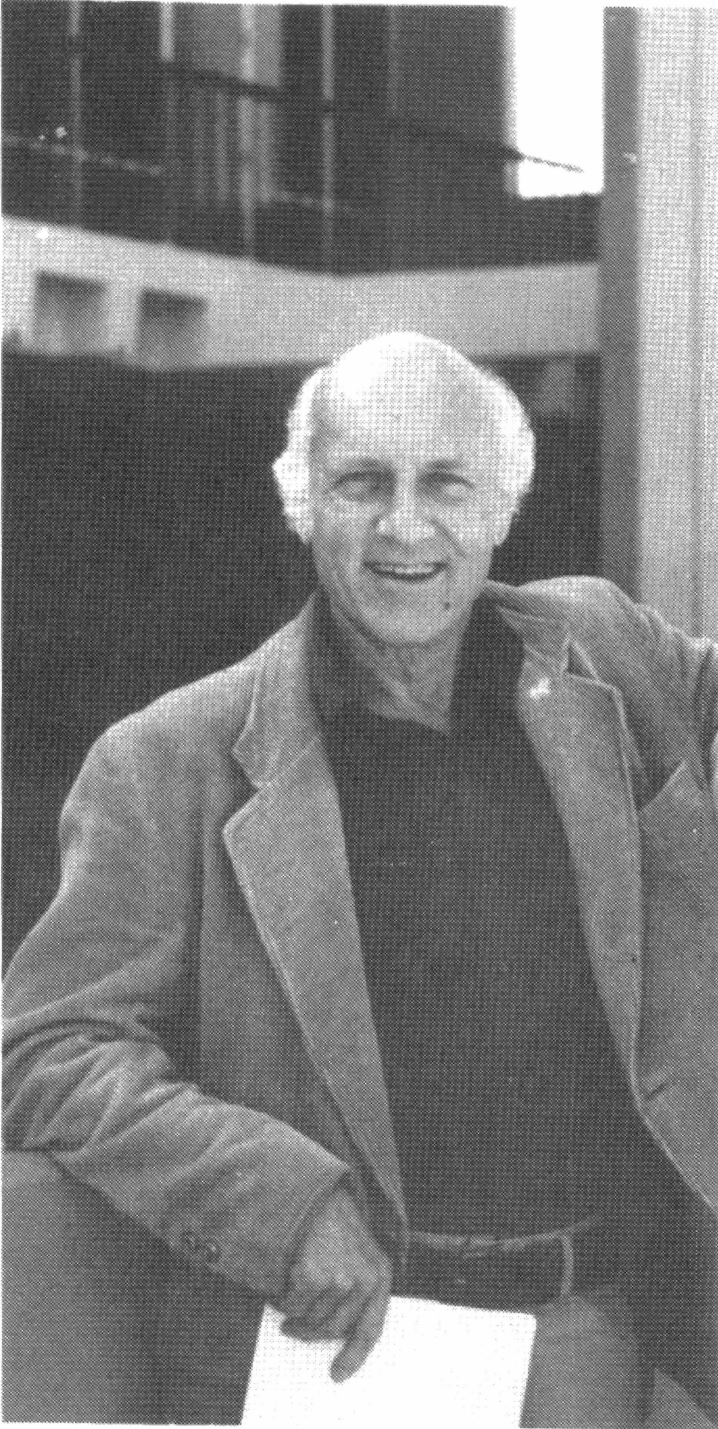HEINZ VON FOERSTER

OBSERVING SYSTEMS

WITH AN INTRODUCTION
BY
FRANCISCO J. VARELA

(Second Edition)

THE SYSTEMS INQUIRY SERIES
Published by Intersystems Publications

HEINZ VON FOERSTER

Gene Youngblood 1977

THE SYSTEMS INQUIRY SERIES

Systems inquiry is grounded in a philosophical base of a systems view of the world. It has formulated theoretical postulates, conceptual images and paradigms, and developed strategies and tools of systems methodology. Systems inquiry is both conclusion oriented (knowledge production) and decision oriented (knowledge utilization). It uses both analytic and synthetic modes of thinking and it enables us to understand and work with ever increasing complexities that surround us and which we are part of.

The Systems Inquiry Series aims to encompass all three domains of systems science: systems philosophy, systems theory and systems technology. Contributions introduced in the Series may focus on any one or combinations of these domains or develop and explain relationships among domains and thus portray the systemic nature of systems inquiry.

Five kinds of presentations are considered in the Series: (1) original work by an author or authors, (2) edited compendium organized around a common theme, (3) edited proceedings of symposia or colloquy, (4) translations from original works, and (5) out of print works of special significance.

Appearing in high quality paperback format, books in the Series will be moderately priced in order to make them accessible to the various publics who have an interest in or are involved in the systems movement.

Series Editors

BELA H. BANATHY and GEORGE KLIR

OBSERVING SYSTEMS

Preface

When Axel Duwe, the man responsible for Intersystems Publications, asked me to write a preface for a second edition of Observing Systems, I was glad to have the chance to answer some questions that have been raised by readers of the previous edition, and to add a few comments.

I was frequently asked what I meant with the title of this book. I intended to leave it to the readers to decide for themselves whether this kaleidoscope of papers is about systems that observe, or about how to observe systems (including systems that observe, for instance, the observant reader).

Then there were the questions about the choice of papers, and about the sequence of their presentation. I could have turned to Francisco Varela who made these choices to provide us with answers, but by going through these articles again I could see a pattern. With the exception of two papers (No. 2 and No. 6) all articles of this collection are based on lectures, addresses, comments, etc., made at various conferences, meetings, or symposia. In hindsight this pattern appears to be obvious; speaking for me is never a monologue, it is always a dialogue: I see myself through the eyes of the other. When I see comprehension, I may have comprehended; when I see puzzlement, I apparently have yet to clarify that point for myself. Moreover, when I speak to a large audience, it is mostly my adrenalin that is speaking; hence, adrenalin seems for me the best ink to write with. Therefore, in this edition occasion and date of presentation is given as a footnote on the titlepage of each paper.

The chosen sequence of papers is (almost) that of the progression dates of presentation. However, as one reader commented, we should have taken more seriously the last phrase of my earlier preface: ". . . the End is supposed to be a Beginning". That is, she suggested reading this collection by beginning with the last paper first, and then going backwards in reverse order, because the full grown organism may contemplate the seed, but not the seed the organism.

On the other hand, with the last paper in this collection, "On Constructing a Reality", I had made common cause with those who prefer to see realities being invented, rather than discovered. In the more than ten years since this paper was given, the number of those who hold this position has grown. Thus I see the notion of an observer-independent "Out There", of "The Reality" fading away very much like other erstwhile notions, "the phologiston", "the imponderable caloric fluid", "the ding-an-sich", "the ether", etc., whose names may be remembered, but whose meanings have been lost.

H.V.F.
Pescadero, California
April, 1984

Preface

Francisco Varela took it upon himself not only to select from my writings the papers for this collection and finding a publisher for it, but also to write an introduction to these articles that were written over a period of twenty-five years.

I am not in the habit of reading my papers, so when I went for the first time through this collection I had difficulties believing that it was I who had supposedly written all that. It was more like being transported in time to those days and nights of questioning, disputing, debating, arguing the issues before us at the Biological Computer Laboratory, a circle of friends, that is, my students, teachers and colleagues.

For me the articles of this collection appear to be frozen instants of an ongoing dialogue, frames of a movie entitled "Observing Systems," whose jerks and jumps backwards and forwards are not to be blamed on the actors, the artists, or the participants, but on the shorcomings of the raporteur, that is me.

When I began writing the credits for this sequence of frames and realized they would always be incomplete, I abandoned this idea in the hope that the references to each article could give a clue as to the main participants in that dialogue. A more complete account is, of course to be found in the Microfiche Collection of the BCL Publications mentioned in Part III.

Whatever the critics may say about this movie, I wish they could see the tremendous joy I had in participating in its creation, and that its End is supposed to be a Beginning.

H.V.F.
Pescadero, California
May, 1982

Acknowledgments

Contents

Introduction

THE AGES OF HEINZ VON FOERSTER

by Francisco J. Varela

Like almost everyone who came into contact with Heinz von Foerster, I owe him a great deal of both learning and support. I say a great deal of learning because Heinz excels at being able to inspire others to think about deep issues which he can point to and express in substantious nutshells. His whole work is marked by this quality, which has a way of staying in one's mind and becoming good food for thought. I also owe him a great deal of support, because one cannot write about Heinz's work without, at the same time, having present his generosity and gentleness, so rare in contemporary academia. As he himself once told me, "To understand something, is to stand under it, so that you may foster its development." And he acted every word of that.

Thus, it is no formality when I say that it is an honour for me to write this Introduction for this collection of essays by Heinz. It is an occasion of celebration for every many of us, and is appropriate, then, that I address myself to the content of this book the reader now holds in his hands.

Considered as a whole, the work of Heinz von Foerster can be taken as a framework for the understanding of cognition. This framework is not so much a fully completed edifice, but rather a clearly shaped space, where the major building lines are established and its access clearly indicated.

Von Foerster's framework has two fundamental principles. First, we are to understand by cognition the described behaviour or a particular class of systems: those which satisfy for their compoenents a specific kind of *internal coherence* (or eigenbehaviour). Second, we are to understand our *own* knowledge as resulting from similar kinds of mechanisms. These two levels are inextricably connected: the study of mechanisms proper of first order systems (those we study), and the study of how second-order systems (those we are) are reflected in such descriptions. This mutually specifying pair, and all its details, constitutes the space where cognition is to be properly understood.

I believe the best way to make this compressed summary into something more understandable and useful for the reader of this book, is to consider some major (st)ages in the development of Heinz's thinking. I shall do so, to return, at the end, with a few remarks about this historical articulation.

* * * * *

The first Heinz stretches from his formative years as a scientist, through his participation in the emergence of cybernetics, until 1958. Intellectual landmarks of this period are his "Quantum mechanical theory of memory," and his editions of the Macy Conferences[1]. I chose the date 1958, because it is the year in which he published "Basic Mechanisms of Homeostasis" which closes this first age.

During this period one key idea has been gradually cultivated and made explicit. To wit: study relations that give rise to processes, independent of their embodiment. In other words: become a cybernetician in its interesting and ample sense of the word.

Lest this sounds too simple to the reader, may I remind him that before this time there was no theoretical domain where we could *ask* questions like, "What is regulation, stability, communication, modelling, . . .?" By casting a horizontal look through disciplines, as it were, cybernetics makes these questions sensical and productive. This is, then, an innovative Heinz who grows from a physicist into a cybernetician by inventing what it is to be one.

At the same time, it is a classic Heinz who searches tools and images for his work in the generalizations from physics, in notions such as entropy, equilibrium, energy exchange, and so on. "I am convinced that in our magnificent task of attempting to unravel the basic laws of incredible complexity of biological systems, these principles will and must be our guides[2]."

A second Heinz grows out from this first age. I chose 1962 as the date for the culmination of this second period, when "On Self-Organizing Systems and Their Environments" was published. These years are dominated by work on the way on which aggregates relate and transform, thus, on self-organization and population dynamics. The cited paper, the volume he edited with G. Zopf, as well as the papers he contributed (with W.R. Ashby and C. Walker) on the connectivity of random networks, are as fresh today as they were when first written.

A key idea of this period is the extension of Shannonian theory to characterize self-organization, in the now well-known principle of order-from-noise, where noise is capable of increasing the redundancy R (e.g. $\partial R/\partial t > 0$). This increase can be understood by noting the many ways in which the components of the system will "select" those perturbations from ambient noise which contribute to increase in order of the system. We move thus from statistical laws (à la Schroedinger) to the consideration of the activity proper to the structure, a theme which, since now, becomes recurrent in von Foerster's work.

In the next age, attention moves on from populations and global properties, to specific cognitive mechanisms. The title which fully announces this transition is telling enough: "A Circuitry of Clues for Platonic Ideation;" it culminates with the comprehensive "Computation in Neural Nets."

