Mechanistic Teleology and Explanation in Neuroethology: Understanding the Origins of Behavior

G. Adrian Horridge

Introduction

Neuroethology is the explanation of behavior in terms of neurons. These words deserve examination.

Explanation, the most difficult word to expand upon, has long fascinated man, and types of explanation are in short supply. First, we can *explain* in terms of antecedent material causes modeled on the blow of a stone that breaks another stone. Second, we can explain by invented unseen abstractions, like the spirit of heat that explains what passes from man to wood in the making of fire. The history of modern science has been the story of material antecedent causes replacing invented abstractions as causes, but when material causes are examined more closely they too are found to be full of invented abstractions. Third, we can explain in terms of future needs. A system is made in a particular way in order to perform a function. Purpose or design of animal parts is perfectly respectable since Darwin's elucidation of natural selection as the basis of purposive design. During evolution, animals grow components in combinations that will subsequently interact in ways that ensure survival of the whole population. Thereby a circle of material cuases is completed and a happening at any time in the cycle can be explained by causes that act at any other time in the cycle, earlier or later.

Behavior includes all that happens when motor and secretory neurons are active, so that salivation, gut movements, and change of color are as much parts of behavior as patterned song, discrimination, or learning. *Neurons* are the only components of nervous systems so far known to be really significant in generating behavior. Where the components of a system can be discerned or inferred, the mechanisms of that system are understood, in the usual sense of that word, only when they are worked out as the consequences of the interactions of the components. This mechanistic explanation of what goes on is a prerequisite for any other form of explanation that may be added. The arrangement of the connections be-

G. Adrian Horridge • Department of Behavioural Biology, Research School of Biological Sciences, The Australian National University, P.O. Box 475, Canberra, A.C.T. 2601, Australia.

tween the neurons is just as relevant to mechanistic explanation as synaptic potentials, spikes, slow waves, synthesis or release of secretions, and so forth. If components other than neurons, such as glial cells, are shown to be active in behavior, they will have to be treated in the same way as the neurons. The problem is to get beyond this reductionist approach.

Some Current Forms of Explanation

Let us now take a look at some of the explanations that we actually use or seek in neuroethology.

Reductionist or Mechanistic Analysis

At the present time, we understand that the discovery of progressively more detailed material causes is the road of progress. First, we list the components and forces of chemistry and physics that are acting in the particular system we study. We analyze every little twist and detail to uncover what acts on what. Later, the interactions are described quantitatively, and we obtain mathematical expressions which not only fit the observed interactions but also embody in the theory a breakdown into the interactions and components that act together in producing the observed result. Theories without this characteristic are suspect. Invertebrate nervous systems are particularly approachable by this kind of analysis; with their identifiable neurons and simple circuits, they reveal principles of interaction that are seminal in the understanding of more complex systems like the human brain, which cannot be analyzed so cleanly.

Drawbacks of Reductionism

Mechanistic analysis with a description of all detail is not, however, of itself sufficient, for several reasons. The question is, how can the function or relevance of neuron activities be introduced into the mechanistic analysis? An undisclosed set of rules guides researchers and journal editors about what is useful or what is significant science. On what principles should these guidelines be based? I take the view that details of neuron activity that are not related to behavior in a rather direct way are not acted on by natural selection, and therefore they cannot be treated as if they are a meaningful part of the system. Most of the vast amount of unquoted results on behavior, especially of man, and of neuron activity and anatomical detail, especially of vertebrates, in the literature of the past 50 years are dead because no functional relevance can be found for them.

In a mechanistic analysis of neuron activity, the detail becomes tedious because the interactions and complexity are potentially infinite and no one beyond the original investigator will ever be interested. Analyzing the interactions of all neurons could occupy the human race for all time, if only because the more they

knew the more ways of analysis they would know, and the more of their own neuronal activity they would have to describe. Even for one species of insect, the whole activity of all neurons is more than we seek. We seek *principles of action,* mainly, and detail becomes significant only when it elucidates principles. We still have the question with us, however, as to what principles of action we accept as fulfillment of the investigator's task, and in what way can the complexity of nervous systems be explained.

