Behind the Mirror
a search for a natural history of human knowledge

Konrad Lorenz
Translated by Ronald Taylor

A Harvest/HBJ Book
A Helen and Kurt Wolff Book
Harcourt Brace Jovanovich
New York and London
Chapter one

Life as a process of learning

1.1 Positive feedback of energy gain

The most amazing function of the life process, and also the one that is in most need of explanation, is that, in apparent contradiction to the laws of probability, it seems to develop from the simple to the complex, from the more probable to the less probable, from systems of a lower order to systems of higher order. However, none of the laws of physics, even the second law of thermodynamics, is broken by this. All life processes are sustained by the flow of energy being 'dissipated' — as physicists say — in the universe. Or, as a Viennese friend of mine once put it, 'Life feeds on negative entropy.'

All living systems are constructed in such a way as to be able to acquire and store energy. Otto Rössler drew an attractive analogy between life sustained by the dissipation of energy in the world and a sandbank formed across a river sideways on to the current: the more sand has piled up, the more the sandbank is able to collect. The more energy living systems have absorbed, the more they will be able to absorb: when a living organism flourishes, it grows and propagates. A large number of big animals will eat more than a small number of little ones. Organisms are thus systems which derive their energy in this cycle of so-called 'positive feedback'.

One finds similar systems in the inorganic world. The term 'snowballing' denotes the same process: an avalanche gets larger and larger; the bigger a fire, the more rapidly it spreads, and so on. The flame occurs often in poetry and in proverbs as a symbol of life and growth.

1.2 Adaptation as acquisition of knowledge

Organic systems differ from these inorganic systems in one vital respect: they owe their ability to acquire and store energy to certain often highly complex physical structures, which, by a process which makes them capable of performing this function, have been phylogenetically developed for the purpose.

Thanks to the discoveries of Darwin and recent research in biochemistry, we are now able to form clear pictures of the processes by which the adaptiveness of organic structures comes about. The structural plan of every species of living organism is laid down in the two strands of chain-like nucleic acid molecules, the sequence of the nucleotides forming a code. This code is replicated at each mitotic division of the cell, the double chain of the molecule splitting in half and each half becoming a double chain again by 'collecting' free nucleotides and joining them to itself in the same order as that of the half that was separated off. Thus new double chains are formed, consisting of the old half and a new, complementary half. Genetic continuity therefore depends on material continuity, with the reservation, however, as Weidel put it, 'that what is passed from generation to generation is a certain structure of matter'. Sometimes slight errors occur in the replication of the chains, with the result that the code of the newly formed molecular strand deviates in certain details from the previous one. This is what is known as genetic mutation.

In all organisms with proper nuclei, the so-called cariotes, which include all higher animals and plants, the genes are grouped together in larger units called chromosomes, which occur in pairs in every nucleus. In each pair of chromosomes there are identical or corresponding genes, grouped in approximately the same order. Before sexual reproduction each pair of chromosomes divides, leaving the reproductive cells with only half a set of chromosomes, which is called the haploid state of the cell; on fertilization the chromosomes form new pairs of which one element derives from the paternal and one from the maternal side; in this way, as well as through mutation in the chromosomes, new combinations of genetic factors can be produced. As a result of these processes, described here in a very simplified way, the external appearance of higher organisms, the so-called phenotype, is never entirely constant.

The extent and frequency of these variations are such that they do not threaten the survival of the species by the production of unviable
aberrations, but they certainly do not always work to the advantage of the individual organisms. Indeed, since all these slight, sometimes infinitesimal, variations are entirely random, their most frequent consequence is to reduce the prospects of energy gain and survival. Only in exceptional cases — but it is these that matter here — does a mutation or new combination of genetic factors enable an organism to adapt more profitably to its environment than its ancestors had done. In such cases the new organism is better fitted to one or the other element of its environment, thus gaining a better chance to acquire energy, or reducing the probability of loss of energy. The chances of survival and reproduction of the favoured organism increase at the expense of, and in competition with, its not similarly favoured fellows, which are thus doomed to extinction. This process is known as natural selection, and the modifications brought about in the organism as adaptation.

