49 *The brain as an engineering problem*

[The following paper represents a breakaway not only from Gestalt thinking, but also from some rather generally accepted neurological doctrines. The title implies that it is *engineering* rather than the more general *physical* analogies which are likely to be appropriate for thinking about brain function. Engineering implies design. Once we know that a system is designed we can consider its efficiency, and how one part serves another part to perform its role. This reflects the extreme importance of the Darwinian revolution in biological thought; for since Darwin, 'purpose' and 'function' can be applied in biology without evoking metaphysics. The same is not, however, true for the physical sciences, except for engineering; which makes engineering – *applied* physics – of special interest to biologists.]

* * *

> *The brain serves to cool the blood*
> Aristotle

> *The brain is like an oven, hot and dry,*
> *Which bakes all sorts of fancies, low and high*
> The Duchess of Newcastle

BIOLOGISTS generally refer to the activity of living organisms as 'behaviour'. When talking about machines, engineers tend to use the word 'performance'. To interchange these words is to raise a smile, perhaps an appreciative smile, but the speaker risks being labelled quixotic. It does appear, however, that the terms 'behaviour' and 'performance' are interchanged much more now than in the past, the reason almost certainly being the influence of cybernetic ideas, which have unified certain aspects of biology and engineering. Some biologists even go so far as to regard their subject as essentially a branch of engineering, and some engineers use examples from biology, such as living servo-systems, to illustrate their principles. The activity of organisms is most often referred to as 'performance' when their efficiency is being considered. Thus play-activity is called 'behaviour', while a skilled worker's activity may be called 'performance'. This change is interesting, for it brings out the influence of the engineering way of thinking upon even lay thought about human and animal activity.

It is worth stressing that physical principles have not always been accepted as appropriate to biology. Aristotle did not make any basic distinction between the living and the non-living, but a sharp distinction was drawn by Kant in the *Kritik der Urtheilskraft* (1790). Perhaps Kant was so influenced by the patent inadequacy of Descartes' attempts to describe organisms in terms of his Natural Philosophy that he was led to say that the behaviour of living systems cannot be governed by causal principles applicable to the physical world. To Kant, living systems are somehow outside the dictates of the laws of nature, and this has been held by some biologists since – certainly as recently as E. S. Russell (1946), who regards 'directiveness' as a special property of living organisms. The influence of Kant's teaching upon biology has been profound and (to the cybernetically inclined) disastrous. Historically, it has led to the creation of special entities to distinguish the living from the non-living, such as Driesch's Entelechy, Bergson's *élan vital* and the Emergent Properties of the Gestalt school of psychology.[1]

We do indeed think of inanimate matter as somehow different from animate matter. If we did not, these words would have no special meaning, for no distinction would be implied. The point is this: is it useful to describe, or to explain, this difference by postulating some *special factor* which is held to be present in animate and absent in inanimate matter? To biologists looking for general explanatory concepts, after the manner of the physical sciences, such postulated special factors must appear harmful. These factors do not enable us to relate phenomena; they do not provide any sort of picture; they do not enable predictions to be made. The trouble with Entelechy, *élan vital* and the rest is that they do not help us to understand. Such terms give a sacrosanct air of life, which may be pleasing, but which

[1] Köhler, in his book *Die physischen Gestalten* (1920), takes a different view from that of most Gestalt writers. He does not suppose that organisms are unique in this respect, but rather that Emergence is to be found in many physical systems. Some philosophers have also taken this view. It leads to the difficulty that 'emergence' is used so generally that it points to nothing special. This point is considered in Gregory, 1953*b* [No. 47].

tends to warn off further enquiry. The Gestaltist's plea for the special nature of 'organic unities' is effectively a warning against attempts at further analysis, the doctrine being that it is *in principle impossible* to analyse the whole in such a manner that its activity can be completely described by the causal relations between the parts. It is, however, just this sort of analysis which is the goal of exploration in the physical sciences. Further, it is important to note the *in principle impossible* here: it is not the complexity of the task which is held to make analysis impossible, but rather the claim that the organic world is such that analysis into parts is doomed to failure, however complete our knowledge of it may be. Curiously, this is regarded by some as an exciting and interesting discovery about living systems. This is an attitude puzzling to those who believe that useful explanations in science should take the form of analysis into simpler elements. Now it *could* be that there is something irreducible about living systems which defies such analysis, but surely we have no right to claim this until the traditional types of explanation have failed for a very long time, and certainly not now while exciting advances are being made in the biological sciences. If we seek the types of explanation found in the physical sciences, *élan vital*, or the concept of Emergence, will appear as doctrines of despair. To postulate such special unanalysable factors is to make a philosophy of pessimism. To say that x is an Emergent Property is to put x into the limbo of the unknown and shut the door upon it, while warning others against peeping through the keyhole.