Methodological Objection to Reductionism

In the study of a small part of a nervous system, it is a temptation to think that the output is understood when all interactions and nonlinearities of neuron activity are known. The operative word is "all," and turning the sentence around soon reveals that the "all" is no more than a hope. "When the output is understood, all interactions and nonlinearities are known." That is to say that we go on looking for interactions in the system while there remains something to be explained in the output. This is an excellent formula for bread-and-butter elucidation of components and their activity. The catch is that the outputs and behavior of the nervous system are always far simpler than the inputs and central processes. Therefore, the behavioral or motor neuron (MN) outputs can be fully explained without necessarily making use of all the activity within, and all the possible interactions of inputs. One can never know *all* interactions inside, or when the complete list of nonlinearities has been found. The system is an open-ended one that we progressively reveal, not a circumscribed one in which the components and processes have already been listed. So much for pure mechanistic analysis: I believe that even as a self-contained exercise on a small subsystem it cannot be complete, and, besides, it is unsatisfactory because relevance and optimization are omitted.

Reductionism Is Description, Not Explanation

When every detail of what goes on in a small part of a nervous system is described, what do we have? We see how the sensory inputs are processed by components on the input side of the watershed of the central nervous system. Interneurons (INs) act on each other to generate behavior patterns that are carried out by MNs. The initial *belief* that the components work the one upon the other, in the way that components of an engine or radio act upon each other, is to a large extent vindicated. Descriptions of interaction, however, fall short of an understanding of why the structure and function of the system are as they are. The engineer putting together components has in mind the *function* of the machine that he makes; he does a good or a bad job insofar as the machine is effective in its intended function. In the living nervous system, natural selection acting over many generations has ensured that components work together: the system fulfills a function. Behavior, from the beat of a jellyfish to the thought processing of man, fits the real world. A deeper understanding of the structure and function appears when the

mechanistic analysis has proceeded far enough for us to propose that all parts of the system are as they are because they are optimized to a particular function or display a compromise between optimizations for more than one function. When that is possible, we really begin to understand what we are studying.

A Clue from the Biophysics of Small Components

When we turn to the small components of nervous systems, we find that much more is known about the way they act. Moreover, many details of their physics and biochemistry can be related in considerable detail to their function. Often the mechanisms can be taken right down to components at the molecular level. *Form and function then cease to be separable ways of thinking about the same system.* At the next level of complexity, the cellular level, the acceptable current view is that the action and the anatomy lie within rigid limits that are set by the laws of physics acting in relation to the supposed *functions* of the organs or cellular organelles. Anatomical examples are the fine details of receptors, particularly auditory and visual receptors, in terms of the properties of sound or light; the spacing and adhesions of membranes; dimensions of nodes, axons, and dendrites in terms of space constants; the forms of synapses in terms of the diffusion laws; and the architecture, on large and small scale, of the effector organs such as cilia or muscle filaments. The two arguments for biophysical explanations of this type are as follows: First, this approach is heuristically successful as shown by the agreement between the efficient performance of the organelle and the laws of physics by which it is supposed to work. We readily accept the conclusion that form and function are together optimized. Second, the same structures occur throughout the animal kingdom. Sensory cells, details of neurons, synapses, cilia, and muscle fibers, including details of their operation, are surprisingly constant throughout the animal kingdom from coelenterates through vertebrates. We accept that they are perpetuated by the historical process of evolution and at the same time determined by the laws of physics in the form that we find them. The issue is whether we use the physical analysis in conjunction with the important assumption that each subsystem, and the whole working together, has been optimized by natural selection for a particular functional compromise, and that components are not merely a legacy, appropriate or otherwise, inherited from previous models. For small components that are well understood, the physical explanation as to why they are as they are is always conceived in terms of the function to be performed. My thesis is that to make nervous systems comprehensible the neurons and their fields must be treated in the same way.

Why Nervous Systems Are as They Are

A major difficulty is to explain many of the easily discoverable facts of neuronal organization: for example, why one neuron should run to two muscles in

the crayfish claw, or why the common inhibitor neuron of the thoracic ganglia of some insects sends a branching axon to many muscles; why the cricket has ears on its elbows; why many insects with binocular vision have eyes that are close together. A new question this year is why should there be nonspiking INs in insect ganglia? The quest appears to be for an explanation as to how that particular arrangement comes about and why it persists. We really want to know why that particular arrangement is the best or an optimum. In standard tests we often find two quite different types of explanation offered, neither of which depends on previous completion of the reductionist analysis of what goes on.

