REMARKS ON THE PROBLEM OF THE LOGICAL DESIGN OF THE VERTEBRATE VISUAL SYSTEM
A Study of the Reliability of Large Complex Systems

Richard E. Warren

Massachusetts Institute of Technology
Instrumentation Laboratory
Cambridge, Massachusetts

ABSTRACT

Most interesting machine tasks, e.g., visual pattern recognition, require large complex systems. Conventional machine designs are not well suited to the reliability demands of large complex systems. Many designs, conceived and tested on a small scale, and theoretically extendable to large scale systems, are still awkward and impractical as a large scale system because reliability appears to vary inversely with the number of components. Animal systems seem to have solved this problem and offer some hope for understanding the problem of how to build large complex systems with high reliablity.

In particular animal systems seem to tolerate a great deal of component variation and noise. It is suggested that animal systems represent designs that actually take advantage of noise, unlike conventional machines, and that some of these animal systems may provide techniques which are applicable to the problem of constructing reliable machines from unreliable components.

DESIGN OF AN EYE

1. INTRODUCTION

At present there is an effort underway to study the problems of building an eye. The long term goal is to build an animal. Some of the details of this effort have been described elsewhere(I); the purpose of this paper is to explain why this effort is considered worthwhile and to focus attention on some of the more interesting philosophical issues.

Recent work in frog's vision is used as a basis for forming general principles of organization that may be applicable to broad areas of the nervous system(2). This is not an attempt to develop a full-blown theory of the vertebrate nervous system. It is rather an attempt to capture something of the style of the animal system in a way that may prove helpful in solving engineering problems. Much of that "style[11] is presented in a series of observations about variability in nature and the relation between structure and function in animal systems.

Certainly one of the more interesting and intriguing aspects of anything in nature is the overwhelming variety. It is often said of things in nature that no two individuals are ever exactly alike; no two snowflakes are ever identical, and no two oak trees ever have quite the same configuration of branches. As C. S. Peirce put it, "The endless variety in the world has not been created by law. It is not the nature of uniformity to originate variation, nor of law to beget circumstance. When we gaze upon the multifariousness of nature we are looking straight into the face of living spontaneity"(3). We find the processes of nature inexplicable and sometimes aweinspiring partly because this wealth of individual variation does not fit into the familiar context of clockwork-like mechanisms which have become the paradigm for explanations. It is not immediately obvious how the outward behavior of two natural mechanisms can be so much alike when their individual components are allowed to vary over such a wide range. It is hoped that this study will help provide a context in which we may begin to find answers to some of these problems. The central problem here is to understand and explain processes of nature in terms of mechanisms whose composition is perhaps not capable of exact duplication even in principle.

2. WHY IMITATE ANIMAL SYSTEMS?

In many areas it is becoming increasingly clear that we must somehow learn how to construct systems in which decisions about future events need not involve a total committment before all the facts are in. We certainly cannot send to Mars, for example, a conventionally designed vehicle with all responses preprogrammed and all its hardware committed to specifically preconceived tasks. We simply do not know enough about the environment to know what responses would be appropriate in all cases. The same may be said of many terrestial problems as well. Even an office manager must often order new equipment a year or two in advance without being very clear in detail about what he will be doing in one or two years. Being forced to make a decision prematurely can have disasterous consequences. In many areas then we see that we must somehow master the technique of building systems with uncommitted or partially committted components, which may later be refined or modified as more data become available.

These problems of reliability and self-maintainance, of learning, and the constraints of real-time all represent exactly the kind of problems the animal nervous system must face on a moment by moment basis.

In the past engineers have faced these problems separately. For example, one thinks of a game like chess as the model of a learning situation, in which the principles of self-organization can be studied in isolation. Here one can almost ignore real-time constraints; and one

neednot bother at all with reliability problems. In this way the logical designer can devote his full resources to advancing the art of self-organizing systems, and let hardware people concern themselves with building fast, reliable components. However, when it comes to the practical task of building an integrated system satisfying a number of different requirements, our technology begins to fail us. The integration of different functions, originally treated separately, may introduce interface problems that overburden the system and detract from its effectiveness. A systems designer who has faced this problem of ever growing complexity in conventional hardware systems can appreciate nature's methods of integration in animal systems. This concern with the eye then has this ulterior purpose: It is hoped that a mastery of the principles of an animal visual system will help shed light on, and promote an understanding of, the functional organization of large complex control systems.

3. THE RELIABILITY PROBLEM

There are at least three distinct senses in which components may be described as "unreliable": 1. Each component may be unstable or unpredictable and vary over time; 2. Individual components may be stable, but the manufacturing process may introduce differences among these individuals; 3. All components may be identical and uniformly bad. For the purposes of this study we can ignore this last sense. Although it is possible for components to be unreliable in this third sense, it is unlikely that we would find them easy to come by. This kind of unreliability assumes a kind of perverted mastery of quality control, and it is just that quality control that we find most difficult to achieve.

