CHAPTER 1

The Problem

1/1. How does the brain produce adaptive behaviour? In attempting to answer the question, scientists have discovered two sets of facts and have had some difficulty in reconciling them. The physiologists have shown in a variety of ways how closely the brain resembles a machine: in its dependence on chemical reactions, in its dependence on the integrity of anatomical paths, and in many other ways. At the same time the psychologists have established beyond doubt that the living organism, whether human or lower, can produce behaviour of the type called 'purposeful' or 'intelligent' or 'adaptive'; for though these words are difficult to define with precision, no one doubts that they refer to a real characteristic of behaviour. These two characteristics of the brain's behaviour have proved difficult to reconcile, and some workers have gone so far as to declare them incompatible.

Such a point of view will not be taken here. I hope to show that a system can be both mechanistic in nature and yet produce behaviour that is adaptive. I hope to show that the essential difference between the brain and any machine yet made is that the brain makes extensive use of a principle hitherto little used in machines. I hope to show that by the use of this principle a machine's behaviour may be made as adaptive as we please, and that the principle may be capable of explaining even the adaptiveness of Man.

But first we must examine more closely the nature of the problem, and this will be commenced in this chapter. The succeeding chapters will develop more accurate concepts, and when we can state the problem with precision we shall be not far from its solution.

Behaviour, reflex and learned

1/2. The activities of the nervous system may be divided more or less distinctly into two types. The dichotomy is perhaps an over-simplification, but it will be sufficient for our purpose.

The first type is reflex behaviour. It is inborn, it is genetically determined in detail, it is a product, in the vertebrates, chiefly of centres in the spinal cord and in the base of the brain, and it is not appreciably modified by individual experience. The second type is learned behaviour. It is not inborn, it is not genetically determined in detail (more fully discussed in S. 1/9), it is a product chiefly of the cerebral cortex, and it is modified markedly by the organism's individual experiences.

1/3. With the first or reflex type of behaviour we shall not be concerned. We assume that each reflex is produced by some neural mechanism whose physico-chemical nature results inevitably in the characteristic form of behaviour, that this mechanism is developed under the control of the gene-pattern and is inborn, and that the pattern of behaviour produced by the mechanism is usually adapted to the animal's environment because natural selection has long since eliminated all non-adapted variations. For example, the complex activity of ' coughing ' is assumed to be due to a special mechanism in the nervous system, inborn and developed by the action of the gene-pattern, and adapted and perfected by the fact that an animal who is less able to clear its trachea of obstruction has a smaller chance of survival.

Although the mechanisms underlying these reflex activities are often difficult to study physiologically and although few are known in all their details, yet it is widely held among physiologists that no difficulty of principle is involved. Such behaviour and such mechanisms will not therefore be considered further.

1/4. It is with the second type of behaviour that we are concerned : the behaviour that is not inborn but learned. Examples of such reactions exist in abundance, and any small selection must seem paltry. Yet I must say what I mean, if only to give the critic a definite target for attack. Several examples will therefore be given.

A dog selected at random for an experiment with a conditioned reflex can be made at will to react to the sound of a bell either with or without salivation. Further, once trained to react in one way it may, with little difficulty, be trained to react later in the opposite way. The salivary response to the sound of a bell cannot, therefore, be due to a mechanism of fixed properties.

A rat selected at random for an experiment in maze-running can be taught to run either to right or left by the use of an appropriately shaped maze. Further, once trained to turn to one side it can be trained later to turn to the other.

A kitten approaching a fire for the first time is unpredictable in its first reactions. The kitten may walk almost into it, or may spit at it, or may dab at it with a paw, or may try to sniff at it, or may crouch and ' stalk ' it. The initial way of behaving is not, therefore, determined by the animal's species.

Perhaps the most striking evidence that animals, after training, can produce behaviour which cannot possibly have been inborn is provided by the circus. A seal balances a ball on its nose for minutes at a time; one bear rides a bicycle, and another walks on roller skates. It would be ridiculous to suppose that these reactions are due to mechanisms both inborn and specially perfected for these tricks.

Man himself provides, of course, the most abundant variety of learned reactions: but only one example will be given here. If one is looking down a compound microscope and finds that the object is not central but to the right, one brings the object to the centre by pushing the slide still farther to the right. The relation between muscular action and consequent visual change is the reverse of the usual. The student's initial bewilderment and clumsiness demonstrate that there is no neural mechanism inborn and ready for the reversed relation. But after a few days co-ordination develops.

These examples, and all the facts of which they are representative, show that the nervous system is able to develop ways of behaving which are not inborn and are not specified in detail by the gene-pattern.