The Legacy of Evolution

The animal has a history. History is a powerful determiner of process. Only certain ways of evolving are possible on account of the constraints against other lines of change for the inherent properties of the system. It may be that nerve ganglia appeared early in evolution of the phyla because only in the ganglia can the nonspiking INs interact closely with each other. More advanced forms may have progressively improved the interaction between nonspiking local INs without being able to get away from the basic mechanism itself. Increased interaction between nonspiking neurons is provided by the fusion of ganglia in the process of cephalization that has happened many times in evolution. The ganglia of mollusks and the ladderlike nerve cords of arthropods, the details of eyes, ears, and in fact all parts of all animals present us with many structures that are explicable in terms of *history*. One can't escape this: animals are not designed from scratch according to principles of optimization. They are optimized in different ways from a historically given ancestor. This, however, is not an explanation of how they work, or how the working mechanisms are optimized.

Design around the Properties of Components

The second type of explanation of why a particular arrangement comes about is made in terms of a mechanical effectiveness of necessity that derives from the properties of the materials of construction, from the geometry of the situation, or from limitations on what can be grown in development. The nervous system is in fact composed of most unpromising materials. Usually our inability to produce this kind of explanation, and therefore our tendency to continue the search, springs from lack of sufficient detail about the way the system works, and therefore the complete mechanistic analysis is a prerequisite. Essentially, this type of explanation is the demonstration that what is found is a good way of using the available materials, and that it makes sense in terms of function. In the same way, we cast a knowledgeable eye over any highly developed human construction, such as a sailing boat, a bicycle, or a cup and saucer, and see that it is so built because the nature of the materials has allowed and encouraged optimization along particular avenues but not along

others. A further twist in this explanation is that in development the proteins can fit together only in certain ways.

The Classical Analysis of Neuron Function

In the classical and much current analysis of nervous systems, first the MNs and INs were isolated as components, and their inputs in some cases and their outputs in others were described. The main contribution was functional anatomy based on the neurons as excitation pathways. In many invertebrates, it is possible to name each neuron uniquely and return to it over and over again in different animals, so that its properties can be progressively worked out as ideas progress. In vertebrates, only classes of neurons can be identified, so that neuron properties are always representative or statistical, and one can rarely return to the same neuron. With difficulty, one can find pairs of neurons, one of which acts directly on the other, but in general in vertebrates the chain of causation is based on probabilities. This itself sets a severe limit on how much the chain of causation can be worked out between components. The invertebrate named neurons provide the seminal ideas.

In this type of analysis, the *input field* of the neuron, together with the output field, is the property to be described, and actions of neurons may be quantified as transfer functions. The field is not yet seen as a joint consequence of the laws of physics and natural selection acting on that part of the neuron activity which eventually appears in behavior. This kind of mechanistic analysis, however, must continue until the components and their basic properties are listed.

The next stage is the description of the interactions between identified neurons. This becomes the study of the flow of excitation to higher-order cells, with three main themes:

1. Fields of sensory neurons, and progressive pattern abstraction as these become fields of higher-order neurons.
2. Patterns of behavior that are generated from the arrangement of the connections between neurons.
3. Analysis of behavior to include feedback loops from the animal's surroundings.

One deficiency of this method is that it omits all the slow, hormonal, secretory aspects of the nervous system, and these are probably as important as the impulses seen on oscilloscopes. Another defect is that the results of recording fields at different levels, even in the best-studied systems, simply do not provide enough data for us to see in detail which neuron acts on which others. Present indications are that this method is not adequate to yield a reductionist description in an isolated and simplified subsystem. The situation is impossible to analyze where there are a number of parallel pathways and where any stimulus is likely to excite an unknown number of neurons, *however indirectly*. The imponderable is the activity in neurons other than those observed by the recording electrode. In these circumstances, an input–output relation, or an interaction of one neuron and another, recorded even in

identified neurons, is *not* an isolated interaction, but is one part of a complex system with other activity going on at the same time. We rarely test for univariance, give proof of which anatomical synapses are responsible for our interaction, or exclude parallel activity. *Analysis depends on having the system extremely restricted.* Where there are pathways in parallel, some leading toward behavioral outputs, it is an essential but almost impossible part of the analysis to discover how the excitation is partitioned between them, and how the partitioning changes under different circumstances. Therefore, except in systems of few fibers, the method of recording at different levels in a nervous system simply fails to give a complete picture. The result is that worker after worker records from a restricted subsystem and then, having gone as far as possible, builds a model with inferences extended beyond the observations, before turning to a new system to analyze.