To regard the brain as a problem in engineering is to look for possible solutions in terms of engineering principles to the questions set by biological enquiry. This chapter is concerned not with answers to specific questions – such questions perhaps as: How are memories stored? How does the eye guide the hand? What are dreams made of? But rather will it attempt to discuss some of the difficulties in taking over engineering methods into biology, and some implications of this approach for the study of the central nervous system.

An alternative to the Kantian doctrine is to say that living systems are *machines*. The cybernetic view is often put in this way, but it has objections. If we use the term 'machine' to include living organisms, it loses its major classificatory use. Further, the term 'machine' is very difficult to define in general terms. We might call a given system a machine though it has no predictable output, displays goal-seeking behaviour, and is in fact indistinguishable in its behaviour from at least simple living systems. If we mean merely that it is man-made, then the distinction is trivial. We cannot get away with an ostensive definition of 'machine' (pointing to all existing machines), for we must allow the possibility of future new kinds of machine, and these could not be included. If animate systems are called 'machines', at least two important things might be meant: (1) that their functioning could be described in terms of known physical principles, or (2) that their functioning could be described, if not in terms of principles known at present, at least in terms of principles which *could* be known to us. This is to say that living organisms are in fact so constituted that we could in principle understand them as engineers or physicists understand their systems. It appears that to call an animal a machine is to indicate that its manner of functioning is not *essentially* different from machines which might be designed or made by men. To deny that animals are machines is, it would appear, to suppose that they *are* essentially different. Those who take the former view feel that existing or possible machines performing similar functions may provide clues as to how animals work, and in particular how their central nervous systems are organized. Those who hold that animals are not machines refuse to accept that this could ever give the whole story. Both types of biologist might well agree that we should go as far as we can in looking for analogies, while being careful not to oversimplify or to accept similarities in a naïve manner.

THE USE OF ENGINEERING CRITERIA FOR DECIDING BETWEEN MODELS OF BRAIN FUNCTION

When a biologist or engineer considers what sort of system might be responsible for producing a given function, he may run up against one of two difficulties: (1) that there does not seem to be *any* known type of system capable of just the observed functions under the given conditions, or (2) that there is a *large number* of possible mechanisms, any of which might provide the required functions. We cannot say anything here about the first contingency, except of course that further observation, experiment or thought might suggest possible mechanisms, but we can say something about the second. It is worth thinking about this, for the principles available for deciding which of various alternative types of mechanism are appropriate are just the principles we need for verifying cybernetic hypotheses. Without such principles we can do no more than guess.

Consider an engineer in a position of doubt about how an unfamiliar machine works. We may take an actual example of a dramatic kind: consider the problem of discovering the manner of function of the control mechanism of an enemy's secret weapon, such as the V1 rockets during the last war. The engineer could make use of the following considerations. First, it was clear that the rocket had been made recently by men in Germany. This knowledge that they were man-made was clearly enormously important, though probably never explicitly stated. Martian rockets would offer many more alternatives, including the high probability of principles quite unknown to us. As it was, new principles were unlikely, though possible. Secondly, examination of rockets which failed to explode revealed many already familiar components such as motors, condensers, valves etc., and a great deal was already known about these. Thirdly, it would seem certain that the rockets must have been designed as efficiently as possible. Now how far does the biologist examining brain function share these assets?

1. Since living organisms are not designed and made by men, any number of new principles might be expected, as in the imaginary case of Martian rockets. As an example, it is now believed that feed-back loops are important in organisms, but these were not known to the engineer until Clerk Maxwell's work in the last century, and there could always be further more or less fundamental principles involved which are so far unknown to engineers.

2. Examination of the brain reveals many identifiable 'components', such as Betz cells and amacrine cells, but the functional properties and circuit potentialities of cells are not as well understood as the functional properties of electronic or mechanical components – and even these have their surprises.

3. Efficiency is a difficult criterion to apply to biological systems for a logical reason: it cannot be assessed without some idea of purpose. It is, however, important to note that the notion of efficiency (and also that of purpose) does not imply specific design for a known end. Thus it might be said that a screw-driver makes a good paint scraper, though it was not designed for that purpose. For something to be said to be efficient, it must be efficient for a stated end though not necessarily for a designed end. Thus if it said that some postulated brain mechanism is more efficient than some other mechanism, we must know what end these mechanisms are supposed to serve, and we must know how to assess relative efficiency towards this end. We may ask, for example, 'how efficient is the eye?' and its efficiency may be measured. Thus its acuity and its sensitivity may be measured and expressed in appropriate units. The difficulty arises when we do not know what to measure through not knowing the functional significance of the structure or system involved. Clearly we could not talk about the efficiency of the eye if we did not know that it subserved vision. If a system is found to be highly efficient, in general but few possi-

bilities are left open when it comes to guessing how it works – not many engineering tricks would be good enough.