1/5. Learned behaviour has many characteristics, but we shall be concerned chiefly with one: when animals and children learn, not only does their behaviour change, but it changes usually for the better. The full meaning of ' better ' will be discussed in

Chapter 5, but in the simpler cases the improvement is obvious enough. 'The burned child dreads the fire': after the experience the child's behaviour towards the fire is not only changed, but is changed to a behaviour which gives a *lessened* chance of its being burned again. We would at once recognise as abnormal any child who used its newly acquired knowledge so as to get to the flames more quickly.

To demonstrate that learning usually changes behaviour from a less to a more beneficial, i.e. survival-promoting, form would need a discussion far exceeding the space available. But in this introduction no exhaustive survey is needed. I require only sufficient illustration to make the meaning clear. For this purpose the previous examples will be examined seriatim.

When a conditioned reflex is established by the giving of food or acid, the amount of salivation changes from less to more. And the change benefits the animal either by providing normal lubrication for chewing or by providing water to dilute and flush away the irritant. When a rat in a maze has changed its behaviour so that it goes directly to the food at the other end, the new behaviour is better than the old because it leads more quickly to the animal's hunger being satisfied. The kitten's behaviour in the presence of a fire changes from being such as may cause injury by burning to an accurately adjusted placing of the body so that the cat's body is warmed by the fire neither too much nor too little. The circus animals' behaviour changes from some random form to one determined by the trainer, who applied punishments and rewards. The animals' later behaviour is such as has decreased the punishments or increased the rewards. In Man, the proposition that behaviour usually changes for the better with learning would need extensive discussion. But in the example of the finger movements and the compound microscope, the later movements, which bring the desired object directly to the centre of the field, are clearly better than the earlier movements, which were disorderly and ineffective.

Our problem may now be stated in preliminary form: what cerebral changes occur during the learning process, and why does the behaviour usually change for the better? What type of mechanistic process could show the same property?

But before the solution is attempted we must first glance at the peculiar difficulties which will be encountered.

1/6. The nervous system is well provided with means for action. Glucose, oxygen, and other metabolites are brought to it by the blood so that free energy is available abundantly. The nerve cells composing the system are not only themselves exquisitely sensitive, but are provided, at the sense organs, with devices of even higher sensitivity. Each nerve cell, by its ramifications, enables a single impulse to become many impulses, each of which is as active as the single impulse from which it originated. And by their control of the muscles, the nerve cells can rouse to activity engines of high mechanical power. The nervous system, then, possesses almost unlimited potentialities for action. But do these potentialities solve our problem? It seems not. We are concerned primarily with the question why, during learning, behaviour changes for the better: and this question is not answered by the fact that a given behaviour can change to one of lesser or greater activity. The examples given in S. 1/5, when examined for the energy changes before and after learning, show that the question of the quantity of activity is usually irrelevant.

But the evidence against regarding mere activity as sufficient for a solution is even stronger: often an increase in the amount of activity is not so much irrelevant as positively harmful.

If a dynamic system is allowed to proceed to vigorous action without special precautions, the activity will usually lead to the destruction of the system itself. A motor car with its tank full of petrol may be set into motion, but if it is released with no driver its activity, far from being beneficial, will probably cause the motor car to destroy itself more quickly than if it had remained inactive. The theme is discussed more thoroughly in S. 20/12; here it may be noted that activity, if inco-ordinated, tends merely to the system's destruction. How then is the brain to achieve success if its potentialities for action are partly potentialities for self-destruction?

The relation of part to part

1/7. It was decided in S. 1/5 that after the learning process the behaviour is usually better adapted than before. We ask, therefore, what property must be possessed by the neurons, or by the parts of a mechanical 'brain', so that the manifestation by the neuron of this property shall result in the whole animal's behaviour being improved.

Even if we allow the neuron all the properties of a living organism, it is still insufficiently provided. For the improvement in the animal's behaviour is often an improvement in relation to entities which have no counterpart in the life of a neuron. Thus when a dog, given food in an experiment on conditioned reflexes, learns to salivate, the behaviour improves because the saliva provides a lubricant for chewing. But in the neuron's existence, since all its food arrives in solution, neither ' chewing ' nor ' lubricant ' can have any direct relevance or meaning. Again, a rat learns to run through a maze without mistakes ; yet the learning has involved neurons which are firmly supported in a close mesh of glial fibres and never move in their lives.

Finally, consider an engine-driver who has just seen a signal and whose hand is on the throttle. If the light is red, the excitation from the retina must be transmitted through the nervous system so that the cells in the motor cortex send impulses down to those muscles whose activity makes the throttle *close*. If the light is green, the excitation from the retina must be transmitted through the nervous system so that the cells in the motor cortex make the throttle open. And the transmission is to be handled, and the safety of the train guaranteed, by neurons which can form no conception of ' red ', ' green ', ' train ', ' signal ', or ' accident ' ! Yet the system works.