During these years, and up until 1970, Heinz will examine in detail the ways in which neural nets are capable of discrimination, learning, and memory. Let me sketch briefly the key idea for each one of these properties.

The core mechanism for discrimination is sought in the necessary connectivities which endow a network (neural or otherwise) with the ability to discriminate edges or sharp transition on its surfaces. This is a powerful perspective to take in the study of sensorial activities, for it takes away the attention from the supposed features of the perturbing agent, and puts it on the structure of the receiving surface. Moreover, such edges can be discriminated not only by sensorial surfaces, but by any neural surface at whichever depth in the brain. This is why von Foerster will devote so much concern to a cascade of neural layers interconnected by means of point-to-point local actions. I believe the advantages of this approach and the formal tools offered, have been surprisingly undeveloped by further research. As one example, let me mention the idea, first proposed by Heinz, that the receptive field of a neural layer receiving excitatory and inhibitory input from a preceding one, could be well described by a difference of two Gaussians

$$f = f_+ - f_- = a_1 e^{-(\frac{x}{\delta_1})^2} - a_2 e^{-(\frac{x}{\delta_2})^2}$$

This local law of interconnection has many powerful features, as has recently been rediscovered[3].

As for learning, the key insight developed was more radical in turn. He argues that learning is not a mapping of some external content, but what the system does in order to transform *its* environment. This idea is particularly well developed in the second part of "Molecular Ethology," (one of my favorite papers). By using the formalism from finite state-transition machines, he shows that in a typical conditioning experiment, for example, "instead of searching for mechanisms in the environment that turn organisms into trivial machines, we have to find the mechanism within the organism that enable them to turn their environment into a trivial machine[4].

With regards to memory, the view taken is better stated, again, in another title of a paper, "Memory without Record." In this, and other places, Heinz develops explicit ways in which a system can exhibit memory, without the need to assume a storage of particular engrams in specific locations. Memory is a mat-

ter of what gets distributed through the change of the system's components.

This period of study of cognitive mechanisms opens very naturally in the next, and current, age of von Foerster. If the previous stage was one concerned with circuits of clues, this one is concerned with the interlocking between cognition and cognizer. This stage opens with the presentation of the Wenner-Green Foundation Conference in 1970 of "Thoughts and Notes on Cognition;" we find again the same themes in the Royaumont meeting in 1972 "Notes on an Epistemology of Living Things." The argument is in two movements. First note that in order for living things to be able to cognize, their organization must be one where mutual reference is mandatory, where the key logic is that of recursive functionals, that is, expression of the form

$$F = \Omega\,(\Omega\,(.\,.\,.(F)\,.\,.\,.)\,)$$

Second, note that such descriptions apply to the describer himself (us), whose own cognition follow also similar recursions and which he can only reveal by acting them out. Thus, the paper concludes: "The environment contains no information. The environment is as it is."

A broad picture of this thinking is presented in a simple form in "On Constructing a Reality." However, my favorite paper of the Heinz of this age, is "Objects: Tokens for Eigen Behaviour," presented for the 80th birthday of Jean Piaget. "Eigenbehavior" is von Foerster's apt term coined to designate the states attained through their recursive mechanisms in cognitive systems. Armed with his vantage point, Heinz offers a delicate rapprochement between his views and Piaget's cycles of perception-action, and development models.

* * * * *

This brings us up to the present on our account of the historical sketch of the development of von Foerster's main ideas, at least as far as the material contained in this book is concerned. I wish to add only a few remarks about the articulation of these ideas.

In this body of work it is immediately obvious we can distinguish two different aspects of it. The first aspect is that which make us younger scientists, realize that what is carved out in this life's work is a lot we tend to take for granted. Of course we should think of relations between components and not tie them to their materiality. Of course self-organization is pervasive in nature. Of course, the nervous system must have mechanism which are both universal and implementable in artificial devices. And so on. Yet we should not allow ourselves

to forget that most of these ideas simply did not exist previous to the wars, and that they came into view because of the tangible work of tangible people. Thus, when we read through the papers and we encounter much that is familiar, it is so because these ideas have spread and permeated the intellectual air we breathe and thus are a testimony to their fertility. In my personal case, it was some of these very papers reprinted here which were instrumental in making me discover the whole world of experimental epistemology, from which I have been never able to stray since. But it is of little consequence whether we found our way into it through Heinz or McCulloch, or Bateson, or von Neuman, or Piaget, or Ashby, or Teuber (or others I am surely leaving out). It is all of one piece, and Heinz's work is both historically and personally linked to it all. His itinerary is then, an itinerary of our own intellectual space.

Yet, there is a second aspect to this body of work. That which has *not* permeated our intellectual preferences and current thinking. That which constitute still a trend or a school, which we do not take simply as ground under us, but as propositions to grapple with. I would say that virtually the whole of the last age of Heinz stands out in this light. There is little doubt that our current models about cognition, the nervous system, and artificial intelligence are severely dominated by the notion that information is represented from an out-there into an in-here, processed, and an output produced. There is still virtually no challenge to the view of objectivity understood as the condition of independence of descriptions, rather than a circle of mutual elucidation. Further, there is little acceptance yet that the key idea to make these points of view scientific programmes is the operational closure of cognizing systems, living or otherwise. These are precisely, the leitmotives of Heinz's last stage.

This book, has then these two complementary and fundamental aspects. On the one hand it is a record of the itinerary of a foundational work. On the other hand, it is a clear statement of a point of view about how are we to understand cognition beyond what we already seem to share as common ground. May there be no mistake in this respect: this is no historical piece, it is present day challenge. Whether we accept its tenets or not, we cannot ignore its standards.

Thus, I am back to my starting point that the work of von Foerster as a whole, amounts to a framework from the understanding of cognition. More than detailed development of detailed empirical mechanism, it is a clear statement of major building lines and staking of grounds. This is a task at which von Foerster is at his best: in presenting a good question clearly, and compressing a deep issue poignantly. Accordingly, I would like to close with a selection of aphorisms by him, which have been formative to me and many others, as scientific koans to bear in mind so that they may give fruits:

★ First order cybernetics: the cybernetics of observed systems;
    Second order cybernetics: the cybernetics of observing systems.

★ The nervous system is organized (or it organizes itself) so as to
    compute a stable reality.

★ There are no primary sensory modalities.

★ Cognition: ⟶ computations of ⟶ ⎤
                              ↑_____⎦

★ Objects: tokens for eigenbehaviour.

★ The logic of the world is the logic of descriptions of the world.

★ Necessity arises from the ability to make infallible deductions;
    Chance arises from the inability to make infallible inductions.

★ Objectivity: the properties of the observer shall not enter in the
    description of his observations.
    Post-objectivity: the description of observations shall reveal the
    properties of the observer.

★ A is better off when B is better off.

★ If you want to see, learn how to act.

NOTES

1. All references for von Foerster's work cited herein can be found in Part III "Publications"
    or reprinted in full in this book.
2. "Basic Mechanisms of Homeostasis," p. 237
3. See for example D. Marr and E. Hildreth, "Theory of Edge Detection,"
    Proc. Roy. Soc. Ser. B, 207: 187, 1980
4. "Molecular Ethology," p. 234

PART I

ON SELF-ORGANIZING SYSTEMS
AND THEIR ENVIRONMENTS*

H. von FOERSTER

Department of Electrical Engineering
University of Illinois, Urbana, Illinois

I AM somewhat hesitant to make the introductory remarks of my presentation, because I am afraid I may hurt the feelings of those who so generously sponsored this conference on self-organizing systems. On the other hand, I believe, I may have a suggestion on how to answer Dr. Weyl's question which he asked in his pertinent and thought-provoking introduction: " What makes a self-organizing system? " Thus, I hope you will forgive me if I open my paper by presenting the following thesis: " There are no such things as self-organizing systems ! "

In the face of the title of this conference I have to give a rather strong proof of this thesis, a task which may not be at all too difficult, if there is not a secret purpose behind this meeting to promote a conspiracy to dispose of the Second Law of Thermodynamics. I shall now prove the non-existence of self-organizing systems by *reductio ad absurdum* of the assumption that there is such a thing as a self-organizing system.

Assume a finite universe, U_0, as small or as large as you wish (see Fig. 1a), which is enclosed in an adiabatic shell which separates this finite universe from any " meta-universe " in which it may be immersed. Assume, furthermore, that in this universe, U_0, there is a closed surface which divides this universe into two mutually exclusive parts: the one part is completely occupied with a self-organizing system S_0, while the other part we may call the environment E_0 of this self-organizing system: S_0 & $E_0 = U_0$.

I may add that it is irrelevant whether we have our self-organizing system inside or outside the closed surface. However, in Fig. 1 the

system is assumed to occupy the interior of the dividing surface. Undoubtedly, if this self-organizing system is permitted to do its
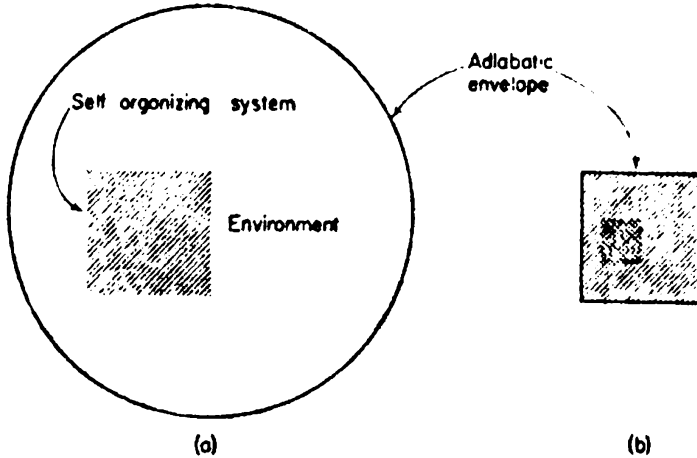
FIG. 1.

job of organizing itself for a little while, its entropy must have decreased during this time:

$$\frac{\delta S_s}{\delta t} < 0,$$

otherwise we would not call it a self-organizing system, but just a mechanical $\delta S_s/\delta t = 0$, or a thermodynamical $\delta S_s/\delta t > 0$ system. In order to accomplish this, the entropy in the remaining part of our finite universe, i.e. the entropy in the environment must have increased

$$\frac{\delta S_E}{\delta t} > 0,$$

otherwise the Second Law of Thermodynamics is violated. If now some of the processes which contributed to the decrease of entropy of the system are irreversible we will find the entropy of the universe U_0 at a higher level than before our system started to organize itself, hence the state of the universe will be more disorganized than before $\delta S_U/\delta t > 0$, in other words, the activity of the system was a disorganizing one, and we may justly call such a system a " disorganizing system."

However, it may be argued that it is unfair to the system to make it responsible for changes in the whole universe and that this apparent inconsistency came about by not only paying attention to the system proper but also including into the consideration the environment of the system. By drawing too large an adiabatic envelope one may include processes not at all relevant to this argument. All right then, let us have the adiabatic envelope coincide with the closed surface which previously separated the system from its environment (Fig. 1b). This step will not only invalidate the above argument, but will also enable me to show that if one assumes that this envelope contains the self-organizing system proper, this system turns out to be not only just a disorganizing system but even a self-disorganizing system.

It is clear from my previous example with the large envelope, that here too—if irreversible processes should occur—the entropy of the system now within the envelope must increase, hence, as time goes on, the system would disorganize itself, although in certain regions the entropy may indeed have decreased. One may now insist that we should have wrapped our envelope just around this region, since it appears to be the proper self-organizing part of our system. But again, I could employ that same argument as before, only to a smaller region, and so we could go on for ever, until our would-be self-organizing system has vanished into the eternal happy hunting grounds of the infinitesimal.

In spite of this suggested proof of the non-existence of self-organizing systems, I propose to continue the use of the term " self-organizing system," whilst being aware of the fact that this term becomes meaningless, unless the system is in close contact with an environment, *which posseses available energy and order*, and with which our system is in a state of perpetual interaction, such that it somehow manages to " live " on the expenses of this environment.