Their insight into these two processes has forced biologists to form two concepts unknown to physicists or chemists. The first is that of ‘fitness for survival’ or teleonomy. As natural selection ‘breeds’ structures which fulfil particularly well a certain survival function, they finally look as though they had been created for the very purpose by a wise and beneficent spirit. One may remark, in parenthesis, that this is not such an entirely wrong impression, for, as I shall show later in this chapter, the constructive power of the human mind derives from processes which are basically similar to those that occur in the genome.

In the last analysis all complex structures of the entire range of organisms have arisen through the selection pressure of survival functions. When a biologist comes across a structure whose function he does not know he naturally asks what purpose it serves. To ask, for example, ‘Why does a cat have curved, pointed claws?’ and to receive the answer, ‘In order to catch mice’, epitomizes the way in which a problem is posed and solved. Colin Pittendrigh coined the word ‘teleonomic’ for this question of survival value, seeking to draw as sharp a distinction between teleonomy and teleology as there is between astronomy and astrology.

The second concept that emerges from this process of adaptation is that of knowledge or ‘information’. The word ‘adapt’ itself implies a process through which a correspondence is brought about between the object that is being adapted and the circumstances to which it is being adapted. What an organism learns in this way about external reality is, quite literally, ‘impressed’ or ‘imprinted’ on it.

The word information primarily means ‘giving form’.

To use the word ‘information’ in the everyday sense could lead to misunderstanding, because information theorists give it a far wider meaning, deliberately avoiding the semantic level, e.g. any question of the content of the information, and particularly any relevance it may have for the biological survival of the organism. Thus in the terminology of information theory we cannot talk, as we do in everyday language, of ‘information about something’. Common parlance, however, has two plain words for this kind of acquiring and possessing information: apprehending and knowing. However, when I speak of information as being the root of all processes of adaptation, I am using the word in its everyday sense to denote something that has a meaning and a purpose for whoever receives or possesses it.

For the benefit of the reader interested in information theory I might add that, as Bernhard Hassenstein has shown, it would be possible also to define this conception of the word ‘information’ in terms of information theory itself. One would then have to say that adaptation is an increment in the transfonnation between the organism and its environment, the increment being caused by processes within the organism, without any perceptible change in the environment. This could be regarded as a special case of the emergence of what Meyer-EPpller called a ‘correspondence’— an asymmetrical or one-sided correspondence, however, inasmuch as it is achieved entirely by changes in only one of the two parties. As the reader can see, the terminology of information theory is hardly suited to describing life processes.

The gain in knowledge, achieved by trial and error by the genome — which retains what is fittest for it — results, as mentioned above (p.6) in the formation of an image of the material world within the living system; an image ‘built up’, as Donald M. Mackay puts it, from the ‘descriptive information content of the situation’. The image thus formed is, in a way, a negative of reality, like a photographic negative, or the plaster cast of a coin. Uexküll described the organism as having a ‘contrapuntal’ relationship to its environment. As already mentioned, there are relationships of this kind between organism and reality in the field of anatomy, such as the sun-like nature of the eye, or the wave-like movements of a fish’s fins. Structures of this kind owe their remarkable efficiency to their ‘descriptive information content’ and serve the

*Quoted from Norbert Bischof.
organism's economy to a marvellous degree, enabling it to tap the most unlikely and inaccessible sources of energy.

The method of the genome, perpetually making experiments, matching their results against reality, and retaining what is fittest, differs from that adopted by man in his scientific quest for knowledge in only one respect, and that not a vital one, namely that the genome learns only from its successes, whereas man learns also from his failures. Otherwise, like the genome, man proceeds by comparing an idea in his mind, a hypothesis he has evolved, with the outside world and 'checks whether it fits'. In his treatise Pattern Matching as an Essential in Distal Knowing Donald T. Campbell has convincingly shown that the greater part of all knowledge, from the simple recognition of an object to the verification of a scientific hypothesis, is arrived at in this way.

This process of 'pattern matching' can only function with groups or configurations of sense data and the relationship between them. The isolated report of the sensory cell is, in principle, always ambiguous. One cannot identify an individual star that shines through a small gap in the clouds; only when there is a larger area of clear sky visible with a number of stars in it is one in a position to compare the pattern one sees with the stellar pattern one knows, and thus identify the star one saw first, provided that it is a fixed star. If it is a planet a great deal more knowledge of the heavenly bodies and the temporal pattern of their distribution is needed to identify it.