When an engineer talks about efficiency he may mean a number of things; he may simply mean that it works well, or that its fuel consumption is low, or that the capital or running cost is low, or a number of other things. If the biologist is to make a reasonable guess at which type of mechanism is responsible for a given type of function, and he wants to use efficiency criteria, he must be clear which criteria it is appropriate to take over, and this raises a number of difficulties. Let us, for the fun of the thing, consider a few engineering efficiency criteria in the context of biology.

(a) Thermal efficiency

This may be used for power systems. The efficiency E of a heat engine is given by $E = W/JQ$, where W is the useful work done by the machine when a quantity of heat, JQ mechanical units, is supplied to it. Since no machine can create energy, W cannot be greater than JQ, so that no engine can have an efficiency greater than 100%.

Now, knowing the total thermal efficiency of a given machine, and knowing the expected efficiency of the type of system by which it is supposed to function, it is clear that if the actual efficiency is higher than predicted, then the hypothesis is false and some other explanation must be sought. If, on the other hand, it is too low some cause for the loss may always be postulated. It follows that where the predicted efficiency is high, more possible solutions are ruled out as too inefficient to be likely, and so the criterion is more useful. This criterion might be used in biology to test a hypothesis about, say, conversion of chemical energy into mechanical energy in muscle. It is hardly applicable to the brain because its thermal properties do not seem important to us, though they did to Aristotle when he regarded its function as cooling the blood.

(b) Information efficiency

Information rate may be defined by the rate of transmission of information defined by binary choices, or 'bits'. The Hartley-Shannon Law, which is basic here, states that

$$C = W \log_2 (1 + S/N),$$

where C is the channel capacity, W the band width, S the average signal power, N the average noise power. Some communication systems are more efficient than others. In particular, a change in the manner of coding the information might make a large difference to the efficiency of the system. Now this does appear to be directly applicable to neurological systems, which is one reason why quantitative estimates of information rate for human subjects, as made by Hick (1952) and Crossman (1953) among others, are of great importance here. The most efficient type of coding is important once we think of the nervous system as handling information and as subject to the same limitations as a man-made system. This criterion has in fact been applied to test between different possible codes adopted by neurones in transmitting information. MacKay and McCulloch (1952) decided, tentatively, in favour of pulse interval modulation for peripheral nerve fibres on this basis.

(c) Capital cost

This is difficult to assess. We might at least say that where general, or some specific, nutriment is in short supply, cells may be 'expensive'. Further, weight may be at a premium, which will limit the permissible number of cells. Also, it might be the case that increase in the number of cells would impose an informational strain on the available information coded in the gametes.

(d) Running cost

Not much can be said about this, beyond the obvious point that if food is scarce it will be an 'expensive' commodity.

(e) Simplicity

This is difficult. The engineer favours the 'neat' solution to a problem, and he dislikes certain 'complicated' types of mechanism. This may be in part due to the aesthetic appeal of simplicity, but simple mechanisms perhaps also tend to be cheaper and more reliable, though not necessarily so. Carburettors for petrol engines have in fact become more and more elaborate, with gain in running economy and overall reliability. Have we any reason to suppose that we should find in nature the 'simplest' way of going about engineering problems? Certainly nature is handicapped by lack of many materials and techniques indispensable to the engineer. It is striking, for example, that flight with flapping wings is for an engineer more complicated, and in every way inferior to, flight with fixed wings, though the former is found in nature. But then nature has not got a suitable engine to provide forward drive independently of the wings, and nature has not got true bearings, or the wheel. This case is far removed from neurology, but neurology also provides examples. A familiar one is that of the retina, which is 'inside out'. The light has to pass through layers of blood vessels, ganglia and supporting cells before it reaches the receptors. This optically shocking arrangement appears to be dictated by embryological, or perhaps basically developmental, considerations. Considerations of this kind make the use of the criterion of simplicity difficult and dangerous to apply.

(f) Length of life

Some types of machine outlive others. There are many reasons for this – choice of materials, friction between moving parts and many more. This criterion for deciding between rival designs can hardly be applied at present to the living machine because of its self-repairing properties.

We conclude that some engineering design criteria can be applied to biological systems in order to ascertain which, among many possible types of mechanism, is the most likely to be operating in any particular case. Efficiency criteria, particularly thermal and information efficiency, seem to be the ones most readily applied in the biological context, but in some cases other criteria might also be used. If this way of linking behaviour study with neurology is adopted, then rather precise 'engineering-type' data will be required. It is unfortunate that it appears difficult to apply the other criteria commonly used by engineers. As a result, cybernetic writing easily becomes science fiction, where the supposed theories and mechanism may be limited in variety only by the imagination of their inventors. This is unfortunate for a vitally important approach to biology.