Although I shall not go into the details of the interesting discussion of the energy flow from the environment into the system and out again, I may briefly mention the two different schools of thought associated with this problem, namely, the one which considers energy flow and signal flow as a strongly linked, single-channel affair (i.e. the message carries also the food, or, if you wish, signal and food are synonymous) while the other viewpoint carefully

separates these two, although there exists in this theory a significant interdependence between signal flow and energy availability.

I confess that I do belong to the latter school of thought and I am particularly happy that later in this meeting Mr. Pask, in his paper *The Natural History of Networks*,[2] will make this point of view much clearer than I will ever be able to do.

What interests me particularly at this moment is not so much the energy from the environment which is digested by the system, but its utilization of environmental order. In other words, the question I would like to answer is: " How much order can our system assimilate from its environment, if any at all? "

Before tackling this question, I have to take two more hurdles, both of which represent problems concerned with the environment. Since you have undoubtedly observed that in my philosophy about self-organizing systems the environment of such systems is a *conditio sine qua non* I am first of all obliged to show in which sense we may talk about the existence of such an environment. Second, I have to show that, if there exists such an environment, it must possess structure.

The first problem I am going to eliminate is perhaps one of the oldest philosophical problems with which mankind has had to live. This problem arises when we, men, consider ourselves to be self-organizing systems. We may insist that introspection does not permit us to decide whether the world as we see it is " real," or just a phantasmagory, a dream, an illusion of our fancy. A decision in this dilemma is in so far pertinent to my discussion, since—if the latter alternative should hold true—my original thesis asserting the nonsensicality of the conception of an isolated self-organizing system would pitiably collapse.

I shall now proceed to show the reality of the world as we see it, by *reductio ad absurdum* of the thesis: this world is only in our imagination and the only reality is the imagining " I ".

Thanks to the artistic assistance of Mr. Pask who so beautifully illustrated this and some of my later assertions,* it will be easy for me to develop my argument.

Assume for the moment that I am the successful business man with the bowler hat in Fig. 2, and I insist that I am the sole reality,
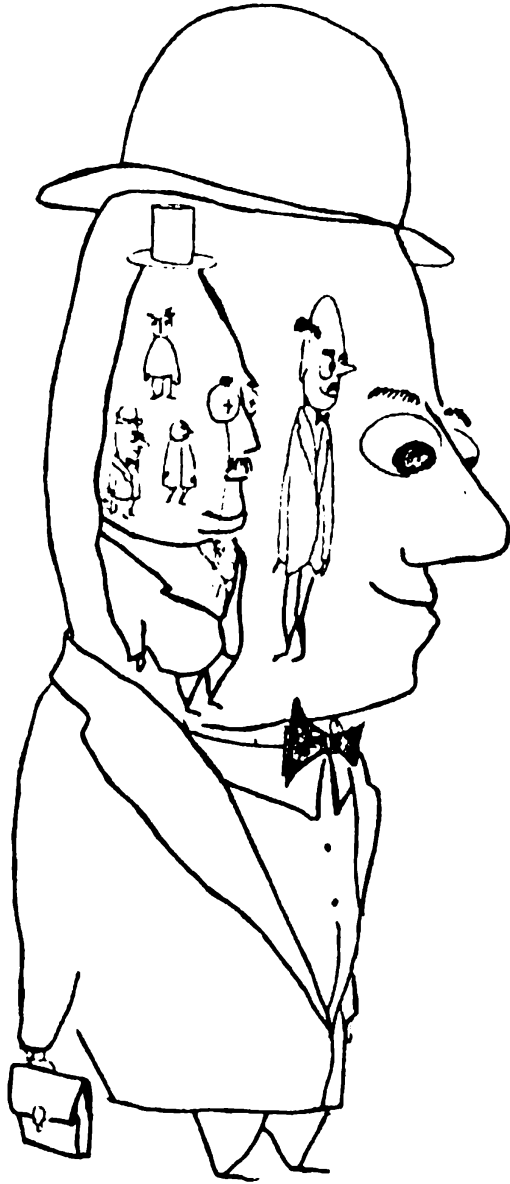
---

*Figures 2, 5 and 6

FIG. 2.

while everything else appears only in my imagination. I cannot deny that in my imagination there will appear people, scientists, other successful businessmen, etc., as for instance in this conference. Since I find these apparitions in many respects similar to myself, I have to grant them the privilege that they themselves may insist that they are the sole reality and everything else is only a concoction of their imagination. On the other hand, they cannot deny that their fantasies will be populated by people—and one of them may be I, with bowler hat and everything!

With this we have closed the circle of our contradiction: If I assume that I am the sole reality, it turns out that I am the imagination of somebody else, who in turn assumes that *he* is the sole reality. Of course, this paradox is easily resolved, by postulating the reality of the world in which we happily thrive.

Having re-established reality, it may be interesting to note that reality appears as a consistent reference frame for at least two observers. This becomes particularly transparent, if it is realized that my " proof " was exactly modeled after the " Principle of Relativity," which roughly states that, if a hypothesis which is applicable to a set of objects holds for one object and it holds for another object, then it holds for both objects simultaneously, the hypothesis is acceptable for all objects of the set. Written in terms of symbolic logic, we have:

$$(Ex) \ [H(a) \ \& \ H(x) \rightarrow H(a + x)] \rightarrow (x) \ H(x) \qquad (1)$$

Copernicus could have used this argument to his advantage, by pointing out that if we insist on a geocentric system, $[H(a)]$, the Venusians, e.g. could insist on a venucentric system $[(Hx)]$. But since we cannot be both, center and epicycloid at the same time $[H(a + x)]$, something must be wrong with a planetocentric system.

However, one should not overlook that the above expression, $\mathscr{R}(H)$ is not a tautology, hence it must be a meaningful statement.* What it does, is to establish a way in which we may talk about the existence of an environment.

---

* This was observed by Wittgenstein,[6] although he applied this consideration to the principle of mathematical induction. However, the close relation between the induction and the relativity principle seems to be quite evident. I would even venture to say that the principle of mathematical induction is the relativity principle in number theory.

Before I can return to my original question of how much order a self-organizing system may assimilate from its environment, I have to show that there is some structure in our environment. This can be done very easily indeed, by pointing out that we are obviously not yet in the dreadful state of Boltzmann's " Heat-Death." Hence, presently still the entropy increases, which means that there must be some order—at least now—otherwise we could not lose it.

Let me briefly summarize the points I have made until now:

(1) By a self-organizing system I mean that part of a system that eats energy and order from its environment.

(2) There is a reality of the environment in a sense suggested by the acceptance of the principle of relativity.

(3) The environment has structure.

Let us now turn to our self-organizing systems. What we expect is that the systems are increasing their internal order. In order to describe this process, first, it would be nice if we would be able to define what we mean by " internal," and second, if we would have some measure of order.

The first problem arises whenever we have to deal with systems which do not come wrapped in a skin. In such cases, it is up to us to define the closed boundary of our system. But this may cause some trouble, because, if we specify a certain region in space as being intuitively the proper place to look for our self-organizing system, it may turn out that this region does not show self-organizing properties at all, and we are forced to make another choice, hoping for more luck this time. It is this kind of difficulty which is encountered, e.g., in connection with the problem of the " localization of functions " in the cerebral cortex.

Of course, we may turn the argument the other way around by saying that we define our boundary at any instant of time as being the envelope of that region in space which shows the desired increase in order. But here we run into some trouble again; because I do not know of any gadget which would indicate whether it is plugged into a self-*dis*organizing or self-organizing region, thus providing us with a sound operational definition.

Another difficulty may arise from the possibility that these self-organizing regions may not only constantly move in space and change in shape, they may appear and disappear spontaneously

here and there, requiring the " ordometer " not only to follow these all-elusive systems, but also to sense the location of their formation.

With this little digression I only wanted to point out that we have to be very cautious in applying the word " inside " in this context, because, even if the position of the observer has been stated, he may have a tough time saying what he sees.

Let us now turn to the other point I mentioned before, namely, trying to find an adequate measure of order. It is my personal feeling that we wish to describe by this term two states of affairs. First, we may wish to account for apparent relationships between elements of a set which would impose some constraints as to the possible arrangements of the elements of this system. As the organization of the system grows, more and more of these relations should become apparent. Second, it seems to me that order has a relative connotation, rather than an absolute one, namely, with respect to the maximum disorder the elements of the set may be able to display. This suggests that it would be convenient if the measure of order would assume values between zero and unity, accounting in the first case for maximum disorder and, in the second case, for maximum order. This eliminates the choice of " neg-entropy " for a measure of order, because neg-entropy always assumes finite values for systems being in complete disorder. However, what Shannon[3] has defined as " redundancy " seems to be tailor-made for describing order as I like to think of it. Using Shannon's definition for redundancy we have:

$$R = 1 - \frac{H}{H_m} \qquad (2)$$

whereby H/H_m is the ratio of the entropy H of an information source to the maximum value, H_m, it could have while still restricted to the same symbols. Shannon calls this ratio the " relative entropy." Clearly, this expression fulfills the requirements for a measure of order as I have listed them before. If the system is in its maximum disorder $H = H_m$, R becomes zero; while, if the elements of the system are arranged such that, given one element, the position of all other elements are determined, the entropy—or the degree of uncertainty—vanishes, and R becomes unity, indicating perfect order.

What we expect from a self-organizing system is, of course, that, given some initial value of order in the system, this order is going to increase as time goes on. With our expression (2) we can at once state the criterion for a system to be self-organizing, namely, that the rate of change of R should be positive:

$$\frac{\delta R}{\delta t} > 0 \qquad (3)$$

Differentiating eq. (2) with respect to time and using the inequality (3) we have:

$$\frac{\delta R}{\delta t} = -\frac{H_m(\delta H/\delta t) - H(\delta H_m/\delta t)}{H_m^2} \qquad (4)$$

Since $H_m^2 > 0$, under all conditions (unless we start out with systems which can only be thought of as being always in perfect order: $H_m = 0$), we find the condition for a system to be self-organizing expressed in terms of entropies:

$$H\frac{\delta H_m}{\delta t} > H_m\frac{\delta H}{\delta t} \qquad (5)$$

In order to see the significance of this equation let me first briefly discuss two special cases, namely those, where in each case one of the two terms H, H_m is assumed to remain constant.

(a) $H_m = $ const.

Let us first consider the case, where H_m, the maximum possible entropy of the system remains constant, because it is the case which is usually visualized when we talk about self-organizing systems. If H_m is supposed to be constant the time derivative of H_m vanishes, and we have from eq. (5):

for $$\frac{\delta H_m}{\delta t} = 0 \ldots \ldots \frac{\delta H}{\delta t} < 0 \qquad (6)$$

This equation simply says that, when time goes on, the entropy of the system should decrease. We knew this already—but now we may ask, how can this be accomplished? Since the entropy of the system is dependent upon the probability distribution of the elements to be found in certain distinguishable states, it is clear that this probability distribution must change such that H is reduced. We

may visualize this, and how this can be accomplished, by paying
attention to the factors which determine the probability distribution.
One of these factors could be that our elements possess certain
properties which would make it more or less likely that an element
is found to be in a certain state. Assume, for instance, the state
under consideration is " to be in a hole of a certain size." The
probability of elements with sizes larger than the hole to be found
in this state is clearly zero. Hence, if the elements are slowly blown
up like little balloons, the probability distribution will constantly
change. Another factor influencing the probability distribution
could be that our elements possess some other properties which
determine the conditional probabilities of an element to be found
in certain states, given the state of other elements in this system.
Again, a change in these conditional probabilities will change the
probability distribution, hence the entropy of the system. Since all
these changes take place internally I'm going to make an " internal
demon " responsible for these changes. He is the one, e.g. being
busy blowing up the little balloons and thus changing the probability
distribution, or shifting conditional probabilities by establishing ties
between elements such that H is going to decrease. Since we have
some familiarity with the task of this demon, I shall leave him for a
moment and turn now to another one, by discussing the second
special case I mentioned before, namely, where H is supposed to
remain constant.

(b) $H = $ const.