In his 'Evolutionary Epistemology' Campbell writes: 'The natural selection paradigm of such knowledge increments can be generalized to other epistemic activities, such as learning, thought and science.' I not only agree with this statement, but I regard it as one of the main tasks of my book to undertake just such a comparison of the various mechanisms by means of which different living systems acquire and store the information relevant to their needs. Most of what science has brought to light about the external world, as Campbell rightly says, has been discovered by 'pattern matching', and since cognitive processes, some on the highest and some on the lowest level, are based on the same principle, one might conclude that there was no other way to acquire knowledge.

This erroneous view — as we shall see it to be — might appear to be strengthened by the fact that all cognitive mechanisms, the oldest and simplest as well as the most recent and most complex, have a further important characteristic in common — namely, that both the mechanism by which the genome acquires knowledge, and that by which man does the same, change with each item of knowledge acquired. Having acquired new information, neither is the same as it was before. Each new acquisition improves its chances of gaining energy and thereby the probability of acquiring further knowledge.

1.3 The acquisition of instantaneous information not intended for storage

But there are also cognitive functions of a quite different kind. As adaptation has created physical structures equipped to absorb and utilize energy, so it has also produced structures whose function is to acquire and exploit knowledge about circumstances which arise suddenly, without notice, in the animal's environment, and which must be instantly taken into account by its behaviour. Behaviour that depends on the operation of these mechanisms takes the form of a relevant response to a particular environmental situation, although this situation has never before arisen in this particular form either in the phylogeny of the species or in the life of the individual organism itself. Significantly, this definition is also applicable to so-called 'insight-controlled behaviour', and is as true of the simplest taxes or orientation responses as it is of the sophisticated operations of the sense organs and nervous system which lie at the root of the a priori forms of human intuition and thought.

The subjective phenomenon of insight, called by Karl Bühler the 'Aha experience', manifests itself in the same way whether we have just grasped the significance of a set of highly complex relationships, or whether through the function of orientation responses the state of disorientation gives way to one of blessed orientatedness, such as when the statolithic apparatus of our inner ear tells us that 'upwards' is a different direction from what it was a moment earlier. I once experienced for myself how powerful this insight can be, when, as I was lying fast asleep in a motor boat on the Danube one night, a friend tipped me overboard. So muddy was the water that even at a shallow depth there was no light to show me which way was down and which was up. I can assure my readers that it was a traumatic experience of the meaning of 'insight' when after a few anxious moments of total disorientation, the statoliths did what was expected of them.
The processes described here are not adaptation processes like those discussed earlier (p.21), but the operations of anatomical, sensory and nervous structures whose adaptation is already completed. Those structures are as little, or even less, open to individual modification than those serving the acquisition not of information but of energy. Also, the repeated function of acquiring instantaneous information must again leave no traces in the physiological mechanism performing it. The basic function of keeping the organism informed of rapid change in its environment presupposes that the underlying apparatus remains capable of cancelling one message and substituting another, which may often be just the opposite.

There is a further and even more important consideration. The organizations which are immune to all changes, i.e. the mechanisms which, on the basis of momentary sensory messages, provide us with our immediate ‘insights’ into the world around us, are the foundation of all experience. They precede all experience, and must do so, if experience is to be possible at all. In this respect they correspond absolutely to Kant’s definition of the a priori.

As we shall discuss in a variety of contexts, the reliable performance of a completely adapted structure must invariably be paid for by the loss of certain degrees of freedom. The mechanisms of these instantaneous cognitive activities are no exception. By being specifically adapted to acquire a particular kind of information, most structures are tied to a very narrow, rigid programme, their inbuilt computing mechanisms containing ‘hypotheses’ to which they blindly adhere. If circumstances arise that were not ‘foreseen’ by the adaptive process that produced them, a structure may transmit false information which it cannot be taught to correct. The various different kinds of sensory illusion provide ample proof of this.

This ‘doctrinaire’ element resulting from the completion of adaptive processes forces our cognitive faculties to accept various hypotheses — or rather, it foists hypotheses upon them without our noticing it. We cannot experience, perceive or think anything except on the basis of assumptions and suppositions of which such hypotheses are innate parts. They are built into our perceiving apparatus. And however hard we try, in a spirit of free inquiry, to discover new hypotheses, we cannot banish from them these ancient a priori hypotheses, which have evolved through mutations and new combinations of genes and been tested over countless ages of ‘pattern matching’, and which, though never absolutely pointless, are always rigid and never entirely relevant.