Explanation with and without Consideration of Design

On the basis of the above discussion, I therefore distinguish between descriptive reductionism and a higher form of mechanistic analysis as follows: A mechanistic or reductionistic explanation in terms of components sets out to describe the causes of the activity of the nervous system, i.e., the origin of behavior in terms of what component acts on what, how it does it, and when. Mechanistic analysis at its best reveals what acts on what and employs all the relevant laws of physics, but, first, it can never be complete, and, second, its ultimate aim is in the wrong direction.

Quite different is the explanation of why the system, or its individual components, or its behavioral output, is as it is, because they are optimized to fit in with each other and with the environment. Both forms of explanation are valid in that they are sought by scientists and accepted when demonstrated. Explaining why things are as they are requires all the relevant laws of physics, but in addition considers the function of the system and the function of each component. With the help of comparative data, where compromises between different demands can be seen to be met in different ways in different animals, we can show that each variant of the system and its components are beautifully adapted to the diversity of the needs that are met. They are optimized within the limitations also set by history, by the properties of materials, but primarily by the laws of physics. Revealing exactly how these compromises, limitations, and optimizations apply is a valid form of explanation that I call "mechanistic teleology" because it is the physical explanation of the stringent design criteria that are needed for selective advantage. Numerous elegant examples are available showing how primary sensory cells, and sense organs such as the eye or ear, for example, are superbly adapted to make best use of the physical principles which govern light or sound. There are also brilliant accounts of the appropriateness of neuron fields in that they select exactly the combination of inputs that are of greatest selective advantage. The MIT foursome (Lettvin *et al.*, 1959) illustrate this very well. However, that pioneering work did not have the influence it deserved. The mainstream of neurophysiological analysis continues to

publish *our* reading of neuron activity, not the nervous system's reading of it. The study of central neuron fields has not been combined with the appropriate physical principles and analysis of optimum compromises that have been applied so successfully to the sense organs. The optimization, however, must apply right through into the behavior, because it is the whole flow of excitation to the behavioral output that is acted on by selection.

Function Finding with Identified Neurons

In the mechanistic analysis, we record from identified neurons and hope to show what acts on what. If a deeper explanation of neuron fields is to be made, the function of each identifiable neuron must be found. Moreover, we have to make the bold assumption that because neurons are identifiable components of the nervous system each will in fact have an identifiable function which in some sense stays with that neuron. As indicated by work in maturation of vertebrate nervous systems, in which neuron properties depend on previous experience, this fundamental assumption is going to break down at some level of analysis, at least in more complex systems, but it holds so far for invertebrates.

Functions of Single Neurons

Throughout the animal kingdom, there are numerous examples where the expression of a behavior pattern is linked in a direct causal sequence with the excitation of an identifiable neuron or of a group of neurons that are permanently connected together in a relatively isolatable system. There are numerous examples where a behavior pattern is brought about by a single MN. In these cases, *without a good reason,* one might as well study the behavior as the neuron. One reason is that intracellular records from the MN soma may reveal synaptic potentials which show part of the input to it. Through-conducting nerve nets and epithelial conducting pathways in coelenterates and giant fiber systems of annelids and of some arthropods are the classical examples of the effective evocation of a behavior pattern by a single neuron. In a few examples, a single sensory neuron sets off a complete behavior pattern, in which case the sensitivity of a single receptor acts as the abstraction mechanism that determines when a particular behavior should be triggered. Much more commonly, a class consisting of many hundreds of receptor neurons, with at least some field properties in common, sets off a stereotyped behavior.