If the entropy of the system is supposed to remain constant, its
time derivative will vanish and we will have from eq. (5)

for $$\frac{\delta H}{\delta t} = 0 \ldots \ldots \frac{\delta H_m}{\delta t} > 0 \qquad (7)$$

Thus, we obtain the peculiar result that, according to our previous
definition of order, we may have a self-organizing system before us,
if its possible maximum disorder is increasing. At first glance, it
seems that to achieve this may turn out to be a rather trivial affair,
because one can easily imagine simple processes where this condition
is fulfilled. Take as a simple example a system composed of N
elements which are capable of assuming certain observable states.
In most cases a probability distribution for the number of elements

in these states can be worked out such that H is maximized and an expression for H_m is obtained. Due to the fact that entropy (or, amount of information) is linked with the logarithm of the probabilities, it is not too difficult to show that expressions for H_m usually follow the general form*:

$$H_m = C_1 + C_2 \log_2 N.$$

This suggests immediately a way of increasing H_m, namely, by just increasing the number of elements constituting the system; in other words a system that grows by incorporating new elements will increase its maximum entropy and, since this fulfills the criterion for a system to be self-organizing (eq. 7), we must, by all fairness, recognize this system as a member of the distinguished family of self-organizing systems.

It may be argued that if just adding elements to a system makes this a self-organizing system, pouring sand into a bucket would make the bucket a self-organizing system. Somehow—to put it mildly—this does not seem to comply with our intuitive esteem for members of our distinguished family. And rightly so, because this argument ignores the premise under which this statement was derived, namely, that during the process of adding new elements to the system the entropy H of the system is to be kept constant. In the case of the bucket full of sand, this might be a ticklish task, which may conceivably be accomplished, e.g. by placing the newly admitted particles precisely in the same order with respect to some distinguishable states, say position, direction, etc. as those present at the instant of admission of the newcomers. Clearly, this task of increasing H_m by keeping H constant asks for superhuman skills and thus we may employ another demon whom I shall call the " external demon," and whose business it is to admit to the system only those elements, the state of which complies with the conditions of, at least, constant internal entropy. As you certainly have noticed, this demon is a close relative of Maxwell's demon, only that to-day these fellows don't come as good as they used to come, because before 1927[4] they could watch an arbitrary small hole through which the newcomer had to pass and could test with arbitrary high accuracy his momentum. To-day, however, demons watching

---

* See also Appendix.

closely a given hole would be unable to make a reliable momentum test, and vice versa. They are, alas, restricted by Heisenberg's uncertainty principle.

Having discussed the two special cases where in each case only one demon is at work while the other one is chained, I shall now briefly describe the general situation where both demons are free to move, thus turning to our general eq. (5) which expressed the criterion for a system to be self-organizing in terms of the two entropies H and H_m. For convenience this equation may be repeated here, indicating at the same time the assignments for the two demons D_i and D_e:

$$ H \quad \times \quad \frac{\delta H_m}{\delta t} \quad > \quad H_m \quad \times \quad \frac{\delta H}{\delta t} \qquad (5) $$

Internal
demon's
results

External
demon's
efforts

External
demon's
results

Internal
demon's
efforts

From this equation we can now easily see that, if the two demons are permitted to work together, they will have a disproportionately easier life compared to when they were forced to work alone. First, it is not necessary that D_i is always decreasing the instantaneous entropy H, or D_e is always increasing the maximum possible entropy H_m; it is only necessary that the product of D_i's results with D_e's efforts is larger than the product of D_e's results with D_i's efforts. Second, if either H or H_m is large, D_e or D_i respectively can take it easy, because their efforts will be multiplied by the appropriate factors. This shows, in a relevant way, the interdependence of these demons. Because, if D_i was very busy in building up a large H, D_e can afford to be lazy, because his efforts will be multiplied by D_i's results, and vice versa. On the other hand, if D_e remains lazy too long, D_i will have nothing to build on and his output will diminish, forcing D_e to resume his activity lest the system ceases to be a self-organizing system.

In addition to this entropic coupling of the two demons, there is also an energetic interaction between the two which is caused by the energy requirements of the internal demon who is supposed to accomplish the shifts in the probability distribution of the elements comprising the system. This requires some energy, as we may remember from our previous example, where somebody has to blow up the little balloons. Since this energy has been taken from the environment, it will affect the activities of the external demon who may be confronted with a problem when he attempts to supply the system with choice-entropy he must gather from an energetically depleted environment.

In concluding the brief exposition of my demonology, a simple diagram may illustrate the double linkage between the internal and the external demon which makes them entropically (H) and energetically (E) interdependent.

For anyone who wants to approach this subject from the point of view of a physicist, and who is conditioned to think in terms of thermodynamics and statistical mechanics, it is impossible not to refer to the beautiful little monograph by Erwin Schrodinger *What is Life*.[5] Those of you who are familiar with this book may remember that Schrodinger admires particularly two remarkable features of living organisms. One is the incredible high order of the genes, the " hereditary code-scripts " as he calls them, and the other one is the marvelous stability of these organized units whose delicate structures remain almost untouched despite their exposure to thermal agitation by being immersed—e.g. in the case of mammals —into a thermostat, set to about 310°K.

In the course of his absorbing discussion, Schrodinger draws our attention to two different basic " mechanisms " by which orderly events can be produced: " The statistical mechanism which produces order from disorder and the ... [other] one producing ' order from order '."

While the former mechanism, the " order from disorder " principle is merely referring to " statistical laws " or, as Schrodinger puts it, to " the magnificent order of exact physical law coming forth from atomic and molecular disorder," the latter mechanism, the " order from order " principle is, again in his words: " the real clue to the understanding of life." Already earlier in his book Schrodinger develops this principle very clearly and states: " What

an organism feeds upon is negative entropy." I think my demons would agree with this, and I do too.

However, by reading recently through Schrodinger's booklet I wondered how it could happen that his keen eyes escaped what I would consider a " second clue " to the understanding of life, or—if it is fair to say—of self-organizing systems. Although the principle I have in mind may, at first glance, be mistaken for Schrodinger's " order from disorder " principle, it has in fact nothing in common with it. Hence, in order to stress the difference between the two, I shall call the principle I am going to introduce to you presently the " order from noise " principle. Thus, in my restaurant self-organizing systems do not only feed upon order, they will also find noise on the menu.

Let me briefly explain what I mean by saying that a self-organizing system feeds upon noise by using an almost trivial, but nevertheless amusing example.

Assume I get myself a large sheet of permanent magnetic material which is strongly magnetized perpendicular to the surface, and I cut from this sheet a large number of little squares (Fig. 3a). These
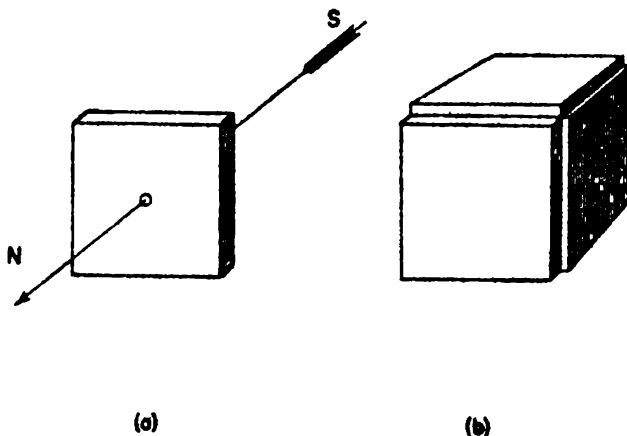

(a)   (b)

FIG. 3. (a) Magnetized square.
(b) Cube, family I.

little squares I glue to all the surfaces of small cubes made of light, unmagnetic material, having the same size as my squares (Fig. 3b). Depending upon the choice of which sides of the cubes have the

magnetic north pole pointing to the outside (Family I), one can produce precisely ten different families of cubes as indicated in Fig. 4.
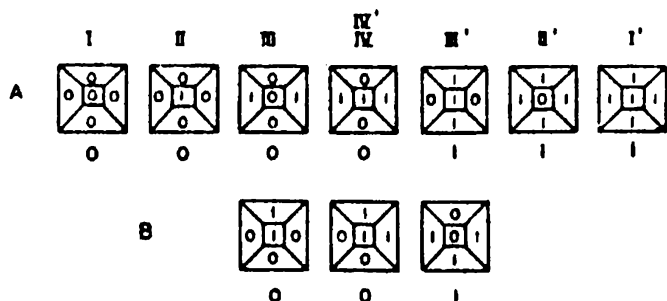

FIG. 4. Ten different families of cubes (see text).

Suppose now I take a large number of cubes, say, of family I, which is characterized by all sides having north poles pointing to the outside (or family I' with all south poles), put them into a large box which is also filled with tiny glass pebbles in order to make these cubes float under friction and start shaking this box. Certainly, nothing very striking is going to happen: since the cubes are all repelling each other, they will tend to distribute themselves in the available space such that none of them will come too close to its fellow-cube. If, by putting the cubes into the box, no particular ordering principle was observed, the entropy of the system will remain constant, or, at worst, increase a small amount.

In order to make this game a little more amusing, suppose now I collect a population of cubes where only half of the elements are again members belonging to family I (or I') while the other half are members of family II (or II') which is characterized by having only one side of different magnetism pointing to the outside. If this population is put into my box and I go on shaking, clearly, those cubes with the single different pole pointing to the outside will tend, with overwhelming probability, to mate with members of the other family, until my cubes have almost all paired up. Since the conditional probabilities of finding a member of family II, given the locus of a member of family I, has very much increased, the entropy of the system has gone down, hence we have more order after the shaking than before. It is easy to show* that in this case

---

* See Appendix.

the amount of order in our system went up from zero to

$$R_{\infty} = \frac{1}{\log_2{(en)}},$$

if one started out with a population density of n cubes per unit volume.

I grant you, that this increase in orderliness is not impressive at all, particularly if the population density is high. All right then, let's take a population made up entirely of members belonging to family IVB, which is characterized by opposite polarity of the two pairs of those three sides which join in two opposite corners. I put these cubes into my box and you shake it. After some time we open the box and, instead of seeing a heap of cubes piled up somewhere in the box (Fig. 5), you may not believe your eyes, but an incredibly ordered structure will emerge, which, I fancy, may pass the grade to be displayed in an exhibition of surrealistic art (Fig. 6).

If I would have left you ignorant with respect to my magnetic-surface trick and you would ask me, what is it that put these cubes into this remarkable order, I would keep a straight face and would answer: The shaking, of course—and some little demons in the box.

With this example, I hope, I have sufficiently illustrated the principle I called "order from noise," because no order was fed to the system, just cheap undirected energy; however, thanks to the little demons in the box, in the long run only those components of the noise were selected which contributed to the increase of order in the system. The occurrence of a mutation e.g. would be a pertinent analogy in the case of gametes being the systems of consideration.

Hence, I would name two mechanisms as important clues to the understanding of self-organizing systems, one we may call the "order from order" principle as Schrodinger suggested, and the other one the "order from noise" principle, both of which require the co-operation of our demons who are created along with the elements of our system, being manifest in some of the intrinsic structural properties of these elements.