1.4 Double feedback of energy and information

The acquisition and storage of relevant information is as basic a function of all living organisms as is the absorption and storing of energy. Both are as old as life itself. Otto Rössler, I believe, was the first biologist to point out that not only do the energy-absorbing processes form a positive feedback cycle in themselves but they also have a positive feedback relationship to the information-acquiring processes.

When, through mutation of a new combination of genetic factors, the probability of absorbing energy increases to the point where natural selection favours the development of the new, improved organism, the number of its descendants also increases. At the same time it also becomes more probable that it will be these descendants who will win the next big prize in the mutation stakes.

This dual feedback cycle characterizes all life, including that of viruses, which, in Weidel’s striking phrase, have only ‘a borrowed life’. It is undeniably true, yet at the same time misleading, to say that living organisms are at the mercy of purely random changes and that evolution only takes place through the elimination of the unfit.

A closer approximation to what really happens in organic nature would be to describe it as follows: life is an eminently active enterprise aimed at acquiring both a fund of energy and a stock of knowledge, the possession of one being instrumental to the acquisition of the other. The immense effectiveness of these two feedback cycles, coupled in multiplying interaction, is the precondition, indeed the explanation, for the fact that life has the power to assert itself against the superior strength of the pitiless inorganic world, and also for the fact that it tends at times to an excessive expansion. The process whereby a large modern industrial company, such as a chemical firm, invests a considerable part of its profits in its laboratories in order to promote new discoveries and thus new sources of profit is not so much a model as a specific case of the process that is going on in all living systems.

It was an important discovery on the part of Otto Rössler that the phylogeny of the organic world does not depend on ‘chance’ or ‘luck’ but that organisms instantly seize on any favourable
Chapter two

The creation of new system characteristics

2.1 Inadequacy of vocabulary

When one attempts to describe the great process of organic growth, one finds oneself hampered by the fact that the language of culture was born at a time when ontogeny, i.e. the evolution of the individual creature, was the only form of development known. Words like development and evolution have the etymological connotation of the unfolding of something that was already there in a compressed or confined form, like the flower in the bud, or the chicken in the egg. For ontogenic processes of this kind such words are perfectly suitable. But they are lamentably inadequate when one attempts to define the nature of an organic creative process through which something entirely new comes into existence, something that was simply not there before.

2.2 Fulguratio, or the 'creative flash'

Theistic philosophers and mystics of the Middle Ages coined the term fulguratio, 'flash of lightning', to denote the act of creation, thereby conveying the notion of a sudden intervention from above, from God. By an etymological accident or perhaps through deeper, unsuspected associations, this term is far more appropriate than those mentioned above for designating the coming into existence of something previously not there. A thunderbolt from Zeus is for the scientist an electric spark like any other, and if we see a spark at
an unexpected point in a system, the first thing we think of is a short circuit, a new connection.

If, for example, two independent systems are coupled together (see Fig. 1, taken from Bernhard Hassenstein), entirely new, unexpected system characteristics will emerge, of whose appearance there was previously not the slightest suggestion. This is the profound truth behind the Gestalt psychologists' principle, mystical in tone but absolutely correct, that the whole is more than the sum of its parts.

There are many ways, as we shall see, in which such new characteristics can arise. One particular way is the following. In a series of subsystems causally linked in a linear chain, with the first functioning only as a cause and the last only as an effect, this last can, by the emergence of a new causal link, influence the first and thus turn the causal chain into a cycle, or loop. We have already met examples of such cycles with positive feedback in our discussion of how energy and information are acquired. Equally important are negative feedback loops, which I shall discuss in later chapters, where they logically belong. Suffice it for the moment to say that, if at some point in a causal cycle a 'negative sign' is built in, i.e. if the effect of one process in the chain becomes weaker as a result of the preceding process becoming stronger, the result is one of regulation. For example, the higher the water rises in a lavatory cistern, the higher it lifts the ball-cock, which then prevents more water from running in. The result is a constancy in the water level.