A striking example is the ability of the bee to navigate by the plane of polarization if the light is ultraviolet. Work by von Helversen and Edrich (1974) and unpublished work by Menzel have shown that the spectral sensitivity curve of discrimination of the *e*-vector is close to the spectral sensitivity curve of the ninth retinular cell in the ommatidia of the bee eye. The optics of the bee ommatidium are such that this spectral sensitivity is close to the absorption spectrum of the protein rhodopsin, with peak sensitivity near 360 nm. Therefore, we have an aspect of the

behavior that is dominated by a clearly defined property of a protein molecule. But again the *explanation* can go deeper. To show why the short wavelengths are most effective and why the polarization of the sky is useful for navigation, I can trace the story back to Tyndall's lecture of 1869, published in *The Forum* in 1888. Tyndall's first experiment was to show "that by the scattering action of minute particles, the blue of the sky can be produced." He did this by forming fine solid particles by chemical reaction between two gases while illuminating from the side. Tyndall then demonstrated that the plane of polarization of the scattered light is at right angles to the illuminating beam. He examined the polarization of the blue of the sky and correctly identified the plane of polarization as perpendicular to the direction of the sun. Tyndall, like the natural selection acting on the bee, found that the shorter the wavelength the greater the scattering and therefore the greater the polarization, because it is the scattered light which is polarized.

Müller's Law

The examples where the input–output relation of a single neuron is closely related to some aspect of behavior have led to the simplification known as Müller's law (Müller, 1835), which is, in brief, in modern terms, that each neuron, as a unit of the nervous system, generates its own definable effect. Normally, each neuron has one axon which carries one message. Change in impulse frequency of the axon, or depolarization of some part of the neuron, represents only the intensity of the signal, and by implication, therefore, the labeling of the activity is by axon identity.

Two major practical problems are set by this axiom. One is that where two neurons *a* and *b* act on a third *c*, the effect of *a* on *c* is contingent on the activity of *b*, which is not necessarily known by the observer to be present when only *a* and *c* are observed. As nonlinearities such as facilitation and inhibition soon emerge from the simplest interactions, the components and their circuits are impossible to define by inference. The presence of *b* is inferred from the lack of fixed relation between *a* and *c*. A modern interpretation of Müller's law allows for this lack of a fixed relation between input and output as part of the properties that can usually be given to an identifiable neuron. The proviso is that the contingent changes which modify the input–output relations are definable by experiment. From being a hypothesis about residual unexplained effects, the contingent changes become another dimension of the field properties of the identified neuron.

The other problem is that neuron fields are never fully mapped. For all neurons except a few specialized receptors, function finding is an open-ended search for further properties of the field. In this context, Müller's law says that contingent changes are restricted, that fields of neurons do not extend indefinitely as the experimenter dreams up new stimuli, that there is a definable function for each neuron, and that the function can be tied down to the anatomical identification. Analysis of invertebrate identified neurons has provided some of the background for these working hypotheses. Apart from their heuristic success, the only basis I know for accepting these principles is that neuron properties appear to be optimized for particular functions, a state of affairs for which there is only one explanation, namely,

that the definable functions of these identifiable neurons are in fact acted on by natural selection, which maintains the optimization. In short, neurons have definable functions on which natural selection acts. If so, mechanistic teleology is a valid approach in the study of neuron fields.

Analyzing Neuron Activity

More often than not, it is impossible to read the message in the axon, to relate IN properties directly to a behavioral output or to the physics of the sense organ. More important, we do not know the whole situation but only a very small fraction of it. This is not surprising, because, in a complex system with many neurons in parallel, the effects of each neuron are contingent on the activity of others. The question is how to allocate functions to the components when faced with this situation.

By sampling at different points in the system with a defined stimulus, one records activity that originates from what we infer to be a causal chain, including final behavior. One way of analyzing such a situation is to work downstream from the properties of the receptor fields to the properties of low- and then high-order IN fields, and so trace the perturbations and combinations of the stimuli into the output. Although there is still a place for random recording from unidentified neurons and for describing the effects of every possible stimulus, that approach leads straight into the morass of recording every damn thing that presents itself, and frustrated efforts to work out what acts on what. I believe that understanding is generated during the progress of this very necessary descriptive work by assuming that Müller's law (each component is designed for a function) holds, and then by mapping the function of that part of the neuron activity which is clearly participating in behavior and therefore optimized by natural selection. I am aware that this is an extraordinarily tedious process, and especially that it requires repeated insights into the workings of the particular animal studied. What the animal abstracts from its own neuronal activity is very slowly discovered, but without it the recording equipment will churn out vast amounts of neuron activity without the discriminating filter of the higher-order neurons.