I may be accused of having presented an almost trivial case in the attempt to develop my order from noise principle. I agree. However, I am convinced that I would maintain a much stronger position, if I would not have given away my neat little trick with
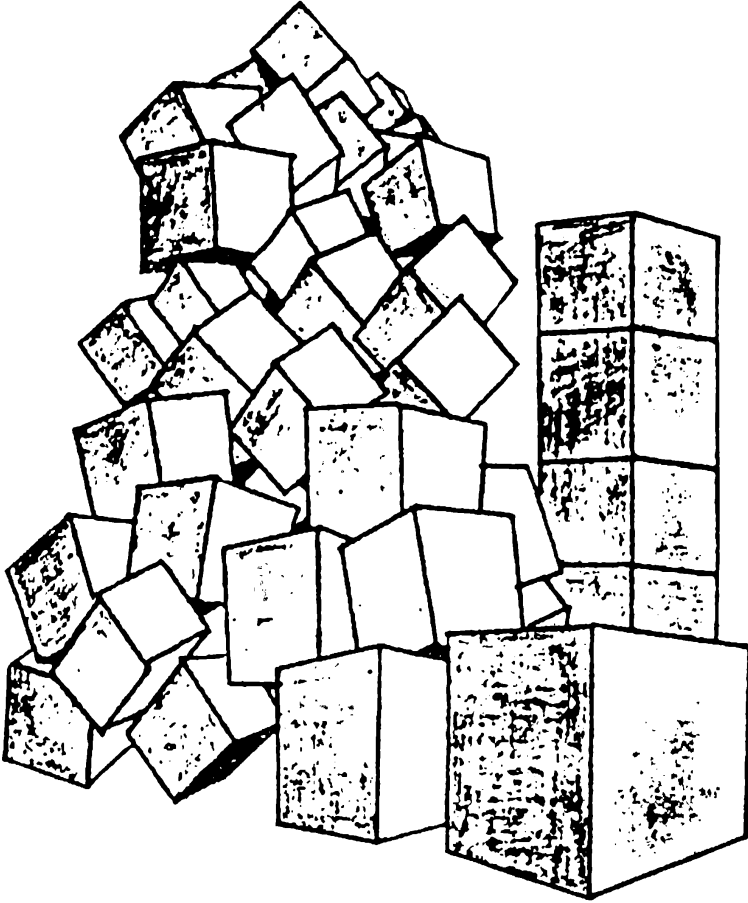
FIG. 5. Before.
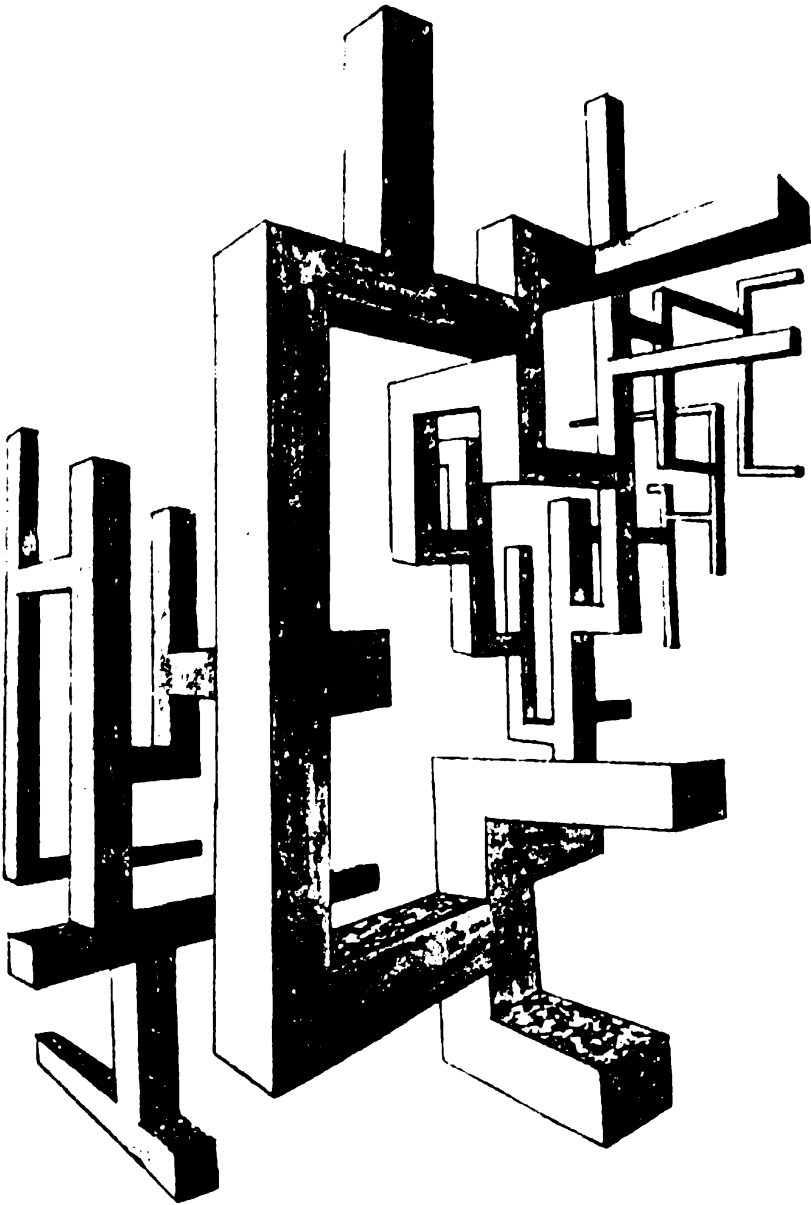
FIG. 6. After.

the magnetized surfaces. Thus, I am very grateful to the sponsors of this conference that they invited Dr. Auerbach[6] who later in this meeting will tell us about his beautiful experiments *in vitro* of the reorganization of cells into predetermined organs after the cells have been completely separated and mixed. If Dr. Auerbach happens to know the trick by which this is accomplished, I hope he does not give it away. Because, if he would remain silent, I could recover my thesis that without having some knowledge of the mechanisms involved, my example was not too trivial after all, and self-organizing systems still remain miraculous things.

APPENDIX

The entropy of a system of given size consisting of N indistinguishable elements will be computed taking only the spatial distribution

(a)

(b)

FIG. 7.

of elements into consideration. We start by subdividing the space into Z cells of equal size and count the number of cells Z_i lodging i elements (see Fig. 7a). Clearly we have

$$\Sigma Z_i = Z \tag{i}$$

$$\Sigma i Z_i = N \tag{ii}$$

The number of distinguishable variations of having a different number of elements in the cells is

$$P = \frac{Z!}{\Pi Z_i!} \tag{iii}$$

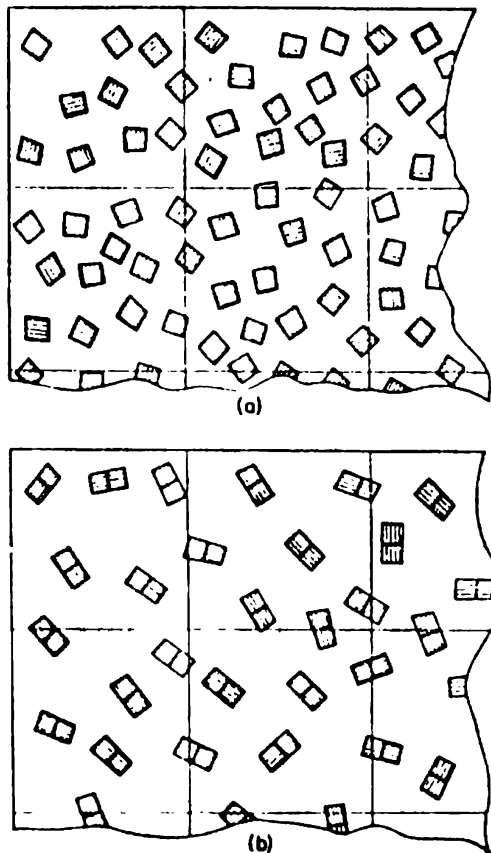whence we obtain the entropy of the system for a large number of cells and elements:

$$H = \ln P = Z \ln Z - \Sigma Z_i \ln Z_i \tag{iv}$$

In the case of maximum entropy H we must have

$$\delta H = 0 \tag{v}$$

observing also the conditions expressed in eqs. (i) and (ii). Applying the method of the Lagrange multipliers we have from (iv) and (v) with (i) and (ii):

$$\Sigma(\ln Z_i + 1)\delta Z_i = 0$$
$$\Sigma i \delta Z_i = 0 \qquad \bigg| \quad \beta$$
$$\Sigma \delta Z_i = 0 \qquad \bigg| \quad -(1 + \ln \alpha)$$

multiplying with the factors indicated and summing up the three equations we note that this sum vanishes if each term vanishes identically. Hence:

$$\ln Z_i + 1 + i\beta - 1 - \ln \alpha = 0 \tag{vi}$$

whence we obtain that distribution which maximizes H:

$$Z_i = \alpha e^{-i\beta} \tag{vii}$$

The two undetermined multipliers α and β can be evaluated from eqs. (i) and (ii):

$$\alpha \Sigma e^{-i\beta} = Z \tag{viii}$$

$$\alpha \Sigma i e^{-i\beta} = N \tag{ix}$$

Remembering that

$$-\frac{\delta}{\delta\beta}\,\Sigma e^{-i\beta} = \Sigma ie^{-i\beta}$$

we obtain from (viii) and (ix) after some manipulation:

$$\alpha = Z(1 - e^{-1/n}) \approx \frac{Z}{n} \qquad\text{(x)}$$

$$\beta = \ln\left(1 + \frac{1}{n}\right) \approx \frac{1}{n} \qquad\text{(xi)}$$

where n, the mean cell population or density N/Z is assumed to be large in order to obtain the simple approximations. In other words, cells are assumed to be large enough to lodge plenty of elements.

Having determined the multipliers α and β, we have arrived at the most probable distribution which, after eq. (vii) now reads:

$$Z_i = \frac{Z}{n}\,e^{-i/n} \qquad\text{(xii)}$$

From eq. (iv) we obtain at once the maximum entropy:

$$\bar{H} = Z\ln(en). \qquad\text{(xiii)}$$

Clearly, if the elements are supposed to be able to fuse into pairs (Fig. 7b), we have

$$\bar{H}' = Z\ln(en/2). \qquad\text{(xiv)}$$

Equating $\bar{H}$ with H_m and $\bar{H}'$ with H, we have for the amount of order after fusion:

$$R = 1 - \frac{Z\ln(en)}{Z\ln(en/2)} = \frac{1}{\log_2(en)} \qquad\text{(xv)}$$

REFERENCES

1. L. WITTGENSTEIN, *Tractatus Logico-Philosophicus*, paragraph 6.31, Humanities Publishing House, New York (1956).
2. G. A. PASK, The natural history of networks. This volume, p. 232.
3. C. SHANNON and W. WEAVER, *The Mathematical Theory of Communication*, p. 25, University of Illinois Press, Urbana, Illinois (1949).
4. W. HEISENBERG, *Z. Phys.* 43, 172 (1927).
5. E. SHRODINGER, *What is Life?* pp. 72, 80, 80, 82, Macmillan, New York (1947).
6. R. AUERBACH, Organization and reorganization of embryonic cells. This volume, p. 101.
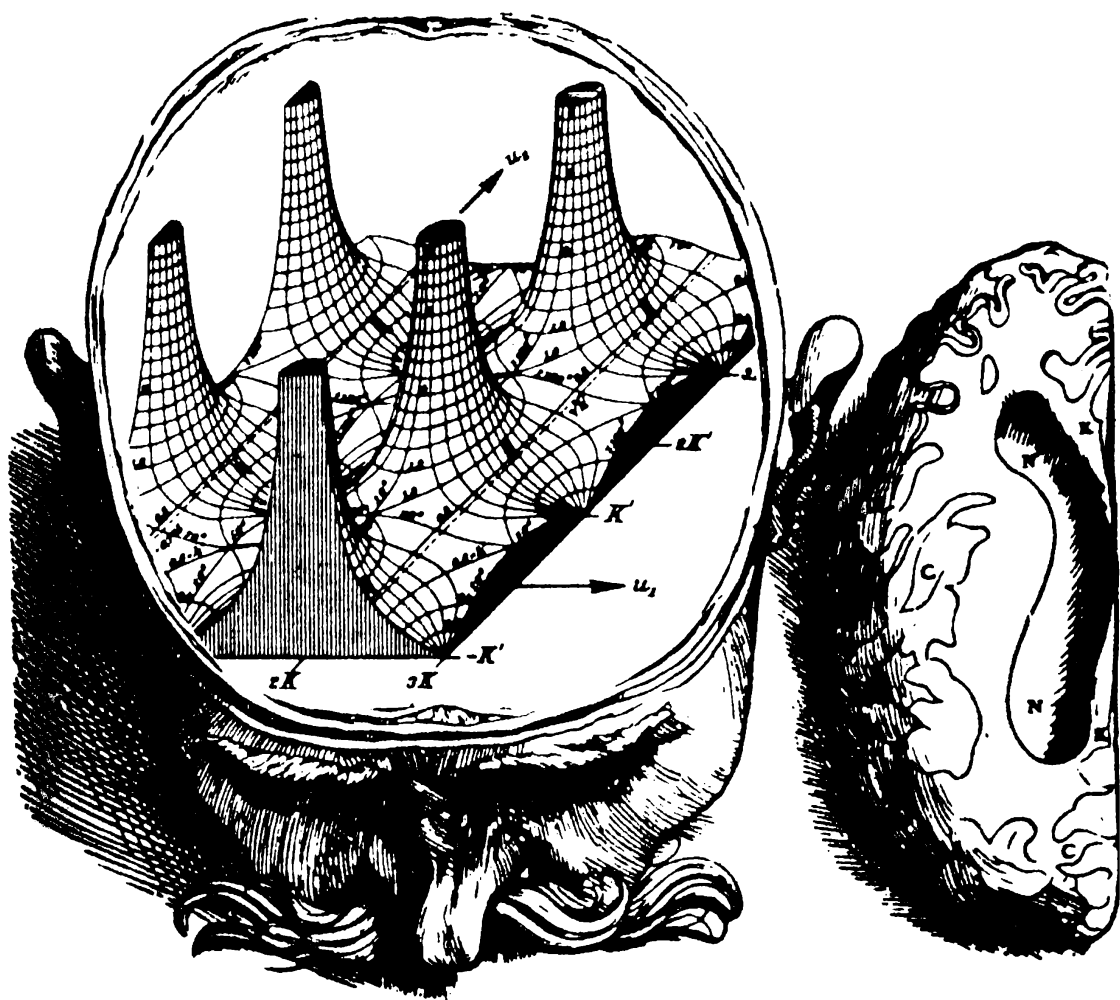
COMPUTATION IN NEURAL NETS