Cybernetics and systems theory have shown that the sudden emergence of new system characteristics has nothing to do with miracles, thereby absolving phylogeneticists from the reproach of vitalism. There is nothing supernatural about a linear causal chain joining up to form a cycle, thus producing a system whose functional properties differ fundamentally from those of all preceding systems. If an event of this kind is phylogenetically unique it may be epoch-making in the literal sense of the word. Some geological epochs have derived their names from such events.

2.3 Unity out of diversity

Many thinkers, philosophers and scientists alike, have recognized that progress in organic development is almost always achieved through the integration of a number of different and independent

---

**Figure 1** Three electric circuits, including an oscillating circuit (c). A current is passed between the poles of a battery with electromotive strength \( E_0 \) or terminal voltage \( V_0 \). \( R \) = resistance. Circuit (a) has a condenser with capacity \( C \); (b) has a coil with inductance \( L \); circuit (c) has both condenser and coil. The voltage \( V \) can be measured at the two terminals.

The diagrams on the right show the changes in voltage when the switch is closed at time zero. In circuit (a) the condenser gradually charges through the resistance until voltage \( V_0 \) is reached. In circuit (b) the current, initially impeded by self-induction, increases until it reaches the strength laid down by Ohm's law; the voltage \( V \) is then theoretically zero, because the total resistance is concentrated in \( R \). In circuit (c) diminishing oscillations are produced. It is clear at a glance that the behaviour of (c) is not the result of superimposing (a) and (b) although (c) can be thought of as having arisen through the addition of (a) and (b). This pattern is valid, for example, for the following values: \( C = 0.7 \times 10^{-9} \text{F} \); \( L = 2 \times 10^{-3} \text{H} \); \( R = 10^3 \omega \); \( \lambda \approx 1.2 \times 10^{-6} \text{sec} \). This last value also defines the time axis, which is the same for all three curves. (Calculations by E. U. von Weizsäcker.)
systems to form a unit of a higher order, the systems becoming modified in the process, the better to equip them for their roles in the new, more advanced system. Goethe defined development as differentiation and subordination of the parts. Ludwig von Bertalanffy has described this process with great precision in his *Theoretische Biologie*, while in *Science, Man and Morals* W. H. Thorpe demonstrated that the most important creative principle in evolution is the emergence of a totality from a mass of different parts which become more and more dissimilar in the process — as well as more dependent on each other. Teilhard de Chardin gave the most succinct and poetic definition: 'Créer, c'est unir' ('To create is to unify'). This is a principle that must have been at work from the very beginning of life.

This creative union of diverse parts to form a new entity in itself implies a complication of the living system. However, the new system often becomes simplified in the course of further evolution through the 'specialization' of the individual parts — that is, through their restriction to particular functions; the other functions which the individual part had to perform while it was still independent are now assigned to other parts of the new system. Even the ganglion cells of our brain, which together perform the highest intellectual functions, are individually far inferior to an amoeba or a paramecium, both in its individual function as a cell and as concerns the relevant information that underlies this function. An amoeba or a paramecium has at its command a whole series of appropriate responses to external stimuli and 'knows' a great number of important things about its environment, whereas the ganglion cell only 'knows' when to 'fire'. Even this it cannot do with varying strength: it either does it or not, following the law of 'all or nothing'. This 'stultification' of an element that has become part of a higher entity has, of course, its good side as well, in that it is indispensable for the operation of the entity as such by guaranteeing the unequivocal transmission of messages. The strength of the message passed on by the cell must not be allowed to depend on the condition in which the cell happens to be, any more than a well-disciplined soldier can be left to decide whether to carry out an order enthusiastically or not.

This process of simplification in the course of integration into a higher entity is a phenomenon found at all stages of evolution. In the psychosocial development of man and his culture it presents special problems. The inevitable development of the practice of the division of labour leads in all areas, but most tragically in the intellectual sphere, to ever increasing specialization. The specialist comes to know more and more about less and less, until finally he knows everything about a mere nothing. There is a serious danger that the specialist, forced to compete with his colleagues in acquiring more and more pieces of more and more specialized knowledge, will become more and more ignorant about other branches of knowledge, until finally he is utterly incapable of forming any judgement on the role and importance of his own sphere within the context of human knowledge and culture as a whole.