LOCALIZATION OF CEREBRAL FUNCTION

What is meant by saying that some feature of behaviour is localized in a part of the brain? It cannot mean that the behaviour itself is to be found in the brain, or that a region of the brain can be sufficient for any behaviour. The intended meaning is that some necessary, though not sufficient, condition for this behaviour is localized in a specific region of the brain.

The evidence for localization is mainly from studies of ablation and stimulation of regions of the brain. If, for example, when a point on the occipital cortex is stimulated, flashes of light are reported by the patient, it is generally held that this region of the cortex must be important for vision. If an area in the left frontal lobe is damaged and speech is found to be disturbed, it may seem that we have found something causally necessary for speech. But have we?

This area may be *necessary* for speech (i.e. if it is removed, speech may disappear) but so also are a number of other parts of the organism, for example the vocal cords, the lungs and the mouth. There is nothing special about the brain here. It may be that the 'speech area' is concerned only with speech, but if so it is not unique in this respect either: if we except coughing, the vocal cords have no other function but to subserve vocalization. Now we may say that the vocal cords are *causally necessary* for speech, and also that the 'speech area' is somehow *causally necessary*, but it is not clear in the second case just what the causal functions are, though we do understand the causal role of the vocal cords. There is an important point here: we may say that *A* is the cause of *B* if *A* is found inductively to be a necessary condition for *B*, and the evidence may be purely inductive for this type of causal argument. No understanding of the mechanisms involved is required to assert the causal relation between *A* and *B*. But we may also say that *A* causes *B* on *deductive* grounds, when we understand (or think we understand) the mechanism by which *A* produces, or causes, *B*.

Once we distinguish these two types of argument from physical structure and function to causal relationship, we should ask which sort of causal argument is being used in discussions about brain function. Take the case of the speech area. It would appear that the reason why this region of the brain is held to be associated with speech is that speech is found to be defective or absent when the region is damaged. This is clearly an inductive argument, and it does not presuppose or imply any knowledge of how the speech area works, or what causal part it plays in the production of speech. Again, we know fairly clearly the causal role of the vocal cords, but not that of the 'speech area'.

Consider now the word 'function'. We may say that it is the *function* of the vocal cords to vibrate in certain ways, producing pulses of air which resonate in cavities . . . we see the causal role of the vocal cords and we come to understand the mechanism of speech production. And now what about the word 'localization'? What is it to say that a *function is localized*? The question is: How can we say that a function is localized until we know what the function (of a given bit of brain tissue) is? To say this we need to know in some detail how the system works. It seems that before we can talk usefully about localization of function we must have some idea of *how* the system works.

It might be interesting to consider how an electronics engineer deals with, and represents, specific functions in a complex device. He uses three types of diagram to represent an electronic machine. These are *(a) blue prints*, showing the physical locations in space of