Heinz VON FOERSTER

Department of Electrical Engineering and Department of Biophysics,
University of Illinois, Urbana, Illinois, USA

A mathematical apparatus is developed that deals with networks of elements which are connected to each other by well defined connection rules and which perform well defined operations on their inputs. The output of these elements either is transmitted to other elements in the network or – should they be terminal elements – represents the outcome of the computation of the network. The discussion is confined to such rules of connection between elements and their operational modalities as they appear to have anatomical and physiological counter parts in neural tissue. The great latitude given today in the interpretation of nervous activity with regard to what constitutes the "signal" is accounted for by giving the mathematical apparatus the necessary and sufficient latitude to cope with various interpretations. Special attention is given to a mathematical formulation of structural and functional properties of networks that compute invariants in the distribution of their stimuli.

1. INTRODUCTION

Ten neurons can be interconnected in precisely 1,267,650,500,228,229,401,703,205,376 different ways. This count excludes the various ways in which each particular neuron may react to its afferent stimuli. Considering this fact, it will be appreciated that today we do not yet possess a general theory of neural nets of even modest complexity.

It is clear that any progress in our understanding of functional and structural properties of nerve nets must be based on the introduction of constraints into potentially hyper-astronomical variations of connecting pathways. These constraints may be introduced from a theoretical point of view, for reasons purely esthetic, in order to develop an "elegant" mathematical apparatus that deals with networks in general, or these constraints may be introduced by neurophysiological and neuroanatomical findings which uncover certain functional or structural details in some specific cases. It is tempting, but –alas– dangerous, to translate uncritically some of the theoretical results into physiological language even in cases of some undeniable correspondences between theory and experiment. The crux of this danger lies in the fact that the overall network response (NR) is uniquely determined by the connective structure (C) of the network elements and the transfer function (TF) of these elements, but the converse is not true. In other words, we have the following inference scheme:

$$[C, TF] \rightarrow NR$$

$$[C, NR] \rightarrow Class [TF]$$

$$[TF, NR] \rightarrow Class [C]$$

Since in most cases we have either some idea of structure and function of a particular network, or some evidence about the transfer function of the neurons of a network giving certain responses, we are left either with a whole class of "neurons" or with a whole class of structures that will match the observed responses. Bad though this may sound, it represents a considerable progress in reducing the sheer combinatorial possibilities mentioned before, and it is hoped that the following account of structure and function in nervous nets will at least escape the Scylla of empty generalities and the Charybdis of doubtful specificities.

The discussion of neural networks will be presented in three different chapters. The first chapter introduces some general notions of networks, irrespective of the "agent" that is transmitted over the connective paths from element to element, and irrespective of the operations that are supposedly carried out at the nodal elements of such generalized networks. This generality has the advantage that no commitments have to be made with respect to the adoption of certain functional properties of neurons, nor with respect to certain theories as to the code in which information is passed from neuron to neuron.

Since the overall behavior of neural networks depends to a strong degree on the operations carried out by its constituents, a second chapter discusses various modalities in the operation of these elements which may respond in a variety of ways from extremely non-linear behavior to simple algebraic summation of the input signal strength. Again no claims are made as to how a neuron "really" behaves, for this – alas – has as yet not been determined. However, the attempt is made to include as much of its known properties as will be necessary to discuss some of the prominent features of networks which filter and process the information that is decisive for the survival of the organism.

The last chapter represents a series of exercises in the application of the principles of connection and operation as discussed in the earlier chapters. It is hoped that the applicability of these concepts to various concrete cases may stimulate further investigations in this fascinating complex of problems whose surface we have barely begun to scratch.

2. GENERAL PROPERTIES OF NETWORKS

In this chapter we shall make some preliminary remarks about networks in general, keeping an eye, however, on our specific needs which will arise when dealing with the physiological situation. A generalized network concept that will suit our purposes involves a set of n "elements", $e_1, e_2, \ldots, e_i, \ldots, e_n$, and the set of all ordered pairs $[e_i, e_j]$ that can be formed with these elements. The term "ordered" refers to the distinction we wish to make between a pair, say $[e_1, e_2]$ and $[e_2, e_1]$. In general:

$$[e_i, e_j] \neq [e_j, e_i].$$

This distinction is dictated by our wish to discriminate between the two cases in which an as yet undefined "agent" is transmitted either from e_i to e_j or from e_j to e_i. Furthermore, we want to incorporate the case in which an element transmits this agent to itself. Hence, the pair $[e_i, e_i]$ is also a legitimate pair. Whenever such a transmission takes place between an ordered pair $[e_i, e_j]$ we say that e_i is "actively connected with", or "acts upon", or "influences" element e_j. This may be indicated by an oriented line (arrow) leading from e_i to e_j.

With these preliminaries our generalized network can be defined as a set of n elements e_i ($i = 1 - n$) each ordered pair of which may or may not be actively connected by an oriented line
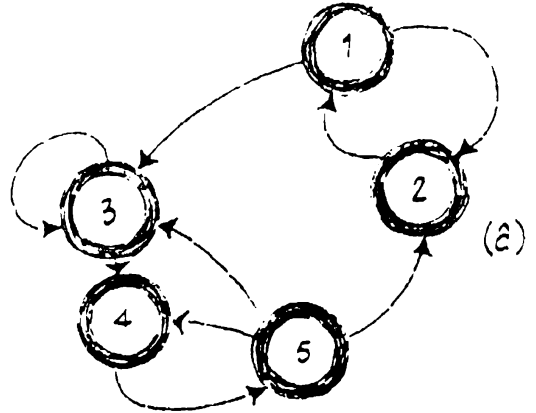
Fig. 1. Network as directed graph (a) and its connection matrix (b).

(arrow). Hence, the connectivity of a set of elements may – in an abstract sense – be represented by a two-valued function $C(e_i, e_j)$ whose arguments are all ordered pairs $[e_i, e_j]$, and whose values are 1 or 0 for actively connected or disconnected ordered pairs respectively.

A simple example of a net consisting of five elements is given in fig. 1a. Here, for instance, the ordered pair $[3, 5]$ appears to be actively disconnected, hence $C(3, 5) = 0$, while the commuted ordered pair $[5, 3]$ shows an active connection path. The function $C(e_i, e_j)$ is, of course, an equivalent representation of any such net structure and may best be represented in the form of a quadratic matrix with n rows and n columns (fig. 1b). The rows carry the names of the transmitting elements e_i and the columns the names of the receiving elements. Active connection is indicated by inserting a "1" into the intersection of a transmitting row with a receiving column, otherwise a "0" is inserted. Hence, the active connection between elements e_5 and e_3.

indicated as an arrow leading from 5 to 3 in fig. 1a, is represented by the matrix element $C_{5,3} = 1$ in row 5 column 3.

This matrix representation permits us at once to draw a variety of conclusions. First, we may obtain an expression for the number of distinguishable networks that can be constructed with n distinguishable elements. Since a connection matrix for n elements has n^2 entries, corresponding to the n^2 ordered pairs, and for each entry there are two choices, namely 0 and 1 for disconnection or active connection respectively, the number of ways in which "zeros" and "ones" can be distributed over n^2 entries is precisely

$$2^{n^2}.$$

For $n = 10$ we have $2^{100} \approx 10^{30}$ different nets and for $n = 100$ we must be prepared to deal with $2^{10000} \approx 10^{3000}$ different nets. To put the reader at ease, we promise not to explore these rich possibilities in an exhaustive manner.

We turn to another property of our connection matrix, which permits us to determine at a glance the "action field" and the "receptor field" of any particular element e_i in the whole network. We define the action field A_i of an element e_i by the set of all elements to which e_i is actively connected. These can be determined at once by going along the row c_i and noting the columns c_j which are designated by a "one". Consequently, the action field of element e_3 in fig. 1 is defined by

$$A_3 = [e_3, e_4].$$

Conversely, we define the receptor field R_i of element e_i by the set of all elements that act upon e_i. These elements can be determined at once by going down column e_i and noting the rows c_j which are designated by a "one". Consequently, the receptor field of element e_3 in fig. 1 is defined by

$$R_3 = [e_1, e_3, e_5].$$

Since the concepts of action field and receptor field will play an important role in the discussion of physiological nerve nets, it may be appropriate to note some special cases. Consider a network of n elements. Under certain circumstances it may be possible to divide these elements into three non-empty classes. One class, N_1, consists of all elements whose receptor field is empty; the second class, N_2, consists of all elements whose action field is empty; and the third class, C, consists of all elements for which neither the action field nor the receptor field is empty. A net for which these three

(a)

(b)
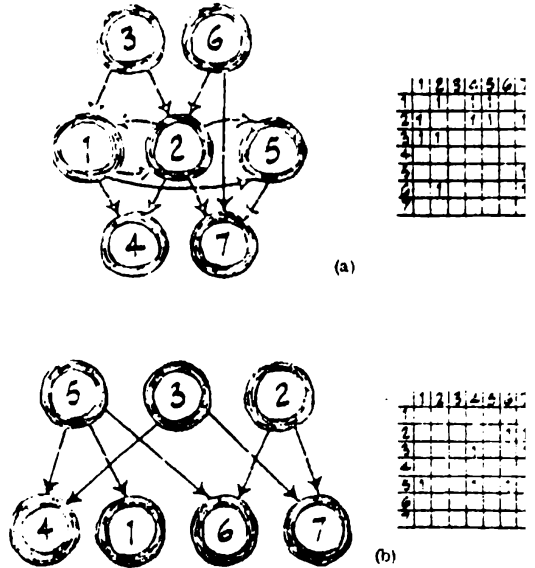
Fig. 2. Hybrid network (a) and action network (b).

classes are non-empty we shall call a "hybrid net". A net which is composed entirely of elements of the third class C we shall call an "interaction net". The net in fig. 1 represents such an interaction net. Finally we define an "action net" which does not possess an element that belongs to class C. Examples of a hybrid net and an action net are given in figs. 2a and 2b with their associated connection matrices.