At each stage of abstraction, even by the sensory neurons at the front end of the nervous system, there is an enormous amount of detail in the neuron activity (especially if we take cross-correlations between different neurons) compared with the puny information content of the ensuing behavior. The best indication of whether or not a stimulus is appropriate is therefore the activity of the next higher IN along the line, and ultimately the behavioral response, but certainly not the recording gear itself. These objections expressed by Horridge (1968, 1969*a*) are largely avoided by working back from the behavioral output to the MNs and thence via central driving neurons into the sensory processing mechanisms as proposed by Hoyle (1964, 1975*a*), because when working upstream in this direction the outputs that are relevant are already partially identified, but the problem is then that the appropriate combinations of stimuli can never be found. Further discussion will be found in Barlow (1961).

Introducing Optimization, Compromise, and Adaptive Design

The rejection of neuron activity that is not picked up by other neurons and is never represented by any wrinkle in the behavior pattern has one consequence of enormous significance for the study of nervous systems. The aspects that appear in the behavior are those that can be acted on by natural selection and therefore we can expect in them a high degree of *optimization,* which is one of the powerful methods of the exact sciences, particularly in the mathematical treatment of physics. If we deal with quantitative aspects of nervous activity that appear in behavior, we can use methods derived from mathematical optimization in order to understand why things are as they are. I called this approach "mechanistic teleology" because the mechanisms are restricted to interactions that are measurable in terms of efficiency in subsequent behavior and can be called "design." The optimization usually refers to a compromise that can be seen as a result of natural selection pushing in two or more incompatible directions.

Over the past few years I have been studying the structure, optics, physiology, and stimulus-gathering and stimulus-splitting properties of the compound eye of insects and crustaceans. The primary photoreceptors are neurons in that they have axons. They divide up the visual world into a large number of information-carrying channels. In the compound eye, the channels are not confounded; i.e., they each carry a message about one parameter at a time. Because of the limited amount of energy available in the stimulus, there is an unavoidable compromise between the number of receptors and the sensitivity (or rather signal-to-noise ratio) of each. There is a related compromise between the resolution of each receptor and its sensitivity because small fields allow high resolution but reduce sensitivity. There is another compromise in that increase in the number of facets on the compound eye gives a greater number of sampling stations but reduces the sensitivity and resolution of each. A further compromise relates to the fraction of a compound eye that can be devoted to a fovea, where the visual axes are crowded together and in this type of eye are therefore farther apart elsewhere. Other compromises also have their basis in simple physical principles. For example, if all light is caught by a receptor, it is black, and so special provisions must be made for color vision; if receptors with different spectral peaks make color vision possible, then sensitivity and perhaps resolution suffer. New life styles are made possible by being able to make use of very dim light. There is also a compromise between seeing a wide range of object sizes and improving performance by concentrating on one size. It may be true that a generalized eye requires as much resolution and sensitivity as possible, and that photon noise and intensity discriminations must be considered, but an object of a particular size subtended at the eye is best seen when the receptor field subtends the same angle as the object, and there may be specialized eyes with some of the field sizes matched to the expected sizes of objects that are looked for. The lesson is that *matching input fields to function* is an effective way of improving performance and reducing numbers of neurons. The principle applies to all neurons.

Most neuron fields, however, are not so clear-cut as those of the eye. It is common to find a class of neurons all of which respond, but in different proportions

and with overlapping fields in many dimensions, to a group of related stimuli. At present, the insights to read the message by our instruments are lacking: only the next neurons down the line can do that. The same kinds of compromises apply; in particular, there is an optimum field overlap in each dimension depending on the number of neuron responses that are seen simultaneously by the next neurons down the line, and the number of discriminations that have to be made.

For the primary photoreceptor neurons of the compound eye, we found that field properties of a photoreceptor could not be explained except in terms of the ultimate needs of the animal, which is another way of describing the selection pressures acting throughout life. In other examples, however, the principle is pretty obvious. A color receptor will have its spectral sensitivity curve in the right place in the spectrum, which implies a match with the wavelengths it is adapted to see. Chemoreceptors are usually designed to respond to a molecule of significance in behavior, mechanoreceptors to a particular loading, auditory organs to threshold levels of a particular carrier frequency, and so on. The teleological nature of this matching of input field with expected stimulus pattern has not previously been stressed. The exponents and the critics of reductionist analysis appear to have missed the point that knowing the function is part of the analysis. My own conviction is that the design of the nervous system, in the sense that engineers use the word "design," will not be comprehensible until we match the properties of the input field with the functions of the output for each neuron. This adds another major obstacle to those which already complicate the study of function finding for neurons.