Another form of simplification is what is called in social contexts 'improved organization'. The first, experimental forms of a new machine are always more complicated than the final version, and the same is frequently true of living systems. Interaction between the various elements, in particular the exchange of information, becomes simplified or is made more direct, and the superfluous remains of earlier stages are removed, or 'become rudimentary', as biologists say.

### 2.4 The one-sided relationship between levels of integration

The manner described above in which smaller systems become integrated to form a new and more advanced system leads to a peculiar kind of one-way relationship, both between the whole and its constituent parts within the organism, and between higher animals and their primitive ancestors. In principle the same relationship also exists between the whole of organic life and the inanimate matter from which it is derived. In ontological terms one can define the relationship by saying that the whole is its parts, and continues to be so even if, as the result of a series of *fulgurationes*, or 'creative flashes', it acquires a number of additional system characteristics in the course of its evolution. The subsidiary systems themselves do not gain any higher characteristics and may even lose some in the process of simplification. None of the laws that govern the subsidiary systems ever suffers any infraction in the new entity, least of all those of inanimate matter, out of whose elements all living systems are built.

The one-way relationship consists on the one hand in the system as a whole possessing all the characteristics, in particular all the weaknesses, of its component parts — for a chain is only as strong as its weakest link. On the other hand, none of these parts possesses the
characteristics specific of the whole. Similarly, every higher organism possesses most of the characteristics of its primitive ancestors, but even the fullest knowledge of an animal's characteristics will not allow us to predict those of its more highly developed descendants.

This does not mean, of course, that highly developed systems are beyond analysis or natural explanation. But the scientist must never forget that the laws and characteristics of any system, like those of the individual subsidiary systems within it, have to be explained on the basis of the laws and characteristics of the systems on the next lower level of integration. And this, in its turn, is only possible when one is familiar with the structure of the higher entity formed by the coalescence of the systems on this level. If one assumes complete knowledge of this structure, it is theoretically possible to explain every living system and all its functions, even the most advanced, in natural terms, i.e. without adducing supernatural factors.

2.5 The irrational residue

However, to claim that a creature is in principle capable of being explained is a valid assertion only if we accept the present structures in its body as our data — in other words, if we maintain that we are not interested in their historical evolution. For, if we ask why a particular organism is structured in one way rather than another, the most important answers will be found in the history of the species concerned. To the question why our ears are situated on each side of our heads, the causal answer is that we are descended from water-breathing ancestors who had at these points a gill slit, the spiraculum, which was retained as an air-conducting passage during the transition to terrestrial life and then, through functional change, became part of the auditory system.

The number of purely historical causes one would need to know in order to fully explain why an organism is as it is may not be infinite but it is sufficiently great to make it impossible ever to trace all causal chains to their end, even supposing they had one. Thus we are always left with what Max Hartmann called an irrational or non-rationalizable residue. That evolution produced oak trees and human beings in the Old World but eucalyptus trees and kangaroos in Australia is simply a result of such undetectable causal sequences, which, in a mood of resignation, we have come to describe as 'chance'.

Although, as we must always emphasize, we scientists cannot believe in miracles — that is, in violations of the universal laws of nature — we are at the same time aware that we can never succeed in giving a complete explanation of how a creature has evolved from its lower ancestors. A higher animal, as Michael Polanyi in particular has pointed out, cannot be 'reduced' to its simpler ancestors; still less can a living system be 'reduced' to inorganic matter and the processes that take place within it.

The same is even true of man-made machines. From the point of view of their material construction they are completely analysable, so that they can be readily manufactured. If, on the other hand, one considers their historical, teleonomic evolution as organs of man, any attempt to explain why they are as they are, and not otherwise, brings one up against the very same non-rationalizable remainder as with living systems.

Polanyi will scarcely have intended to postulate the presence of vitalist factors, but in order to remove any possibility of misunderstanding, I would prefer to say that no system on a higher level of integration can be deduced from a lower system, however fully one may understand this lower system. We know with certainty that higher systems have arisen from lower ones, absorbing them and containing them like bricks in a building. We also know, with absolute certainty, the earlier stages in development from which higher living beings emerged. But each step forward has consisted of a fulguratio, a historically unique event in phylogeny which has always had a chance quality about it — the quality, one might say, of something invented.