In the net of fig. 2a:

$$N_1 = [e_3, e_6],$$
$$N_2 = [e_4, e_7],$$
$$C = [e_1, e_2, e_5].$$

In the net of fig. 2b:

$$N_1 = [e_2, e_3, e_5],$$
$$N_2 = [e_1, e_4, e_6, e_7],$$
$$C = [0].$$

For obvious reasons we shall call all elements belonging to class N_1 "generalized receptors" and all elements belonging to class N_2 "generalized effectors". The justification of this terminology may be derived from general usage which refers to a receptor as an element that is

Fig. 3. Replacement of elements by networkers.
Periodic networks.

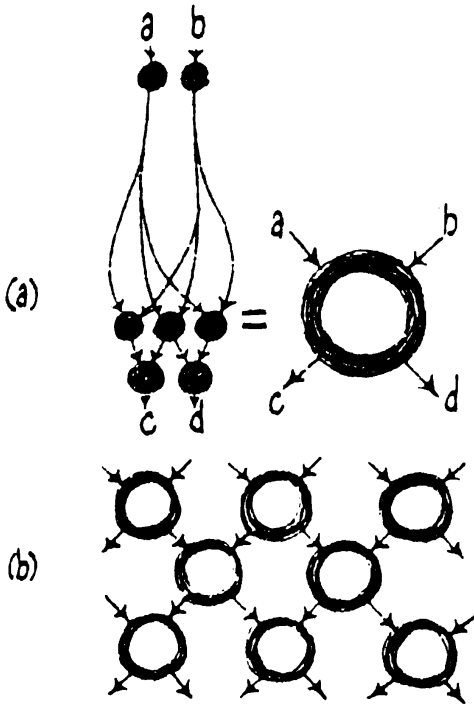Fig. 4. Network without feedback. Action net.

not stimulated by an element of its own network but rather by some outside agent. Likewise, an effector is usually thought of as an element that produces an effect outside of its own network.

This observation permits us to use hybrid networks or action networks as compound elements in networks that show some repetition in their connection scheme. An example is given in fig. 3 in which the net suggested in 3a is to be inserted into the nodes of the net indicated in 3b. The repetition of this process gives rise to the concept of periodic networks, features almost ubiquitous in the physiological situation. To expect such periodicity is not too far-fetched if one realizes for a moment that many net structures are genetically programmed. The maximum amount of information necessary to program a net of n elements is $H_n = n^2$. If this net is made up of k periods of n/k elements each, the maximum information required is only $H_{k,n} = k(n/k)^2 = n^2/k$. Consequently, periodicity — or redundancy — represents genetic economy.

Keeping this point in mind let us investigate

further constraints in the structure of networks.

Consider for the moment an action net consisting of $n = 2m$ elements where the number of elements in set N_1, the generalized receptors, equals the number of elements in N_2, the generalized effectors. In this case the connection matrix has precisely half of its rows and columns empty (0), and the other half filled (1). This makes it possible to re-label all elements, letting the receptors as well as effectors run through labels $1 \rightarrow m$. Consequently, a new matrix can be set up, an "action matrix", which has precisely the same property as our old connection matrix, with the only difference that the elements of the effector set — which define again the columns — are labeled with the same indices as the receptor elements defining the rows. Figs. 4a and 4b illustrate this transformation in a simple example.

The first advantage we may take of an action matrix is its possibility to give us an answer to the question whether or not several action nets in cascade may be replaced by a single action

net; and if yes, what is the structure of this net?

The possibility of transforming a network into the form of an action-matrix has considerable advantages, because an action matrix has the properties of an algebraic square matrix, so the whole machinery of matrix manipulation that has been developed in this branch of mathematics can be applied to our network structures. Of the many possibilities that can be discussed in connection with matrix representation of networks, we shall give two examples to illustrate the power of this method.

In algebra a square matrix A_m of order m is a quadratic array of numbers arranged precisely according the pattern of our connection matrix, or our action matrix. The number found at the intersection of the ith row with the jth column is called element a_{ij}, which gives rise to another symbolism for writing a matrix:

$$A_m = \| a_{ij} \|_m .$$

Addition and subtraction of two matrices of like order is simply carried out by adding or subtracting corresponding elements:

$$A_m \pm B_m = C_m = \| c_{ij} \|_m , \qquad (1)$$

$$c_{ij} = a_{ij} \pm b_{ij} . \qquad (2)$$

It is easy to see that matrix addition or subtraction corresponds to superposition or subposition in our networks. However, we should be prepared to obtain in some of the new entries numbers that are larger than unity or less than zero, if by change the network superimposed over an existent one has between two elements a connection that was there in the first place. Hence, the new entry will show a number "2" which is not permitted according to our old rules of representing connections in matrices (which admitted only "ones" and "zeros" as entries). We may, at our present state of insight, gracefully ignore this peculiarity by insisting that we cannot do more than connect, which gives a "one" in the matrix. Reluctantly, we may therefore adopt the rule that whenever matrix manipulation produces entries $c_{ij} \cdot 1$, we shall substitute "1" and for entries $c_{ij} < 0$, we shall substitute "0". Nevertheless, this tour de force leaves us with an unsatisfactory aftertaste and we may look for another interpretation. Clearly, the numbers c_{ij} that will appear in the entries of the matrix indicate the numbers of parallel paths that connect element e_i with element e_j. In a situation where some "agent" is passed between these elements, this multiplicity of parallel

pathways can easily be interpreted by assuming that a proportionate multiple of this agent is being passed between these elements. The present skeleton of our description of networks does not yet permit us to cope with this situation, simply because our elements are presently only symbolic blobs, indicating the convergence and divergence of lines, but incapable of any operations. However, it is significant that the mere manipulations of the concepts of our skeleton compel us to bestow our "elements" with more vitality than we were willing to grant them originally. We shall return to this point at the end of the chapter; presently, however, we shall adopt the pedestrian solution to the problem of multiple entries as suggested above, namely, by simply chopping all values down to "0" and "1" in accordance with our previous recommendations.

Having eliminated some of the scruples which otherwise may have spoiled unrestricted use of matrix calculus in dealing with our networks, we may now approach a problem that has considerable significance in the physiological case, namely, the treatment of cascades of action networks. By a cascade of two action networks A_m and B_m, symbolically represented by $Cas(AB)_m$ we simply define a network consisting of $3m$ elements, in which all general effectors of A_m are identical with the general receptors of B_m. Fig. 5a gives a simple example. The question arises as to whether or not such a cascade can be represented by an equivalent single action net. "Equivalent" here means that a connecting pathway between a receptor in A and an effector in B should again be represented by a connection, and the same should hold for no connections.

The answer to this question is in the affirmative; the resulting action matrix C_m is the matrix product of A_m and B_m:

$$\| c_{ij} \|_m = \| a_{ij} \|_m \times \| b_{ij} \|_m , \qquad (3)$$

where according to the rules of matrix multiplication the elements c_{ij} are defined by

$$c_{ij} = \sum_{k=1}^{m} a_{ik} b_{kj} . \qquad (4)$$

Fig. 5b shows the transformation of the two cascaded nets into the single action net. Clearly, this process can be repeated over and over again, and we have

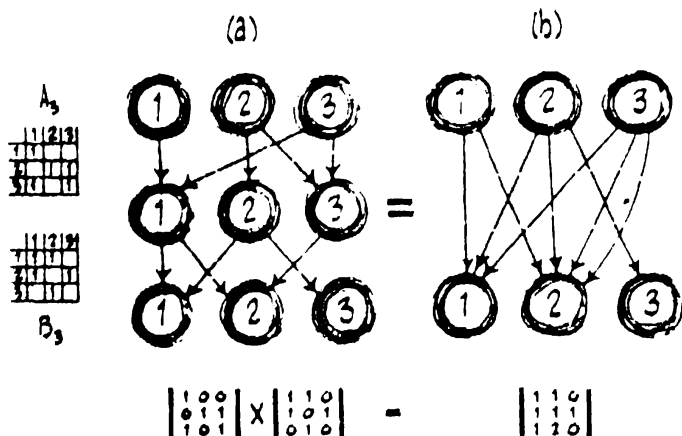$$Cas(A_1 A_2 A_3 A_4 \ldots A_k)_m = \prod_{1}^{k} A_{mi} . \qquad (5)$$

(a)   (b)

Fig 5. Cascade of action networks (a) and equivalent action network (b).

$$\begin{Vmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{Vmatrix} \times \begin{Vmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{Vmatrix} = \begin{Vmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 0 \end{Vmatrix}$$

Here we have one indication of the difficulty of establishing uniquely the receptor field of a particular element, because an observer who is aware of the presence of cascades would maintain that elements e_1 and e_3 in the second layer of fig. 5a constitute the receptor field of element e_2 in layer III, while an observer who is unaware of the intermediate layer (fig. 5b) will argue that elements e_1, e_2 and e_3 in the first layer define the receptor field of this element.

In passing, it may be pointed out that matrix multiplication preserves the multiplicity of pathways as seen in fig. 5, where in the cascaded system element e_2, bottom row, can be reached from e_3, top row, via e_1 as well as via e_3, middle row. All other connections are single-valued.

As a final example, we will apply an interesting result in matrix algebra to cascades of action networks. It can be shown that a square matrix whose rows are all alike

$$a_{ij} = a_{kj} , \tag{6}$$

and each row of which adds up to unity

$$\sum_{j=1}^{m} a_{ij} = 1 \tag{7}$$

generates the same matrix, when multiplied by itself:

$$\|a_{ij}\|_m \times \|a_{ij}\|_m = \|a_{ij}\|_m , \tag{8}$$

or

$$A_m^2 = A_m . \tag{9}$$

Translated into network language, this says that an action network with all receptors connected to the same effectors remains invariant when cascaded an arbitrary number of times. As an example, consider the action matrix

$$A_6 = \begin{Vmatrix} 0 & 1 & 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 & 0 & 1 \end{Vmatrix} = 4 \times \begin{Vmatrix} 0 & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & 0 & \frac{1}{4} \\ 0 & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & 0 & \frac{1}{4} \\ 0 & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & 0 & \frac{1}{4} \\ 0 & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & 0 & \frac{1}{4} \\ 0 & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & 0 & \frac{1}{4} \\ 0 & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & 0 & \frac{1}{4} \end{Vmatrix} .$$

for which the equivalent network is represented in fig. 6. Call p the number of effectors contacted ($p = 4$ in fig. 6), and $N(A_m)$ the normalized matrix whose elements are

$$n_{ij} = \frac{1}{p} a_{ij} . \tag{10}$$

Clearly,

$$\sum_1^m n_{ij} = 1 ,$$
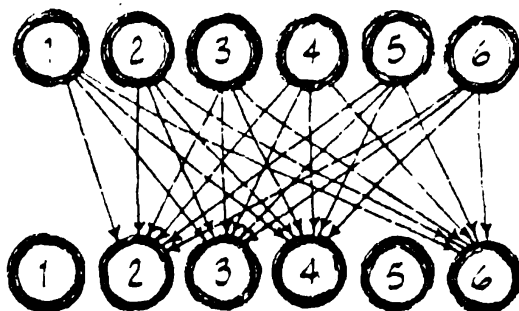

Fig. 6. Stochastic action network.

and k cascades give

$$A_m^k = p^k N(A_m) . \qquad (11)$$

Consequently, the multiplicity of connections will grow with p^{k-1}, while the connection scheme remains invariant.

This observation, which at this level may have the ring of triviality, will later prove to be of considerable utility when we consider variable amounts of an "agent" being passed on from element to element. This again requires a concept of what happens at the site of the elements, a question which leads us, of course, straight into the discussion of "What is a neuron"?