There are some who will realize that I am trying to go further than the enormous past and current efforts to map sensory fields, or find the optimum stimulus for any named neuron which may be sensory, inter-, or motor. That is merely describing what is there, and such work fills the literature. Most of those recordable properties of neurons are not seen at a later stage down the line, are not of interest to the animal, and have not been honed down by natural selection. The work is therefore much simplified when we concentrate on those aspects of the fields that show through into a behavior pattern. The need is for biologically trained neurobiologists to replace the incentive to *describe* by an incentive to *explain* the design. One way to do this is to adopt the attitudes of the few imaginative engineers or informal physicists (e.g., Glegg, 1969; Feynman *et al.*, 1971), and to move from the anecdotal toward the principles of optimization.

Return to Compromises in Design

Consideration of the physical factors affecting field sizes and their overlap, and hence the number of receptors in the compound eye, showed us, in our research, exactly how the compromises relate to function. It is easy to define adaptation for one function, but given numerous functions it is very difficult, even knowing the final structure, to separate the different priorities and kinds of weighting that have been acted on by natural selection. It is well-nigh impossible if the functions and the

structure have not all been specified with complete clarity. That is why explaining why things are as they are is a test of the completeness of the descriptive analysis. A curious fact about the consideration of adaptation is that approximate measurements are often adequate for the explanation but they must be the right measurements. The problem is that the functions have to be discerned before the proper descriptive measurements are made. One can only go so far without individual naming of neurons. Often one gets more insights by introducing a wide range of comparative data into the mechanistic analysis, as in the study of the fields of the compound eye. Furthermore, design considerations always apply to variants on a definite model, which is equivalent to a drawingboard specification, or a working model, and never to a statistical statement about populations of neurons or of animals. This is one reason why studies on vertebrates lag behind those on invertebrates, and identified neurons place the seminal ideas before us.

Laws of Physics vs. History

It is often said that at the molecular level the laws of physics and chemistry are the only ones that play a role. At the next level, as natural selection acts on the complex products of protein molecules, the fact of adaptation becomes the acceptable way of explaining purposive design. At the higher levels, the physicochemical laws can no longer explain most of the biological phenomena, which are too far removed from physics. They still invite explanation.

"Explain biological phenomena" here means "explaining the details in terms of function" and can sometimes be sharpened down to something like "explain why these particular features are the optimum (selected) ones in terms of what the structure does in the life of the animal." This latter kind of explanation is what the physicist seeks when he looks at components in living animals, but it can be presented only when the component's function is known. Similarly, the physics and chemistry of a particular sense organ can be understood to be as they are only when it is known when the next neuron down the line, and ultimately the animal, responds to that sensory neuron. The same holds even for the receptor molecules which react with the stimulus, because they have *been selected for their effects*. Similarly, the field of every neuron is selected because having that field the neuron works in the system as a whole. Therefore, for every sensory cell and neuron we must take those effects into account. With Edelman (1974), "My purpose is not epistemological, nor shall I attempt an analysis of the reductionist or holist *position,* for I believe that in their most extreme forms there is a fruitless opposition implied by these terms." My purpose is not to attack current neurobiology but to make teleology respectable and try to make descriptions of neuron actions more relevant to understanding why neurons are as they are.

Relevance and Purpose: A Red Herring

A cybernetic revolution swept through biology in the postwar period as a result of analogies with control systems, and discoveries of real control systems as com-

ponents of animals. As the output of a feedback loop returns to be added to the input, the input–output of any component can be related to the behavior of the whole animal only when the loop circuit has been worked out. There can be an impression of purpose when a control mechanism compensates for perturbations that are artificially introduced into the system. Feedback loops, however, are a red herring in this context: first, within the loop each neuron has its own input–output–relevance; second, *the loop as a whole can be regarded as a single component* with a single input–output–relevance.