A realistic solution to the general problem of reliability must solve both the problem of the unstable component and the problem of variation in several components designed for the same task. The past attempts to solve the general problem may have been hindered by a failure to keep both of these aspects of the problem in mind. Most of the conventional reliability measures, like the use of Hamming codes and redundancy bits, are directed at the so-called problem of the "noisy channel". This is just one aspect of the more general problem of the noisy component. Our experience in this area has already made it clear that even if one solves the component problem and succeeds in making fairly reliable components, one is still faced with the problem of the overall reliability of the larger system. All other things being equal, the larger the system, the less reliable it is. The conventional strategy here is to concentrate first on the problem of component reliability and when we run out of ideas, we turn in desperation to ad hoc systems reliability measures which for the most part are awkward and are hard to integrate into the system to which we have already painfully committed ourselves. We have known for some time that this strategy is only marginally successful, but we have not abandoned it because there has been no clear alternative.

In view of this one is tempted to say that the problem should be approached from the other end. Unfortunately, that is easier said than done. It is not immediately obvious what, or where, the "other end" is. In a general way, of course, the suggested strategy here is to solve the systems-reliability problem first, and then design the individual components that turn out to be necessary for the job. Perhaps for large systems at least the hardware and the system design have to go hand in hand. To put it more generally; the reliability requirements of large systems forces an integration of structure and function.

Ordinarily it has not been easy, nor even desirable, to design a machine in which the structure and function are intimately connected. Often for economic reasons, designs have been encouraged which could be implemented without a committment to a particular kind of hardware; in this way the manufacturer could take advantage of hardware improvements without having to modify radically the logical design. This is a reasonable strategy as long as the systems are small and simple. However, it is becoming increasingly clear that for large systems the separation of the hardware reliability problem from the rest of the system pays off in diminishing returns with the increase in system size. In the current literature there is a growing number of Jeremiahs forecasting doom if we do not repent our practice of disintegrated design. For example. Steel and Kircher write in The Crisis We Face.(4)

> To sum up the crisis in automation, we are pursuing a course that leads to a severe overcomplexity. The nature of this complexity arises not from the basic requirements of automatic control, but entirely from our disintegrated approach to invention, development, and production of military and commercial automatic control devices and business computers.

The suspicion that animal systems and perhaps even human social structures might provide an insight into the design of reliable systems with unreliable components has prompted control system designers to take interest in bionics and cybernetics. This should give rise to a reformulation of the reliability problem in more practical terms. Instead of asking how to keep a machine error-free, one asks the weaker question: How to design machines in which errors simply go hand in hand with hardware failures; that is, how to prevent structural losses bringing about a disproportionate loss of function.

The work of Lettvin, Maturana, McCulloch, and Pitts is enormously important in this regard, because it is one of the first serious attempts to explain the function of the frog's retina in terms of the shape and structure of retinal ganglion cells. If it can be shown that structure and function are intimately related, that a specific function depends on a specific shape, then one can begin to see in

a general way at least how it might be that a small deterioration in structure brings about only a proportionally small deterioration of function.

4. THE PROBLEM OF DESIGN

In all vertebrates the primary processing of visual information takes place in the three cellular layers of the retina: i.e., the photoreceptors, the bipolar cells, and the ganglion cells. The optical stalk consists of axons, or output fibers, from the ganglionic layer. The problem of the design of the eye reduces then largely to the problem of assigning a plausible function to each of these three layers of the retina. The key word here is 'plausible'. The design not only must work, it must fit into the context of the nervous system. This leads one initially into a consideration of the kinds of constraints within which a nervous system must operate and the theoretical basis for supposing that some designs are better than others.

The development of the theoretical context in which one may explain and interpret the vertebrate eye as a mechanism will not be a simple achievement; certainly it will not be a simple extension of existing theory. Even a casual look at the anatomy of the vertebrate eye reveals features that cannot be explained in conventional terms. And a closer look gives rise to the suspicion that virtually everything runs contrary to what is now accepted as good engineering practice. The striking thing here is that the layer of photoreceptors, the transducers in the system, is farthest from the light source. The image is erected on the back side of the retina, after having been filtered through literally a maze of blood vessels and cell bodies. There is no evidence that the light interacts chemically with anything in the bipolar or ganglion layers, before it strikes the photoreceptor layer. One can almost imagine a malevolent deity who turned things around and put these layers in front of the photoreceptors just to deteriorate the image, scatter the light, and confound our attempt to understand how it works. Even the lens system is substandard by conventional standards; in the words of Helmholz, "The monochromatic aberrations in the optical system of the eye are not, like the spherical aberration of glass lenses, symmetrical about an axis. They are much more unsymmetrical and of a kind that is not permissible in well constructed optical instruments"(5). Moreover, one glance at the ways neurons are connected to one another - literally a jungle of interconnections - firmly impresses one with the impossibility of classical circuit analysis.