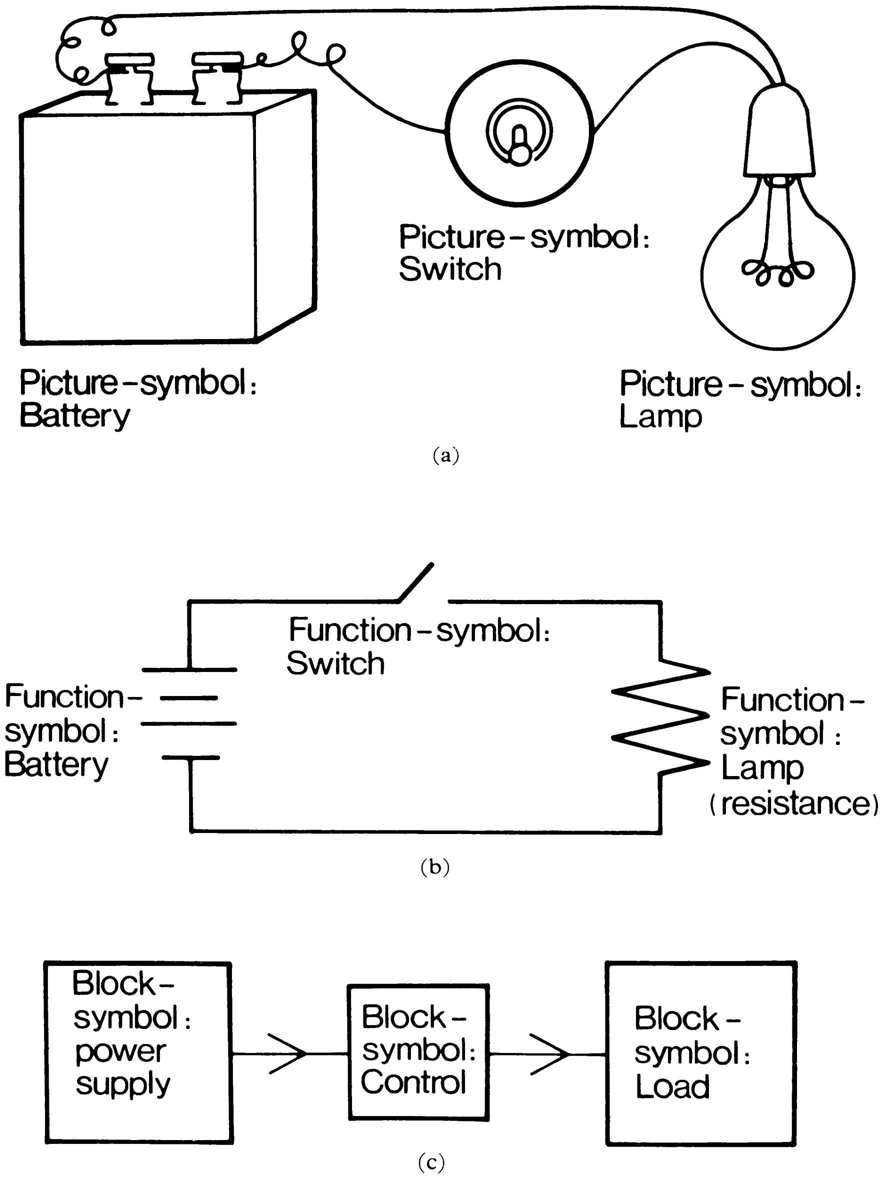

FIGURE 49.1

(a) *Shows a simple blue print type of diagram. Pictorial representation of components are linked with paths of conductivity (wires), also shown pictorially. This may be compared with histological descriptions of the brain.*

(b) *Shows a circuit diagram. The symbols show conventional functional properties of the components. This may be compared with the physiological descriptions of the brain.*

(c) *Shows a block diagram. The blocks show the functional units of the system, indicating the causal processes in terms of the flow of power or information. This may be compared with cybernetic descriptions of the brain.*

the components, with their sizes and shapes. These drawings will give dimensions and describe the structure, so to say the cyto-architecture, of the machine. *(b) Circuit diagrams*, in which each component is shown in diagrammatic form with its connections (usually paths of perfect conductivity) with other components having idealized properties. *(c) Block diagrams*, in which there are a number of boxes connected together with flow lines, each box being labelled with its 'function'. Thus one might be labelled 'Radio frequency amplifier', another 'mixer' and so on. Now what knowledge of the system is conveyed by *(a)*, *(b)* and *(c)*? And what knowledge, therefore, must the man who designs the diagrams have of the system? *(a)* requires a knowledge of the look of the thing; *(b)* implies information about certain selected properties of the 'components' of the system. (A capacitor in series marked 0·1 μf., for instance, conveys a great deal of information about what will happen for various conditions – for example, that direct current will be blocked while high frequencies will hardly be affected, except for a phase shift.) But the general effect of these changes produced by a component will not be apparent except in terms of *(c)*, the block or function diagram. The condenser may then turn out to be part of an oscillator, and then its purpose within that functional section of the device could be stated. For example it might provide feed-back between anode and grid of a valve. Now when a trained engineer looks at a circuit he can very often see almost at a glance what it will do and how it does it. With an unfamiliar circuit this might be difficult, but he could probably predict its function and performance given the circuit valves and a slide rule. To do this he will make use of a number of generalizations, or Laws, such as the conditions for obtaining oscillation, or linear amplification or whatever it is. The point is that *(c)* implies these general principles while *(b)* does not. Only a skilled engineer could design or work out a block diagram, but anyone who can recognize components by their appearance could draw a circuit diagram. The question for neurology is: Do we want diagrams of structure of the 'circuit' with component characteristics, or functional block diagrams? I would suggest that only the last describe how a system works. They are comparatively simple, but alone convey (and imply for their drawing) a knowledge of the function of the system.

The neurologist can discover the properties of *his* components, the neurones; he may discover how the components are wired up. Can he discover the causal mechanisms involved? Can he find out where they are situated in the brain?

There seems to be a widespread hope that, by ablation and stimulation of parts of the brain, functional regions may be discovered, these being logically the same as the boxes in a block diagram. Is this a reasonable hope?

ABLATION AND STIMULATION AS TECHNIQUES FOR DISCOVERING FUNCTIONAL REGIONS OF THE BRAIN

Suppose we ablated or stimulated various parts of a complex man-made device, say a television receiving set. And suppose we had no prior knowledge of the manner of function of the type of device or machine involved. Could we by these means discover its manner of working?

In the first place, and most important, to remove a part of a machine, even a discrete component, is not in general to remove a necessary condition for some feature of the output. If a part is removed from a complex machine, we do not in general find that simple elements or units are now missing from the output. It should be noted here that the functional processes taking place in the components, or groups of components, of a machine are generally quite different from anything in the output. Thus we do not see the spark is a car engine represented in its output – we see wheels turning and the car moving: no spark. If a component is removed almost anything may happen: a radio set may emit piercing whistles or deep growls, a television set may produce curious patterns, a car engine may back-fire, or blow up or simply stop. To understand the *reason* for these 'behavioural' changes we must know at least the basic principles of radio, or television, or car engines, or whatever it is, and also some of the details of the particular design. Of course, if we already know about radio, or engines, then these abnormal manifestations may well lead to correct diagnosis of a fault: the difficulty is to reverse the procedure.