Bringing Optimization Theory into Neurobiology

An explanation of the size and properties of neuronal fields in terms of the part played by their *outputs* in normal behavior is not just an academic exercise to satisfy man's demand for elegant or relevant explanations. This particular issue becomes prominent where a physical argument is based on a process of optimization. For two decades there has been a question as to whether information theory, which refers to the amount of information carried by a transmission system, applies to the nervous system. Practicing neurobiologists ignore information theory because it is about optimized systems and its main conclusions appear inapplicable. In particular, there appear to be far more channels in nervous systems than an engineer would need for controlling the same behavior, and the rate of information transfer in single neurons appears to be far below optimum at most times. This difficulty is in part overcome when we use optimization theory only for those forms and activities of receptors and neurons which show through in the behavior, and when we try to work out the compromises in the adaptations for particular purposes, given only limited components. There are three aspects of the relevance: (1) Why the input–output function of that neuron has certain properties will be understood only when the chain of behavior, of which it forms one link, is also known. (2) The output of that neuron will mostly be thrown away at the next stage of analysis; only a small part of the information carried by that neuron is carried to the next neurons down the line. (3) Lines must be kept free for rarely encountered options. In brief, information theory can be applied to the real system if we consider only that part of the information *at the input which can be used in the output*. In the past, information theory has been rejected by electrophysiologists because they have related it to all that they can record. However, in dealing with optimum systems we consider only the information in the input which can be used. For this, it is necessary to know which fraction is used.

Every neuron has a *field,* which is represented quantitatively as the pattern of contours of sensitivity, sometimes in many dimensions of the stimulus. Only a part of this field is useful for behavior, therefore subjected to natural selection for optimization, and therefore amenable to information theory, or interesting for explaining nervous system functioning. In recording from neurons, let us adopt an attitude of mind which concentrates on field properties that appear in the output of later neurons and in behavior.

What Is the "Problem of Reduction" in the Nervous System?

The attempt to reduce the phenomena of life to physical interpretation has been progressively more fruitful over the past centuries. There is no attempt here to reverse this trend. I define "mechanistic teleology" as the philosophical approach by which the experimentalist must look at the later events subsequently caused by a living process (in this case, activity of a neuron) as well as the direct mechanical causes of the events themselves. In other terminology the input–output function needs to be extended to the input–output–relevance function. No reasonably well-informed neurobiologist can have any doubts that most of our subject, as presented to the world, falls desperately short of the interest it would contain if the relevance were completely known.

Types of Explanation Available—Summary

Three characteristic forms of explanation are in use in neuroethology.

The first type of explanation, *mechanistic reductionism,* explains behavior as the product of interactions between components of the nervous system. This is the philosophy required to motivate the actual work of analysis. Mechanistic reductionism really only describes what goes on. Except for small localized and restricted regions where numerous small neurons and hormonal or similar effects are excluded, descriptions obtained by mechanistic reductionism are unsatisfactory in that they are always incomplete, and also they do not explain why components and interactions are as they are.

The second type of explanation brings in the *legacy of evolution.* By this approach the components and interactions are as they are because selection has acted on what the ancestors have provided. Experimental analysis takes this for granted.

Although at worst a cloak for ignorance, at best the evolutionary approach reveals the compromises of the past, and the shortfall from perfect performance because evolution is not complete.

The third type of explanation, *mechanistic teleology,* in setting out to explain why components and interactions are as they are, interprets the knowledge of how the system works in terms of the efficiency in performing a function. Mechanistic teleology relates mechanisms to ends on the assumption that adaptive evolution acting over long periods of time has evolved structures and physiological properties to be efficient for particular functions. With this assumption, the principles of physics and chemistry can help to explain why things are as they are, because they meet the quantitative requirements for an optimized physical system. Mechanistic teleology therefore requires a more exact and deeper understanding than any other approach. Explanations of varying degrees of rigor, relating structure to function, have been common in biology for many years, but analysis of the nervous system has hardly yet reached the level of analysis where quantitative mechanistic teleology

is possible. We are now reaching a stage, however, where the fields of receptor neurons can be treated in this way, and a start can be made on central neurons.

When sufficient detail is known of both input field and the output as seen by the next stage down the line, it is profitable and enjoyable to apply principles of optimization and the laws of physics to the design of the system. Through the laws of physics, mechanistic teleology relates the functionally significant part of the input to the functionally relevant part of the output. No other type of explanation provides a comparable understanding of why the system is as it is. This type of explanation is not a figment of the human mind because the optimization as the quantitative expression of the evolutionary process, fits the properties of each neuron to the demands set on behavior by the outside world.