5. DESIGN STRATEGY

One approach to the problem of designing a reliable machine is to face the main issue directly. Noisy components and imperfect quality control are the facts of life. If natural mechanisms take advantage of everything at hand, as they seem to, perhaps they even take advantage of noise. Accordingly it seems to make sense to consider a machine which works well only when the components are in some sense faulty, or stamped out of an imperfect and flexible mold. The philosopher, feeling the frustration of dealing with the practical man, rationalizes: "It takes all kinds to make a world". Here we are talking about an analogous situation in which it may be said that "It takes all kinds to make a reliable machine". The importance of this passing reference to social models can be appreciated more fully when we begin to look at some possible designs. In our development of these designs we have taken the social model seriously, and have more or less consciously tried to picture a mechanism in terms of "societies" of individual computers, voting mechanisms, societies of peers, the judgment of experts, and so on. Social models are instructive because they seem to be examples of mechanisms which somehow rely on individual variation, and which have the same sort of loose coupling between individuals and the same kind of many-to-many connections that one finds in the nervous system.

This approach to machine reliability is of course in marked contrast to the conventional approaches. In fact, it may even be said that we are introducing a new concept of Machine'. The traditional concept of a machine is linked with the notion of a mechanism with precisely describable components; all the gear teeth have to be precisely matched to be able to mesh. There is no room for individuality of parts. The traditional approach to machine manufacture is modeled after the 'clockmaker'. The clockmaker-engineer first lays out his plan on paper; when he makes a part, he knows exactly what he wants. If he fails to, he throws it away, tries again until he succeeds in making something that satisfies the plan exactly. Nature on the other hand rarely rejects any system component, even though many of the components appear to be "stamped out from a very inexact and flexible mold". It is as though nature first made the components and then later looked around to see how to use them. If one accepts the challenge of nature and attempts to design and build a reliable machine from unreliable components, one is doing violence to the traditional concept of a machine, especially if factors like noise and unreliability become explicit components of design and appear, as it were, on the blueprints.

This study is certainly not the first to violate the traditional concept of a machine, although not as many have covered this ground as one might think. Nearly every philosopher since Descartes has considered the possibility of mechanizing thought processes and building automata, but this typically has amounted to a reduction of thought processes to machine processes. In other words, this has been a redefinition of 'thought', rather than a redefinition of 'machine'. The first time the traditional concept of a machine was violated was when someone thought it would be nice if machines could detect and correct their own errors. The use of Hamming codes in machine design has made it possible for components to be "inexact" and noisy. McCulloch and Pitts made a real advance when they described a network of threshold elements which compute the same function under

different thresholds. Reliability techniques like these help modify our concept of a machine because they make it possible to describe a mechanism without specifying the components in exact detail. The exact nature of the micro-structure becomes less important if it can be shown to have little effect on the macro-structure. And this in turn points the way to designs in which the quality control of individual components may be relaxed without compromising the overall performance.

In general, the McCulloch-Pitts nets represent an important step forward because they show that machines are possible in which, in some aspects at least, the function depends upon ordering relations without depending upon a particular metric. That such designs are possible should not be too surprising; one should be able to see this much simply by gazing, with Peirce, at "the multifariousness of nature". If there can be so many detail differences among individuals of a given species, then these differences, and the metrics associated with them, can have little to do with the basic mechanism. Mechanisms which somehow depend on orderings without metrics are not only easier to build, in the sense that the component quality control may be relaxed, but also such mechanisms may be more reliable, in the sense that they are less disturbed by noise, especially if it can be shown that the noise in the system affects only the metrics and not the orderings.

Another important milestone in the breakdown of the traditional concept of a machine is the 30-called "Perceptron"(6). In its simplest forma perceptron consists of a number of random threshold elements tied together in a random network. Inputs and outputs are connected to randomly distributed junctions in the network. The process of "learning" consists in representing patterns of inputs and 'rewarding" the network for issuing the desired output. The "reward" results in some fairly simple internal modifications of thresholds. For example, the threshold of all elements that fired in the case to be rewarded s lowered, or perhaps the threshold of all those that did not fire is increased. Although early expectations that the perceptron could be a practical device have not been realized, it is nevertheless an important theoretical contribution. Since a perceptron is an almost structureless machine, it can be viewed as an answer to the question: What is the most function realizable from the least structure? Although the perceptron is far from a practical device, the fact that it can do anything at all is simply astounding from the point of view of traditional machines. It is important then because it gives one the confidence to face the otherwise unsettling question of the traditionalist: How is function possible at all in the absence of structure.

By exploring the middle ground between highly structured conventional machines and almost structureless perceptrons, we may learn how to take advantage of both. We may learn how to avoid highly structured machines whose complexity produces an undesirable sensitivity to noise and component failure.

6. THE CONCEPT OF LAYERED COMPUTATION

One of the first things that strikes the student of biology is the layered structure of animal tissue; this is especially striking in the "higher", more organized species. Seen through a microscope, almost any tissue from these higher forms is easily resolved into cell layers. This layered structure is apparent even in the nervous system. The one exception is the reticular formation, which is only one layer deep. Here the primary flow of information is in the horizontal dimension.

The picture that emerges is one of many units within a layer, all operating in parallel. The units, or neurons, accept inputs from corresponding units in the previous layer, and issue outputs to the corresponding units in the subsequent layer.