Consider a television set which has, of course, two quite distinct outputs – sound and vision. Some 'ablations', or 'extirpations', may quickly reveal which parts are *necessary* for each output, and also which parts are *common* to the two outputs. In the case of the brain, there is a large number of inputs and outputs: the limb movements, the face with its various expressions, the voice, and so on. It may be a fairly simple matter to discover regions of the brain which are necessary for these various outputs, and in general they will lie near the peripheral output of the system. The inputs, the senses and their projection areas, we might also expect to locate in this way without undue difficulty. What I suspect *is* difficult, indeed impossible, is to locate functional regions of the system. It seems to me that this conclusion is forced upon us by considering the possibility of isolating elements of a complex output in a single channel in the case of man-made machines. In a serial system the various identifiable elements of the output are not separately represented by discrete parts of the system. Damage to a part may indeed introduce quite new factors into the situation, and these could only be comprehensible when we are provided with a model indicating the function of the parts. If the brain consisted of a series of independent parallel elements with separate output terminals for each, like a piano, it might be possible to identify behavioural elements with particular parts of the system, as the various notes of the piano might be regarded as being 'localized' in the piano; but where output is the result of a number of causally necessary operations taking place in a series, then this is not possible. The removal, or the activation, of a single stage in a series might have almost any effect on the output of a machine, and so presumably also for the brain. To deduce the function of a part from the effect upon the output of removing or stimulating this part we must know at least in general terms how the machine works. The point here, perhaps, is not so much that the piano is a parallel rather than a serial system, but that it is a set of largely independent machines in one box. Where they do interact, as in the pedal systems, then one 'ablation' may affect all the notes. Parts of the brain could be independent.

The effects of removing or modifying, say, the line scan time-base of a television receiver would be incomprehensible if we did not know the engineering principles involved. Further, it seems unlikely that we should discover the necessary principles from scratch simply by pulling bits out of television sets, or stimulating bits with various voltages and wave forms. The data derived in this way might well lead to hypotheses once we knew something of the problem in engineering terms.

But we should, in some systems, be able to map projection areas and delimit pathways, and this is a good deal. Analogy with familiar physical systems strongly suggests that to go further these studies should be used to test rival hypotheses of brain function, rather than to attempt to isolate functional regions. This brings us back to the idea of physical model explanations, with ablation and stimulation studies as one way of trying to decide between rival models. We are left with the difficulties besetting this approach: in particular, the brain might work on some novel principle, and then its true manner of function would never come up for testing by any experimental technique. It would clearly require a most highly sophisticated set of techniques to discover a quite new principle in the living brain, but this is conceivable. Perhaps the principle of scanning, or heterodyning, could be discovered by these techniques, even in a jelly.

It is a common finding that with electronic equipment several very different faults may produce much the same 'symptom'. For example, anything which produces a change in the supply voltage will first affect the part of the system most susceptible to supply changes, and so anything affecting the supply will tend to produce the same fault. To aggravate the position, faults affecting the supply voltage are not limited to the power pack supplying the voltage to the various parts of the system, but may be in any of these parts, increasing or decreasing the load and so affecting all the other parts in greater or lesser degree. Thus the removal of any of several widely spaced resistors may cause a radio set to emit howls, but it does not follow that howls are immediately associated with these resistors, or indeed that the causal relation is anything but the most indirect. In particular, we should not say that the function of the resistors in the normal circuit is to inhibit howling. Neurophysiologists, when faced with a comparable situation, have postulated 'suppressor regions'.

Although the effect of a particular type of ablation may be specific and repeatable, it does not follow that the causal connection is simple, or even that the region of the brain affected would, if we knew more, be regarded as functionally important for the output – such as memory or speech – which is observed to be upset. It could be the case that some important part of the mechanism subserving the behaviour is upset by the damage although it is most indirectly related, and it is just this which makes the discovery of a fault in a complex machine so difficult.

We may consider one or two further points. Since learning is important in at least the mammalian nervous system, it is clear that where animals and men have had different past experiences their brains are likely to be in some ways different. What is 'stored' must at any rate vary between individuals of the same species. It is known that for man surgical removal of some areas of the brain, e.g. the frontal lobe, may pass almost unnoticed in some individuals, while in others it produces serious defect of function. This might perhaps be due to the different importance of specific causal mechanisms in individuals employing different 'strategies', or possibly to the unequal importance of various pieces of stored information. In any case we should expect, and do in fact find, individual differences. This is a complicating factor in interpreting ablation studies which would hardly concern an engineer using man-made machines, except indeed for certain electronic computers.

