The two sets of facts were sometimes treated as irreconcilable.

Yet Sherrington in 1906 had shown by the spinal reflexes that the nervous system was neither divided into permanently separated parts nor so wholly joined that every event always influenced every other. Rather, it showed a richer, and a more intricate picture - one in which interactions and independencies fluctuated.

The separation into many parts and the union into a single whole are simply the two extremes on the scale of 'degree of connectedness'.

Adaptation thus demands not only the integration of related activities but the independence of unrelated activities.

A polystable system

A set of systems of special importance is the set of those systems that are made of parts that have a high proportion of their states equilibrial, and are made by parts being joined at random.

If to this specification we add the restriction that the original parts are rich in states of equilibrium, we get a type of system.

For lack of a better name I shall call it a **polystable** system.

Briefly, it is any system whose parts have many equilibria and that has been formed by taking parts at random and joining them at random

The polystable system, if composed of parts whose states of equilibrium are distributed independently of the states of their inputs, goes to a final equilibrium in a way that depends much on the amount of functional connexion.

When the connexion is rich, the line of behaviour tends to be complex and, if n is large, exceedingly long; so the whole tends to take an exceedingly long time to come to equilibrium. When the line meets at a state at which an unusually large number of the variables are stable, it cannot retain the excess over the average.

When the connexion is poor (either by few primary joins or by many constancies in the parts), the line of behaviour tends to be short, so that the whole arrives at a state of equilibrium soon. When the line meets at a state at which an unusually large number of the variables are stable, it tends to retain the excess for a time, and thus to progress to total equilibrium by an accumulation of local equilibria.

Adaptation, the organism and the environment

The environment of the living organism tends typically to consist of parts that are rich in states of equilibrium. Associated with this constancy is the fact that most variables of the environment have an immediate effect on only a few of the totality of variables.

I suggest, therefore, that most of the environments encountered on this earth by living organisms contain many parts-functions. Conversely, a system of part-functions adequately represents a very wide class of commonly occurring environments.

As an animal interacts with its environment, the observer will see that the activity in the environment is limited now to this set, now to that. If one set persists active for a long time and the rest remains inactive and inconspicuous, the observer may, if he pleases, call the first set 'the' environment. And if later the activity changes to another set he may, if he pleases, call it a 'second' environment. It is the presence of part-functions and dispersion that makes this change of view reasonable.

An organism that tries to adapt to an environment composed largely of part-functions will find that the environment is composed of subsystems which sometimes are independent but which from time to time show linkage.

Brain, environment, communication and adaptation

Let us dispose once for all of the ideas, fostered in almost every book on the brain written in the last century, that the more communication there is within the brain the better.

In adapting systems, there are occasions when an increase in the amount of communication can be harmful.

Given that co-ordination between the reacting parts is demanded, does this imply that the reacting parts must be in direct communication? It does not; for communication between them is already available *through the environment.*

The anatomist may be excused for thinking that communication between part and part in the brain can take place only through some anatomically or histologically demonstrable tract or fibres. The student of function will, however, be aware that channels are also possible through the environment. An elementary example occurs when the brain monitors the acts of the vocal cords by a feedback that passes, partly at least, through the air before reaching the brain.

Brain connexions: advantages and disadvantages

Other things being equal, the fewer the joins, the fewer are the modes of behaviour available to the system. From this point of view, extra connexions within the brain can be advantageous, for they make possible a greater repertoire of behaviours.

Another way by which the same fact can be seen is to consider the reacting parts before they were joined. The parameters used in the joining must, before the joining, have had fixed values (for otherwise the parts would not have been state-determined). Thus before the joining each

parameter must have been fixed at some *one* of its possible values; after the joining the parameter would be capable of variation as it was affected by the other part. With the variation would have come, to the part, a corresponding variety in its fields, and ways of behaving. Thus joining by mobilising parameters that would otherwise be fixed, adds to the variety of possible behaviours.

If increased connexions bring advantages, they also bring the disadvantage of lengthening, perhaps to a very great degree, the time required for adaptation.

Doubtless there are even more factors to be reckoned in the balance, but what we have seen is sufficient to show that *richness of connexion between the parts in the brain has both advantages and disadvantages*   .

Clearly the organism must develop so that its brain finds, in this respect, an optimum.

It is not suggested that what is wanted is the optimum in the strict sense. Finding an optimum is a much more complex operation than finding a value that is acceptable (according to a given criterion).

To demand the optimum, then, may be excessive; all that is required in biological systems is that the organism finds a state or value *between given limits*.

Summary

The primary fact is that *all* isolated state-determined dynamic systems are selective: from whatever state they have initially, they go towards states of equilibrium. The states of equilibrium are always characterised, in their relation to the change-inducing laws of the system, by being exceptionally resistant.

(Specially resistant are those forms whose occurrence leads, by whatever method, to the occurrence of further replicates of the same form - the so-called 'reproducing' forms.)

If the system permits the formation of local equilibria, these will take the form of dynamic subsystems, exceptionally resistant to the disruptive effects of events occurring locally.

When such a stable dynamic subsystem is examined internally, it will be found to have parts that are *co-ordinated* in their defence against disturbance.

If the class of disturbance changes from generation to generation but is constant within each generation, even more resistant are those forms that are born with a mechanism such that the *environment*
will make it act in a regulatory way against the particular environment - the 'learning' organisms.

This book has been largely concerned with the last stage of the process. It has shown, by consideration of specially clear and simple cases, how the gene-pattern can provide a mechanism (with both basic and ancillary parts) that, when acted on by any given environment, will inevitably tend to adapt to that particular environment.