One of the features of this picture is the wealth of possible feedback. While most of the cells in a given layer are designed to pass information in only one direction, a few cells are able to send information back to the previous layers, from which the layer in question gets its inputs. In this scheme each layer could do a small amount of computation, then pass the results on to the next layer and at the same time issue a few feedback control commands to the previous layer to modify thresholds and generally tailor the computation to the demands of the input. In fact, it may just be this feedback potentiality which is the point of the layered structure of the nervous system. This may be the answer to the question: How does the animal control system act in real-time with such slow components.

Talking about the nervous system in terms of layers and computations with possible feedback is a way of emphasizing its role as a control computer and guidance system. In this sense it is to be contrasted with the conventional data-processing computer which is typically a problem solver without real-time constraints. It answers questions like how much is 5 plus 7, or how many biscuits do we have in the warehouse, or what are the odds Maine will vote democratic. The converiiiu lal computer solves peoblem — which can be structured in this simple question-and-answer fashion. Faced with the decision, the designer of conventional data-processing computers will always sacrifice speed for accuracy, and this is why conventional computers make poor control computers.

It is possible to approach the problem of object recognition using the method of conventional data processing; in fact, most of the current object-recognition schemes do just that. In this method, for example, we might scan an area for an "object" (i.e., closed edge) then, having discovered one, inquire after its properties and look for a match in a list of properties to see if this object is of interest. The difficulty with this method is that in any real-world application, edges are rarely closed and the object will move around; by the time one has recognized an object as an object, it may have moved. In fact, in animals it is just this motion that makes an object attractive. This suggests that an object-recognition device

can only recognize moving objects if it can somehow servo its reference axes with the moving object; this is probably part of the mechanism of paying attention. The aspect of motion here introduces real-time constraints which the conventional object-recognition schemes are ill-suited to handle.

The need for rapid feedback capabilities arises in almost any control system requiring responses in real-time. A control system must be able to function within a "tight" feedback loop to achieve accuracy and fine grain control. Conventional data-processing computers achieve fast computation times in a variety of ways; most of these involve large immobile pieces of hardware with large power requirements. Control computers, on the other hand, are often intended for applications where size and mobility are crucial, and so are required to find answers with simpler hardware. Here we find specialization, in which the hardware design and the logical design are closely intertwined with a specific application. Word lengths are kept short to reduce carry propagation times in arithmetic units. In general, the techniques rely on relatively shallow computation. (The depth of computation here can be measured roughly by the number of significant parentheses in the expression of the function to be computed.) Deep computation which may give more accurate results takes more time and slows down response.

It is not unreasonable to suppose that the presence of shallow layers found in the animal nervous system reflects a committment to shallow computation. By reducing a fairly shallow complicated or deep computation to a number of successive shallow computations, each one of which may yield information for feedback, as well as information to be passed on the next "layer" of computation, one begins to see how the layered structure of the nervous system might explain how fantastic response times can be achieved with relatively slow components.

There is some temptation to explain the superiority of the animal's response over conventional computer systems in terms of the number of components available for the task. One is tempted to say that the difference here is simply the difference between serial and parallel operation. Here one might say that conventional systems must operate within the constraints of serial processing, and that is why they are slower. But this is surely a misleading description. It is not in general obvious that all problems, which are now solved in a serial fashion, could be solved faster in a parallel fashion. (It would be something like expecting two ships to cross the Atlantic faster than one.) For one thing, any problem which can be solved in a parallel way must be representable as a function whose terms are commutative; that is, the order in which the terms are computed is not important. For example, one may indeed hasten the process of adding a column of numbers by separating that column into two smaller columns; this makes it possible to add the two columns in parallel, and then finally add the two sub-totals to get the final total. This technique is clearly not possible for all functions. In many computations some terms have to be computed before others. For example, it is well known that in computations involving both multiplication and addition, (e.g., ab±c) the order of computation is important. However, the problem appears in many other areas as well. Many pattern recognition problems involve the recognition of entities which are highly context dependent, as for example in the translation of natural languages. Here it is important to establish the context before deciding on the meaning of particular words"! In this sense some computations are essentially serial. In general there will always be some computations which cannot be reduced to corresponding parallel computations.

The order in which the layers of the nervous system are arranged is undoubtedly related to the essentially serial aspects of the computations that the nervous sytems is designed to perform. Part of our task is to explore this fast-feedback aspect of layered computation as a possible method of achieving fine control.

If it is true that the layered structure of the nervous system has something to do with the distribution of feedback, then one might expect that the kinds of computation performed in a given layer would all be similar, so that information which was fed back to a given layer would be appropriate to anything that might be going on in that layer. It would also satisfy our sense of economy if it turned out that neurons within a given layer were all simple variations on a single theme; it would explain how variations could be rich with a relatively simple genetic code. There is some reason to believe this may be the case.