1/8. In some cases there may be a simple mechanism which uses the method that a red light activates a chain of nerve-cells leading to the muscles which close the throttle while a green light activates another chain of nerve-cells leading to the muscles which make it open. In this way the effect of the colour of the signal might be transmitted through the nervous system in the appropriate way.

The simplicity of the arrangement is due to the fact that we are supposing that the two reactions are using two completely separate and independent mechanisms. This separation may well occur in the simpler reactions, but it is insufficient to explain the events of the more complex reactions. In most cases the ' correct ' and the ' incorrect ' neural activities are alike composed of excitations, of inhibitions, and of other changes that are all physiological, so that the correctness is determined not by the process itself but by the relations which it bears to the other processes.

This dependence of the ' correctness ' of what is happening at one point in the nervous system on what is happening at other points would be shown if the engine-driver were to move over to the other side of the cab. For if previously a flexion of the elbow had closed the throttle, the same action will now open it; and what was the correct pairing of red and green to push and pull must now be reversed. So the local action in the nervous system can no longer be regarded as ' correct ' or ' incorrect ', and the first simple solution breaks down.

Another example is given by the activity of chewing in so far as it involves the tongue and teeth in movements which must be related so that the teeth do not bite the tongue. No movement of the tongue can by itself be regarded as wholly wrong, for a movement which may be wrong when the teeth are just meeting may be right when they are parting and food is to be driven on to their line. Consequently the activities in the neurons which control the movement of the tongue cannot be described as either ' correct ' or ' incorrect ': only when these activities are related to those of the neurons which control the jaw movements can a correctness be determined; and this property now belongs, not to either separately, but only to the activity of the two in combination.

These considerations reveal the main peculiarity of the problem. When the nervous system learns, it undergoes changes which result in its behaviour becoming better adapted to the environment. The behaviour depends on the activities of the various parts whose individual actions compound for better or worse into the whole action. Why, in the living brain, do they always compound for the better?

If we wish to build an artificial brain the parts must be specified in their nature and properties. But how can we specify the ' correct ' properties for each part if the correctness depends not on the behaviour of each part but on its relations to the other parts? Our problem is to get the parts properly co-ordinated. The brain does this automatically. What sort of a machine can be *self*-co-ordinating?

This is our problem. It will be stated with more precision in S. 1/12. But before this statement is reached, some minor topics must be discussed.

The genetic control of cerebral function

1/9. The various species of the animal kingdom differ widely in their powers of learning: Man's intelligence, for instance, is clearly a species-characteristic, for the higher apes, however well trained, never show an intelligence equal to that of the average human being. Clearly the power of learning is determined to some extent by the inherited gene-pattern. In what way does the gene-pattern exert its effect on the learning process? In particular, what part does it play in the adjustments of part to part which the previous section showed to be fundamental? Does the gene-pattern determine these adjustments in detail?

In Man, the genes number about 50,000 and the neurons number about 10,000,000,000. The genes are therefore far too few to specify every neuronic interconnection. (The possibility that a gene may control several phenotypic features is to some extent balanced by the fact that a single phenotypic feature may require several genes for its determination.)

But the strongest evidence against the suggestion that the genes exert, in the higher animals, a detailed control over the adjustments of part to part is provided by the evidence of S. 1/4. A dog, for instance, can be made to respond to the sound of a bell either with or without salivation, regardless of its particular gene-pattern. It is impossible, therefore, to relate the control of salivation to the particular genes possessed by the dog. This example, and all the other facts of which it is typical, show that the effect of the gene-pattern on the details of the learning process cannot be direct.

The effect, then, must be indirect: the genes fix permanently certain function-rules, but do not interfere with the function-rules in their detailed application to particular situations. Three examples of this type of control will be given in order to illustrate its nature.

In the game of chess, the laws (the function rules) are few and have been fixed for a century; but their effects are as numerous as the number of positions to which they can be applied. The result is that games of chess can differ from one another though controlled by constant laws.

A second example is given by the process of evolution through natural selection. Here again the function-rule (the principle of the survival of the fittest) is fixed, yet its influence has an infinite variety when applied to an infinite variety of particular organisms in particular environments.

A final example is given in the body by the process of inflammation. The function-rules which govern the process are genetically determined and are constant in one species. Yet these rules, when applied to an infinite variety of individual injuries, provide an infinite variety in the details of the process at particular points and times.

Our aim is now clear: we must find the function-rules. They must be few in number, much fewer than 50,000, and we must show that these few function-rules, when applied to an almost infinite number of circumstances and to 10,000,000,000 neurons, are capable of directing adequately the events in all these circumstances. The function-rules must be fixed, their applications flexible.