A further point that might be made is this: Suppose we ablate or stimulate some part of the brain, and lose or evoke something in behaviour, then it is not clear – even quite apart from previous considerations – that this region is the seat of the behaviour in question. Might it not lie along a 'trunk line' or 'association pathway'? A cut telephone line might affect communication over a wide area, principally behind the region of damage. This has at least two important implications: first, unless the region is known not to lie on a 'cable' the region cannot be identified with a brain 'centre' responsible for some aspect of behaviour, since the 'centre' responsible but cut off might lie anywhere from this region along the trunk line. This is further complicated by the consideration that it might be cut off in some conditions but not in others: it might conceivably depend upon whether the animal is motivated in a particular way, receiving information from a particular 'store', or countless other possibilities, whether this block will matter; and the same is true of damage to, or stimulation of, a 'centre', even if this word is taken as meaningful. In many machines it might be possible to remove large parts without any effect except under certain working conditions.

There are two points here: (1) damage might produce, so to say, a shadow within which brain function is lost to regions of the brain on the 'other side' of the damage. If the better analogy is a short-circuited power line, the effect may extend both ways along the cable. (2) The damage may be important only under certain critical circumstances. It does not matter that a car's trafficators are not functional until the driver wishes to turn a corner in traffic – or that his brakes do not work until he tries to stop.

This view of what we mean by 'function' is important in considering brain 'centres'. These are supposed loci for particular types of behaviour: thus Hess has a 'sleep centre' for the cat, in the hypothalamus. This idea of 'centres' has been taken over by Ethology and is particularly important in Tinbergen's writings. But we may well feel worried about the concept of functional centres when we do not know what is going on, in functional terms, in the region concerned. The above considerations apply here *mutatis mutandis*. Why, if stimulation of a given region produces sleep, should this region be regarded as a 'sleep centre'? To take a facetious example: if a bang on the back of the head produces stars and a headache, is this a 'centre' for stars and headaches?

In summary:

1. It might be argued that 'localization of function' means that some feature of behaviour has certain vital (but not sufficient) causal mechanisms located in a given region of the brain. But before we know

how, in general terms, the brain works we cannot say what these supposed causal mechanisms are, and thus it is very difficult to say what we mean by 'localization of function'.

2. Stimulation and ablation experiments may give direct information about pathways and projection areas, but their interpretation would seem to be extremely difficult, on logical grounds, where a mechanism is one of many interrelated systems, for then changes in the output will not in general be simply the loss of the contribution normally made by the extirpated area. The system may now show quite different properties.

3. It would seem that ablation and stimulation data can only be interpreted given a model, or a 'block diagram', showing the functional organization of the brain in causal, or engineering, terms. Such data may be useful in suggesting or testing possible theoretical models.

4. These models are explanations in the engineering sense of 'explanation'.

CONCLUSION

It would be nice to say something more constructive about the use of engineering thinking in biology. Given that there are certain difficulties in taking over engineering ideas of design into biology, can we not still use engineering techniques and devices to make some better-than-random guesses about how the brain works?

We have throughout looked at the brain as an engineering problem in a general way: we have not considered any particular engineering techniques, or mechanisms, or machines which might throw light on biological function. We have mentioned radio sets and car engines when thinking about localization of brain function, yet it is at least clear that brains are very different from these. We could certainly think of machines more like brains – and this might be worth doing. What about computers? Obviously we should expect more similarities between computers and brains than between car engines and brains, for the inputs and outputs are similar for the one though not for the other. Now we might go further and ask: what *sort* of computer is most like the brain? There are many different types of practical computer. As is well known, they are divided into two main classes: analogue and digital. Each has certain advantages. The former are usually simpler in construction, they are fast, and are generally subject to rather large random errors. The best-known example is the slide rule. Their inputs and outputs are usually continuously variable, though this is not always so: a slide rule might be made with click stops and still be called analogue. The essential point is that the input variables are represented by the magnitude of some physical variable, such as a length or a voltage. Digital computers, on the other hand, are generally slower, and their answers tend to be either correct or wildly wrong. They work in discrete steps, and according to some fixed rules or calculus. The functional units (essentially switches) of a digital computer take up certain discrete semi-stable states according to a code. For some purposes the analogue type would be chosen by the engineer, and for others the digital type. Thus we may now ask: which would be the most suitable type of computer for a brain, an analogue or a digital computer? Or perhaps a mixture? To answer this question we may make a list of the relevant properties of the brain and try to decide which type of computer fits best [see No. 47]. Some of the difficulties we anticipated at the beginning: we found that engineering criteria are not easy to apply, and that some are indeed inappropriate. The basic efficiency criteria evidently may be applied, but they have their difficulties unless we know a good deal about the functional properties and efficiencies of the components of the brain. Thus it is not possible, for example, to say whether the brain works too fast to be a digital computer unless we know the rate at which the components can change their states, or count. If we also knew the minimum number of steps logically required to reach a given solution with the available data it would be possible to say whether the brain *could* work digitally.