Lettvin's work on vision in frogs suggests that there are roughly five different functions performed by retinal ganglion cells, and that these correspond roughly to five different kinds of anatomically distinguishable cells. These five types, however, represent a convenient way of characterizing a population that has many intermediate types. It is this problem of the intermediate type that makes simple precise analytic models of each of the types of neurons an almost pointless endeavor. A good model of the frog's retina needs to show something like a family relationship between the various types and to exhibit the ways in which these intermediates are something like a variation on a simple theme.

Four of the functions performed in the frog's retina are normally characterized as edge detection, moving convex edge detection (i.e., bugs), event detection, and dimming; there is some question about the function performed by the fifth kind, because they are so rare that only a few have been studied. On the face of it, it may not seem as though edge detection and event detection are variations on a single theme, or even have anything at all in common. To understand the family relationship among these apparently diverse functions one needs to consider detectors of this sort from the stand point of the logical function they compute. From this point of view one sees that to detect an edge visually one needs to compute a difference in light intensity over some special area. In the simplest case, this is achieved with a Boolean 'exclusive or' element with two

input sin different parts of the visual field. If one input is on and the other off, we know there is a visual edge or gradient between the two inputs. An event on the other hand involves a difference over time rather than space. In this case an exclusive-or element with two inputs coming from the same area of the visual field, but a delay introduced in one of the inputs, would compute change over time. This shows that edge detection and event detection are related in the sense that both involve a difference detector. Now consider what would happen if we made a number of exclusive-or elements and allowed differences of delay in the two inputs. Those elements with large delays in one input would respond to slow changes in the visual field; and those with short delays would detect relatively rapid changes. Suppose further that we allowed the position of the two inputs to vary, so that some pairs of inputs came from widely different areas in the visual field and some came from relatively close areas. Relatively sharp edges (steep gradients in light intensity) would be detected by elements whose inputs were close together, and less distinct edges would be detected by inputs that were farther apart. Elements whose inputs were close together but with slightly different delays would detect either very sharp edges or slight movement. In this way, it becomes clear that one can represent all varieties of edge and motion detection in terms of two-dimensional abstract space whose coordinates range over the special relation between inputs and the difference in delays in inputs.

Elsewhere(I) the author has presented a detailed model of the f r og' s edge and bug dectector s which suggests that edge and moving convex edge detection are also variations on a simple theme, in which the size of the input area deterimines the sensitivity to convex edges. There it is suggested that a convex edge detector is one in which the two inputs to an exclusive or are concentric fields, one inside the other. As the size of the fields is diminished, the section of a given convex edge within the fields becomes less distinguishable from a straight line. The controlling assumption here is that the inferior lens that Helmholz speaks of, and the fact that light has to penetrate two layers of neurons to reach the photoreceptors, transforms a stright line into a kind of uneven, or "wiggly" edge whose small convexities are detectable by a small convex edge detector; so that what we call an edge detector is actually a small convex edge detector. A larger cell has more room for delays between the concentric input areas, and so is more likely to detect motion as well, as in the moving bug detector. This is an example of the way in which structure and function may be intimately related in the animal nervous system.

7. MODELS OF HORIZONTAL ASSOCIATION

In considering designs which tolerate a wide variation of properties in the individual components, it is useful to consider mechanisms which seem to thrive on variation. Human societies, as we have already suggested, are obvious examples of such mechanisms. However, we can find rudimentary examples of similar mechanisms in other areas. In high-fidelity sound reproduction, for example, it is well-known that a number of inexpensive and even poorly-made speakers, connected together in a series-parallel network, will often produce a satisfactory system. Furthermore, it is often pointed out that the system is ever so much better when the speakers come from different manufacturers; the point here, of course, is that we do not want all the speakers to have the same peak in their response curves, say at 100 cps. It is obvious that the wider the range of individual differences, the better the the overall frequency response, although to be sure the transient response and efficiency may suffer without some further refinements.

A slightly more sophisticated use of randomness was suggested by Albert Novikoff(7). Using the techniques of integral geometry normally associated with the "[Buffon needle problem' , Novikoff shows how patterns might be "recognized", that is transformed to a normal form, in a way that is independent of the effects of translation and rotation. Specifically, each pattern is uniquely identified with a certain probability distribution. A pattern is projected onto a field of randomly distributed line segments, or "needles", of randomly varying lengths; for each pattern one may tabulate the points of intersection between the pattern and show how they are distributed among the various lengths of line segment; each unique pattern will have its own characteristic distribution, independent of translation and rotation within the field of line segments. What is particularly interesting about this scheme is that it may provide a way of making sense out of the apparent random jungle of neurons and dendrites that one finds, for example, in the visual cortex of higher vertebrates. This example, and the one above, at any rate show that the concept of a system which is in some way dependent on, or even thrives on, the random individuality of its components is not wholly unheard of.

However, there is no question that the social model provides the most fertile source for inspiration in this area. It is just that we mist be cautious about the way we use the social model. It can show things in a new light, but it can never act as an explanation of a mechanism, because we understand less about social mechanisms than the control systems we are trying to explain and build. This is so mainly because we lack a theory of large partially structured domains.