(The gene-pattern is discussed again in S. 9/9.)

Restrictions on the concepts to be used

1/10. Throughout the book I shall adhere to certain basic assumptions and to certain principles of method.

The nervous system, and living matter in general, will be assumed to be identical with all other matter. So no use of any ' vital ' property or tendency will be made, and no *deus ex machina* will be invoked. No psychological concept will be used unless it can be shown in objective form in non-living systems; and when used it will be considered to refer solely to its objective form. Related is the restriction that every concept used must be capable of objective demonstration. In the study of man this restriction raises formidable difficulties extending from the practical to the metaphysical. But as most of the discussion will be concerned with the observed behaviour of animals and machines, the peculiar difficulties will seldom arise.

No teleological explanation for behaviour will be used. It will be assumed throughout that a machine or an animal behaved in a certain way at a certain moment because its physical and chemical nature at that moment allowed it no other action. Never will we use the explanation that the action is performed because it will later be advantageous to the animal. Any such explanation would, of course, involve a circular argument; for our purpose is to explain the origin of behaviour which appears to be teleologically directed.

It will be further assumed that the nervous system, living matter, and the matter of the environment are all strictly determinate: that if on two occasions they are brought to the same state, the same behaviour will follow. Since at the atomic level of size the assumption is known to be false, the assumption implies that the functional units of the nervous system must be sufficiently large to be immune to this source of variation. For this there is some evidence, since recordings of nervous activity, even of single impulses, show no evidence of appreciable thermal noise. But we need not prejudge the question. The work to be described is an attempt to follow the assumption of determinacy wherever it leads. When it leads to obvious error will be time to question its validity.

Consciousness

1/11. The previous section has demanded that we shall make no use of the subjective elements of experience; and I can anticipate by saying that in fact the book makes no such use. At times its rigid adherence to the objective point of view may jar on the reader and may expose me to the accusation that I am ignoring an essential factor. A few words in explanation may save misunderstanding.

Throughout the book, consciousness and its related subjective elements are not used for the simple reason that at no point have I found their introduction necessary. This is not surprising, for the book deals with only one of the aspects of the mind-body relation, and with an aspect—learning—that has long been recognised to have no necessary dependence on consciousness. Here is an example to illustrate their independence. If a cyclist wishes to turn to the left, his first action must be to turn the front wheel to the *right* (otherwise he will fall outwards by centrifugal force). Every practised cyclist makes this movement every time he turns, yet many cyclists, even after they have made the movement hundreds of times, are quite unconscious of making it. The direct intervention of consciousness is evidently not necessary for adaptive learning.

Such an observation, showing that consciousness is sometimes not necessary, gives us no right to deduce that consciousness does not exist. The truth is quite otherwise, for the fact of the existence of consciousness is prior to all other facts. If I perceive—am aware of—a chair, I may later be persuaded, by other evidence, that the appearance was produced only by a trick of lighting; I may be persuaded that it occurred in a dream, or even that it was an hallucination; but there is no evidence in existence that could persuade me that my awareness itself was mistaken—that I had not really been aware at all. This knowledge of personal awareness, therefore, is prior to all other forms of knowledge.

If consciousness is the most fundamental fact of all, why is it not used in this book? The answer, in my opinion, is that Science deals, and can deal, only with what one man can *demonstrate* to another. Vivid though consciousness may be to its possessor, there is as yet no method known by which he can demonstrate his experience to another. And until such a method, or its equivalent, is found, the facts of consciousness cannot be used in scientific method.

The problem

1/12. It is now time to state the problem. Later, when more exact concepts have been developed, it will be possible to state the problem more precisely (S. 8/1).

It will be convenient, throughout the discussion, to have some well-known, practical problem to act as type-problem, so that general statements can always be referred to it. I select the following. When a kitten first approaches a fire its reactions are unpredictable and usually inappropriate. Later, however, when adult, its reactions are different. It approaches the fire and seats itself at that place where the heat is moderate. If the fire burns low, it moves nearer. If a hot coal falls out, it jumps away. I might have taken as type-problem some experiment published by a psychological laboratory, but the present example has several advantages. It is well known; it is representative of a wide class of important phenomena; and it is not likely to be called in question by the discovery of some small technical flaw.

With this as specific example, we may state the problem generally. We commence with the concepts that the organism is mechanistic in action, that it is composed of parts, and that the behaviour of the whole is the outcome of the compounded actions of the parts. Organisms change their behaviour by learning, and change it so that the later behaviour is better adapted to their environment than the earlier. Our problem is, first, **to identify the nature of the change which shows as learning,** and secondly, **to find why such changes should tend to cause better adaptation for the whole organism.**