Similar considerations apply to testing the hypothesis that the brain is an analogue machine. We may ask: is the brain too accurate for an analogue machine? We cannot answer this until we know how the 'templates' representing the variables work; we need to know more about the actual ironmongery available; 'ideal' considerations are not adequate here, we must know the properties of the components. If we invoke feed-back principles the brain might be an analogue device given rather variable templates – there are many such 'saving' possibilities. In fact this view that the brain is in essential respects analogue is perhaps borne out by the type of errors observed in control situations. The point is that engineering here supplies the hypotheses for testing, and also (up to a point) the manner of testing them, but to make these decisions it is important to know in detail the functional limitations of the components of the brain. It is also important to have 'engineering' performance data. Much experimental work in psychology is in fact undertaken for this purpose. It may well be vital for linking psychology with neurology, and we should use engineering concepts both to suggest appropriate experiments and to integrate and interpret the available data. For example, studies on tremor take on a new significance within the context of servo-theory, for all error-correcting servos are subject to 'hunting'.

A rather different approach, which we might do well to adopt, is the following: We might look for what we are virtually certain to find and then measure it. Two, rather different, examples must suffice. First, we believe that a system cannot itself gain knowledge without inductive generalization, and we know that this is impossible without probability estimates. This involves some form of counting, and some form of store for count rates or relative count rates. This at once suggests that the brain should be looked at as in part an inductive machine (e.g. Gregory, 1952*b* [No. 46]). Probably no one had actually built an inductive machine until Uttley (1954*a* and *b*) built his, specifically as a possible model of brain function, but the man-made induction machine follows standard engineering principles. To go to the next stage and ask whether the brain is the *same sort* of induction machine as Uttley's raises all sorts of difficulties, some of which we have already discussed. The point here is that we believe on *very general grounds* that probabilities must be important to achieve adapted behaviour, and so induction and probability mechanisms really must be found if we look for them.

The second example of this approach is the interesting though more specific problem of 'noise' in the nervous system. It is well known that all communication systems are ultimately limited by random noise, which tends to cause confusion between signals. It seems impossible that the nervous system can be an exception, and so it is hardly a discovery that there is 'noise' in nerve fibres, and in the brain. The assessment of the actual 'noise' level in the various parts of the nervous system (Gregory and Cane 1955 [No. 7]; Gregory, 1956; Barlow, 1956, 1957*a*) and of changes in 'noise' level due to ageing or brain damage [No. 8] may throw some light on neural function, if only by helping us to apply efficiency criteria to test between rival explanatory models. It is interesting in this connection that Granit (1955) has recently summarized the evidence for random firing of the optic nerve but has not interpreted this as a background 'noise' level against which visual signals must be discriminated, but rather regards it as necessary for keeping the higher centres active. Thus the same observation might be regarded as a necessary evil or a special and useful part of the mechanism. Here the very general properties of communication systems would lead us to the former interpretation, but without these general considerations there would have been no reason to suppose that random firing is not useful to the organism and, so to speak, part of the design. Given the engineering viewpoint, we should ask how the system is designed to *minimize* the effect of the background noise, and this is quite a different sort of question, leading to quite different experiments.

Information rates and noise levels will not in themselves tell us how the ear or the eye gives us useful information – how they work – but such measures are in conformity with the engineer's insistence upon knowing the performance limits, and the reasons for the limits, of his systems. Experimental psychology is currently, and for practical reasons, concerned with the limits of human ability in many directions, e.g. in steering and guiding. These measures may be vital in deciding how the guiding or steering is done. In many cases it is only limits, such as sensory thresholds, which can be used to provide 'engineering' data from complex organisms. Now this idea of looking for properties which are found in all, or at least in most, engineering control systems, and then obtaining quantitative measurements of them under various operating conditions is rather different from the idea of thinking of a physical model as a possible 'analogy' to a behaviour mechanism and then testing this model with observation or experiment. Before we attempt seriously to test specific models of brain function – types of memory store and the like – we might do well to make careful estimates of such things as neural 'noise' levels which we are virtually certain must be there to be found. Having done this, we may be in a stronger position to test specific hypotheses, for we should be able to apply engineering criteria with sufficient rigour to make some hypotheses highly improbable, while others might be shown to be quite possibly true.

These considerations have some relevance to the progress of experimental psychology. If we have no idea of the sort of system we are dealing with, controlled experiment becomes impossible, for we cannot know what to control. On the other hand, a too fixed and particular model tends to blinker the mental eye, making us blind to surprising results and ideas without which advance is impossible.