In our limited and simple-minded experiments with social models, we have fallen onto two processes which we think are of fundamental significance for any device which uses voting mechanisms. These two processes are: 1. the formation of a peer group, and 2. the recognition of experts within that peer group. Stated briefly, the function of the peer group is to link together similar units as part of a voting mechanism and to isolate extremely deviant units, and the function of an expert is to allow units of known reliability to settle differences of opinion or resolve close decisions during the voting

process. This process can be compared to that of forming a crossover network in the series-parallel network of speakers mentioned above to overcome the defects in transient response and inefficiency.

The peer group is the vehicle of the expert's influence. For example, expert lawyers influence only other lawyers, not plumbers or doctors. The expert plays a crucial role in any voting mechanism. He helps swing the bias the right way in close elections, because he can influence peers without being influenced by them. Although we state these principles in distinctly social terms, our claim is that that we can describe a simple mechanism in which each of these social terms makes sense in a relevant and non-trivial way.

Consider the device in figure 1. This device consists of four sub-systems: inputs, logic elements, outputs and association elements. The logic elements are active in the sense that they have gain and switching properties. The association elements are inactive in the sense that they are simply conductors; as conductors they have the properties of a resistor or a diode.

Now let us consider a network of these units in which several different Boolean functions are computed. Some will have outputs only when both inputs are active; some will have outputs when either input is active. Initially every element is connected to every other element through chain of association elements whose initial resistance is zero ohms. Every output is a positive voltage. Let the inputs overlap one another so that neighboring elements are presented with roughly the same inputs. If two neighbors fire together, no current flows along their common association element; however, if they do not fire together, a resulting difference in voltage allows a current flow across the association element. Let us suppose that the association elements have this additional property: The resistance goes up a little each time a current flows through the associative element. Now let us present a random pattern of inputs to our network. When two neighboring units fire differently. The connection between them will deteriorate slightly. Only when both have a plus voltage together is there no current flow from one to the other. After a period of time, clearly most of the logic elements which compute a boolean sum will be tightly coupled

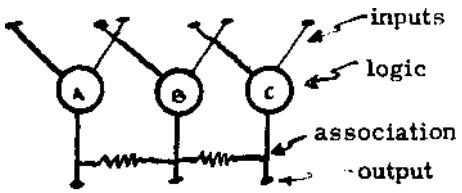
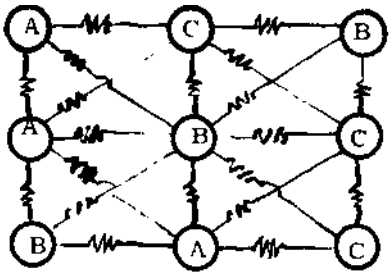
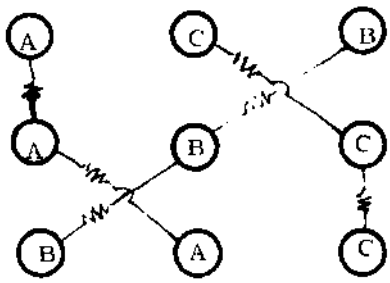


Fig. 1



Before



After

Fig. 2 Top View

together,and loosely coupled, or not at all, to elements which compute other functions. Similarly, in the case of those units that compute a product, only those that fire together will be linked together. This process we call the "formation of peergroup". If there are a few randomly scattered units that always issue an output, and some that never do, these will probably not be linked with anything. Figure 2 represents the conditions before and after this process.

The formation of peergroups in the above sense is really only half the battle. The interconnections between the various members of a peergroup constitute the channel of communication between the members. The whole point of that channel is that the more reliable components can be allowed to "communicate" with, and somehow offset the effects of, the less reliable components. We now have to say something about the way in which the more reliable components can be given additional weight.

It is natural to expect that one's quality control measures will always leave something to be desired. Within each peergroup, some logic units will naturally be better than others. This raises two important problems: 1. What does it mean to be "better", or to be an "expert"? and 2. How does the mechanism recognize one? The first is easier to answer than the second. In fact

I do not think the second can be answered at all generally; there is an implied criterion for truth in any general answer, and there is just no adequate general criterion for truth.

An expert is one who influences his peers, but is not influenced by them. From the standpoint of our model, we call a unit an "expert[11] if the horizontal association elements around it form some non-linear element, like a diode. In terms of our model, this is what it means to be an "expert". The problem of the criteria by which the mechanism is to decide which elements are to become surrounded by diodes is fairly complicated. The specific criteria for "expertise ' will probably vary radically with the specific task. In any case the specific criteria are less important, at this point at least, if we can say something about the manner in which they are applied; and this we can do in a general way.

The first thoughts on how "experts" might be recognized arose out of a consideration of the difficulties encountered in a series of attempts, by various members of the biology department at M.I.T., to make reliable recordings of neurons firing in the retina of various vertebrates. Much of this work is unpublished because the results were negative or inconclusive. It was found that it was virtually impossible to estimate the reliability or accuracy of the experiments because the results were not in general repeatable. It appears that much of the difficulty was due to feedback from other parts of the nervous system and to influences from other systems in the organism which at various times are more or less loosely coupled to the visual system.

This sort of problem is hardly new or unexpected. For theneurophysiologist, "feedback from other areas...", has become one of the facts of life, an occupational hazard to be endured without complaint. Largely because of the work of Hernandez-Peon(8), it is now well known that there are control centers in the brain (i.e., the reticular formation) in vertebrates which can regulate the output rates of afferent, or sensory, neurons. It is estimated that in the frog about 10% of the fibers in the optical stalk are channels for feedback to control the firing rates of ganglion cells in well. It is very likely that this feedback is used to control the amount of information delivered to the brain. It is not surprising that the brain cannot attend to or process all the inputs that are presented to it. At any one time many inputs represent merely background noise and can be ignored without any harm. It naturally occurred to us that the mechanism for recognizing background noise and eliminating it could also be used to recognize faulty neurons. Furthermore, the mechanism whereby the decision is made to attend more closely to an object could also be used to recognize those neurons which seem more reliable indicators of objects worth attending to. The mechanism of attention-control appears to involve feedback commands which can select a small set of neurons whose output is to be enhanced or inhibited. It is as though a central control system could determine which neurons to amplify and which to turn off. If we knew how the animal system did this, we might have a scheme for recognizing the reliable and unreliable components.

One clue was suggested in the following statement, made by Dr. McCulloch explaining some of the difficulties in interpreting the outputs of electrodes implanted in the visual system:

".. . The mouse, which does not turn its eyes and keeps them open, is another nice animal to work on. His retina is the same all over, and whether you get a response from a particular ganglion cell or from a particular axon depends upon whether the mouse is hungry or whether it has smelled its cheese. If it has, then it bothers to look, but it will not look the rest of the time. The mouse shows very little response to any visual stimulus. The situation is far too complicated to be solved with a set of electrodes"(9).

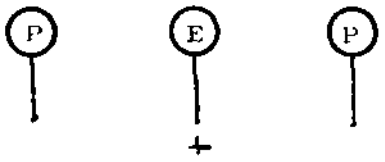
The striking thing here is that one sense modality (e.g., olfaction) may regulate another (e.g., vision). It suggests that neurons reporting the presence of an object have their output rates increased or decreased according to whether or not the object is reported by more than one sense. That is, neurons identified with objects that are both seen and heard, or seen and smelled, tend to be those that are in some more "real", and so there are the ones that have their output rates enhanced.

The requirement that two or more sense modalities agree may be related to the fact that noise on one sensory channel is not likely to have any interesting relation to noise on the other sensory channels. Two sense modalities, considered as communication channels, are not likely to be susceptible to the same kind of noise; things which distort visual inputs probably do not distort auditory ones.

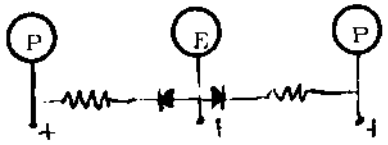
It seems then that in a certain sense we are linking our method of expert-recognition with a kind of "coherence theory of truth . When two different channels agree on the presence of an object, the gain on those neurons agreeing is turned up. If in the act of modifying the gain on those few neurons, a diode (or some similar non-linear device) is formed around those neurons, our mechanism is complete. This is a general account of the method we propose to pursue. It still leaves a lot unanswered. For example, the way in which it is decided whether two sense modalities "agree" is not at all a trivial matter. One can easily see why the criteria for the coherence of inputs has to be handled in terms of the specific inputs. It is one thing to see that the results of two different methods of computation agree, and quite another to see that two different forms of input somehow match. It could be a little like trying to match two person's automobiles and wardrobes on the basis of theories about underlying personalities.

A specific set of criteria for agreement in the case involving the retina will be considered later when we begin to apply some of these principles to the problem of the design of an eye. At this point we will assume that somehow, the mechanism has correctly identified the "experts", or the more reliable components. Figure 3 shows how a peergroup might be expected to improve

the system response with only a small number of "experts". Clearly any voting device, which sums the outputs to determine the majority decision, will be correct more often if it can give this kind of weight to the components which are more likely to be correct.



No horizontal connections:
Only expert fires.



Horizontal connections: All peer group members fire.

Fig. 3

Although we have proposed a model of peergroup functions as a solution to the reliability problem, it is fairLy :L - *tr* "i*t it also could pass as^a theory of learning and self-organization in general. In this regard a comparison with some of the other efforts in this area is instructive. The most publicised effort in the field is the "perceptron" approach.. The perceptron is structurally the simplest of all the learning devices proposed thus far. The unfortunate thing about the perceptron is that it is not immediately obvious that it works. We need a proof to convince us that this process of threshold modification in a randomly connected network actually converges on anything. The present suspicion is that it does not, at least in the interesting cases.

The peer group model proposed here is too structured a mechanism to count as a perceptron, at least in the ordinary sense. While the perceptron is an attempt to answer the question, how much function is possible with how little structure, the peer group mechanism is an attempt to face the problem of how to find a middle ground between a conventional highly structed but unreliable machine and a relatively structureless machine which finesses the quality control problem. It is clear that we know how to achieve good quality control in some things, and it is important to take advantage of that asset when we can. On the other hand, it is clear that sometimes we cannot achieve the quality control that we might like, and here it is important to learn how to take advantage of the other side of the coin as well.

Unlike the perceptron, we do not need a proof to convince us that the peer group process is convergent. (In some sense, it is clearly not convergent since the process passes through, but does not necessarily stop at, the desired point.) Rather we need a demonstration that the learning process does not rapidly deteriorate into an aging process. Modifications in the network are made only by breaking connections, not by making new ones. This breaking of connections is the basic mechanism for dividing groups of similar logic units into peer groups. It is easy to see that after a long period of time the connections between two units, which are very similar (e.g., they may compute the same logical function but have different thresholds), and which should be in the same peer group, finally will be broken. If no two units are ever exactly alike, even two that are very similar will eventually have fired differently enough times so that the association elements between them will have deteriorated. This shows that the process as we have outlined it has an inherent aging problem. After the initial formation of the peer groups, the groups will continue to divide and get smaller. When the groups get small enough to reduce the probability of there being an expert within each group, then it is obvious that the system will begin to fail.

There are several obvious ways in which this aging problem could be overcome, or at least postponed. For example, if new horizontal association elements could be made to grow and replace those that had deteriorated, it is easy to see how peer groups would become more stable. In fact the animal may do just this. If we identify these association elements with horizontal or glial cells in the nervous system, such a regrowth could be explained. In the nervous system, neurons are not regenerated; after birth they only deteriorate. Any theory of learning which is attributable to animal systems must take this into account. However, glial cells are not neurons; they are structural cells which help hold a layer of neurons together. Glial cells are in fact known to be regenerated. In the past it was hard to fit these glial cells into a convincing theory of the nervous system because, as glial cells, they are incapable of performing any of the interesting tasks which we can attribute to neurons. As passive elements, they can have only the properties of materials like copper wire and resistors; they cannot be active elements like transistors.

Although the regeneration of these association elements is an obvious way to improve the system, in this study we have avoided relying on this expedient because we want to exhibit a design which is capable of being built within the framework of current technology. We are attempting to formulate a design whose manufacture consists in dumping a number of micro-elements

made with very poor quality control, to insure variety, into a container, shaking the container to level out the pile into a layer, and then pouring a glue-like material in to fix it. The glue presumably has the properties of our association elements. It seems unlikely that we could come up with a "regenerative" glue, in the required sense, and so we have directed our efforts at other expedients which help stablize peer groups and result in a long and useful life span before they deteriorate.

Some computer studies of the peer group mechanism have been carried out with a small robot designed to learn to solve a simple maze problem. The results indicate that this sort of mechanism, built with inferior quality control by ordinary standards, can actually begin to learn something about an unkown environment by determining which of its many and various coponents best best correllate with one another in that environment. This principle of relaxing quality control and "covering all bets" in an unkown environment sometimes pays off in surprizing ways. What was particularly interesting in these experiments was that many so-called imperfections, such as loose connections, actually turned out to play an important role in the processing of sensory inputs. A loose connection for example was shown often to be a very good wall sensor when it happened that the "noise spike" it generated systematically correlated with other sensory inputs explicitly designed to report impact with obstacles. A full report on these robot experiments is expected to be published in the near future.

NOTES AND REFERENCES

Note 1. 1963 Advanced Sensor Investigations, L. Sutro, R. Warren, etal, MIT Instrumentation Fab report R-436, 1964.

Note 2. The two most important sources are (1) "What the Frog's Eye Tells the Frog's Brain", (Fettvin, Maturana, McCulloch, and Pitts, Proceedings of the IRE, vol. 47, no. 11, November 1959); and (2) "Anatomy and Physiology of Vision in the Frog", Maturana, Fettvin, McCulloch, and Pitts; Journal of General Physiology, July 1960, vol. 43, no. 6, pp 129-175.

Note 3. C.S. Peirce, "Science and Immortality". First appeared in a symposium in the Christian Register, Boston, April 7, 1887.

Note 4 Steel and Kirchner, The Crisis We Face, McGraw-Hill, 1960, *p.21*

Note 5. Helmholtz, Physiological Optics, Optical Society of America, 1924, vol. 1, p. 189.

Note 6. F. Rosenblatt; "Perceptron Simulation Experiments", Proc. IRE vol. 43 , 1960.

Note 7. A. Novikoff; "integral Geometry in Pattern Recognition", Principles of Pattern Recognition, ed. H. von Foerster and G. Zopf, Pergamon Press, New York, 1962.

Note 8. Hernandez-Peon; "Reticular Mechanisms of Sensory Control", Sensory Communications, (ed. RosenblithT^ MIT Press, 1961.

Note 9. McCulloch; "Anastomatic Nets Combating Noise", Information Storage and Neural Control, (Fields and Abbott eds.), Charles C. Thomas (publisher), Springfield, 111., 1963.