Before we attempt to tackle this quite difficult question — which will be approached in the next chapter — we owe our patient reader an explanation of the term "connection of two elements" for which we offered only a symbolic representation of an "oriented line", along which we occasionally passed a mysterious "agent" without even alluding to concrete entities which may be represented by these abstract concepts. We have reserved this discussion for the end of this chapter because a commitment to a particular interpretation of the term "connection" will immediately force us to make certain assumptions about some properties of our elements, and hence will lead us to the next chapter whose central theme is the discussion of precisely these properties. In our earlier remarks about networks in general we suggested that the statement "element e_i is actively connected to element e_j" may also be interpreted as "e_i acts upon e_j" or "influences e_j". This, of course, presupposes that each of our elements is capable of at least two states, otherwise even the best intentions of "influencing" may end in frustration [*]. Let us denote a particular state of element e_i by $S_i^{(\lambda)}$, where the superscript λ:

$$\lambda = 1, 2, 3, \ldots, s_i .$$

labels all states of element e_i, which is capable of assuming precisely s_i different states. In order to establish that element e_i may indeed have any influence on e_j we have to demand that there is at least one state of e_i that produces a state change in e_j within a prescribed interval of time, say Δt. This may be written symbolically

$$[S_j^{(\lambda)}(t) \rightarrow S_j^{(\lambda')}(t + \Delta t)] = \Phi[S_i^{(\mu)}(t)] . \qquad (12)$$

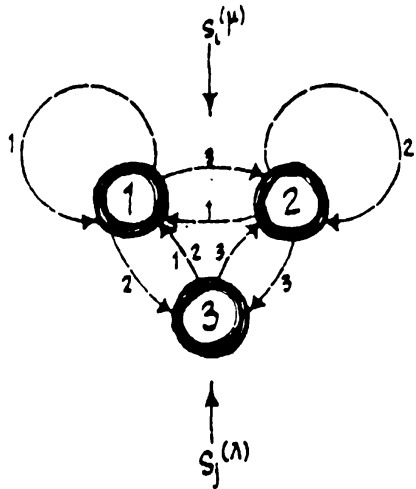[*] An excellent account on finite state systems can be found in Ashby (1956).

Fig. 7. State transition diagram.

In this equation the function Φ relates the states $S_i^{(\mu)}$ in e_i that produce a transition in e_j from $S_j^{(\lambda)}$ to $S_j^{(\lambda')}$. Consequently, Φ can be written in terms of superscripts only:

$$\lambda' = \Phi_{ij}(\lambda, \mu) . \qquad (13)$$

In other words Φ relates the subsequent state of e_j to its present state and to the present state of the acting element e_i. Take, for instance, $s_1 = s_2 = 3$; a hypothetical transition matrix may read as follows:

		μ		
$\lambda' = \Phi(\lambda, \mu)$		1	2	3
	1	1	3	2
λ	2	1	2	3
	3	1	1	2

The associated transition diagram is given in fig. 7, where the nodes represent the states of the reacting element, the arrows the transitions, and the labels on the arrows the states of the acting element which causes the corresponding transition. Inspection of the transition matrix of an element e_i will tell us at once whether or not another element, say e_i, is actively connected to e_j, because if there is not a single state in e_i that produces a state change in e_j we must conclude that e_i does not have any effect on e_j. Again for $s_1 = s_2 = 3$, the "transition" matrix for such an ineffective connection looks as follows

$\lambda' = \Phi(\lambda, \mu)$	μ		
	1	2	3
1	1	1	1
λ 2	2	2	2
3	3	3	3

In general, if we have, for all μ:

$$\Phi_{ij}(\lambda, \mu) = \lambda ,$$

we have in the connection matrix

$$c_{ij} = 0 .$$

If we wish to establish whether or not an element e_i is actively connected to itself, we again set up a transition matrix, only replacing μ by λ. An example of the hypothetical transition matrix of a self-connected element, capable of three states:

$\lambda' = \Phi(\lambda, \lambda)$	λ		
	1	2	3
1	2	-	-
λ 2	-	1	-
3	-	-	3

This element oscillates between states 1 and 2 but stays calmly in 3 when in 3.

The state transition matrix that describes the action of, say, $(k-1)$ elements on some other element is, of course, of k dimensions. In this case the state labels for the ith element may be called λ_i. An example of such a matrix for two elements e_2, e_3, acting on e_1 is given in fig. 8.

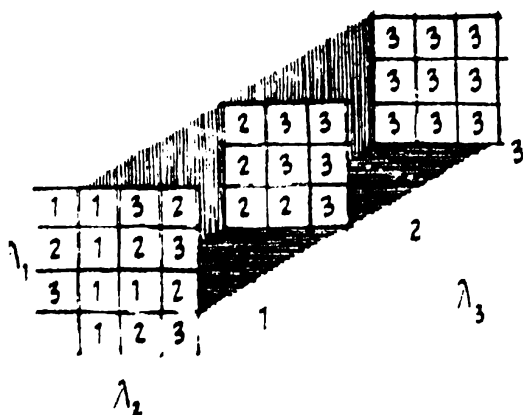If the states of our elements represent some

Fig. 8. State transition matrix for three elements.

physical variable which may undergo continuous changes, for instance, if these states represent the magnitudes of an electrical potential, or of a pressure, or of a pulse frequency, the symbol S_i itself may be taken to represent this magnitude and eq. (12), which described the state transitions of element e_j under the influence of element e_i, assumes now the form of a differential equation

$$\frac{dS_j}{dt} = \Phi_{ij}[S_i(t)] , \qquad (14)$$

which can be solved if the time course of the activity of S_i is known. This, however, may depend on the state of elements acting upon e_i, which in turn may depend on states of elements acting upon those, etc. If we consider such a hierarchy of k levels we may eventually get:

$$\frac{dS_j}{dt} = \Phi_{ij}\{\Psi_{hi}[\Psi_{gh}[\ldots S_0(t - k\Delta t)]]\} ,$$

which is clearly a mess. Nevertheless, there are methods of solving this telescopic set of equations under certain simplifying assumptions, the most popular one being the assumption of linear dependencies.

This brief excursion into the conceptual machinery that permit us to manipulate the various states of individual elements was undertaken solely for the purpose of showing the close interdependence of the concepts of "active connection" and "elements". A crucial role in this analysis is played by the time interval Δt within which we expected some changes to take place in the reacting element as a consequence of some states of the acting element. Clearly, if we enlarge this time interval, say, to $2\Delta t, 3\Delta t, 4\Delta t, \ldots$ we shall catch more and more elements in a network which may eventually contribute to some changes of our element. This observation permits us to define "action neighbours" of the kth order, irrespective of their topographical neighborhood. Hence, in e_j we simply have an action neighbor of kth order for e_i if at least one of the states of e_j at time $t - k\Delta t$ causes a state transition in e_i at time t.

With these remarks about networks in general we are sufficiently prepared to deal with some structural properties of the networks whose operations we wish to discuss in our third chapter. In our outline of the structural skeleton of networks we kept abstract the two concepts "element" and "agent", for which we carefully avoided reference to concrete entities. However, the abstract framework of the interplay of these

concepts permits us to interpret them according to our needs, taking for instance, "general receptors" for receptors proper (e.g., cones, rods, outer hair cells, Meissner's corpuscles, Krause's end-bulbs, Merkel's discs; or for intermediate relays receiving afferent information, bipolar cells, cells of the cochlear nucleus, or for cells in various cortical layers). "General effectors" may be interpreted as effectors proper (e.g., muscle fibers), and also as glia cells, which act in one way on neurons but in another way on each other.

Furthermore, we are free to interpret "agent" in a variety of ways, for instance, as a single volley on a neuron, as a pulse frequency, as a single burst of pulses, as pressure, as light intensity. This freedom is necessary, because in some instances we do not yet know precisely which physical property causes the change of state in some elements, nevertheless, we know which element causes this state change. A commitment to a particular interpretation would favor a particular hypothesis, and would thus mar the general applicability of our concepts.

Our next task is, of course, to give a description of the operational possibilities of our elements in order to put some life into the as yet dead structure of connective pathways.

3. GENERAL PROPERTIES OF NETWORK ELEMENTS

Today our globe is populated by approximately 3×10^9 people, each with his own cherished personality, his experiences and his peculiarities. The human brain is estimated to have approximately 10×10^9 neurons in operation, each with its own structural peculiarities, its scars and its metabolic and neuronal neighborhood. Each neuron, in turn, is made up of approximately 4×10^9 various building blocks — large organic compounds of about 10^6 atoms each — to which we deny individuality, either of ignorance or of necessity. When reducing neurons to a common denominator we may end up with a result that is not unlike Aristotle's reduction of man to a featherless biped. However, since it is possible to set up categories of man, say, *homo politicus*, *homo sapiens* and *homo faber*, categories which do not overlap but do present some human features, it might be possible to set up categories in the operation of neurons which do not overlap but do represent adequately in certain domains the activity of individual neurons. This is the method we shall employ in the following paragraphs.

We shall select some operational modalities as they have been reasonably well established to hold for single neurons under specified conditions, and shall derive from these operational modalities all that may be of significance in the subsequent discussion of neural nets.

Peculiar as it may seem, the neuron is usually associated with two operational principles that are mutually exclusive. One is known as the "All or Nothing" law, which certainly goes back to Bowditch (1871) and which states that a neuron will respond with a single pulse whose amplitude is independent of the strength of stimulus if, and only if, the stimulus equals or exceeds a certain threshold value. Clearly, this description of the behavior of a neuron attaches two states to this basic element, namely, "zero" for producing no pulse, and "one" for producing the pulse. Since modern computer jargon has crept into neurophysiology, this neural property is usually referred to as its "digital" characteristic, for if a record of the activity of a neuron in these terms is made, the record will present itself in form of a binary number whose digits are "ones" and "zeros":

$$\ldots 01100111011110 \ldots$$

When we adopt this operational modality of a neuron as being crucial in its processing of information, we also associate with the string of "ones" and "zeros" the code in which information is transmitted in a network.

The other operational principle, which is diametrically opposed to the one just mentioned, derives its legitimacy from the observation that — at least in sensory fibers — information is coded into the lengths of time intervals between pulses. Since the length of a time interval is a continuous variable, and since under certain conditions this interval may represent monotonically a continuously varying stimulus, this behavior of a neuron is usually referred to as — by again invoking computer jargon — as its "analog" characteristic. Under these conditions we may regard the behavior of a neuron as the transfer function of a more or less linear element whose input and output signal is a pulse interval code and whose function is pulse-interval modulation. A "Weber-Fechner neuron" simply has a logarithmic transfer function, a "Stevens neuron" a power-law response (Stevens, 1957) and a "Sherrington neuron" has neat, almost linear properties with threshold (Sherrington, 1906).

Although it is not at all difficult — as we shall see — to propose a single mechanism that reconciles all types of operation in neurons discussed

so far (analog as well as digital), it is important to separate these operational modalities, because the overall performance of a network may change drastically if its elements move, from one operational modality to another. Consequently, we shall discuss these different modalities under two different headings: first "The Neuron as an 'All or Nothing' Element" with special attention to synchronous and a-synchronous operations, and second "The Neuron as an 'Integrating Element'". After this we shall be prepared to investigate the behavior of networks under various operation conditions of its constituents.

3.1. *The neuron as an "all or nothing" element*

3.1.1. Synchronism

This exposé follows essentially the concepts of a "formal neuron" as proposed by McCulloch and his school (McCulloch and Pitts, 1943; McCulloch, 1962), who define this element in terms of four rules of connection and four rules of operation.

Rules of connection:
A "McCulloch formal neuron":

i) receives N input fibers X_i $(i = 1, 2, N)$, and has precisely one output fiber *.

ii) Each input fiber X_i may branch into n_i facilitatory (+) or inhibitory (-) synaptic junctions, but fibers may not combine with other fibers.

iii) Through the neuron, signals may travel in one direction only.

iv) Associated with this neuron is an integer θ $(-\infty < \theta < +\infty)$, which represents a threshold.

Rules of operation:
v) Each input fiber X_i may be in only one of two states $(x_i = 0, 1)$, being either OFF (0) or ON (1).

vi) The internal state Z of the neuron is defined by

$$Z = \sum n_i x_i - \theta . \qquad (15)$$

vii) The single output Y is two-valued, either ON (1) or OFF (0) $(y = 0, 1)$ and its value is determined by

$$y = \phi(Z) = \begin{cases} 0 \text{ for } Z < -\epsilon, \\ 1 \text{ for } Z > -\epsilon, \end{cases}$$

where ϵ is a positive number smaller than unity:

$$0 < \epsilon < 1 .$$

* The ith fiber will be denoted by a capital X_i, while its state by a lower case x_i.
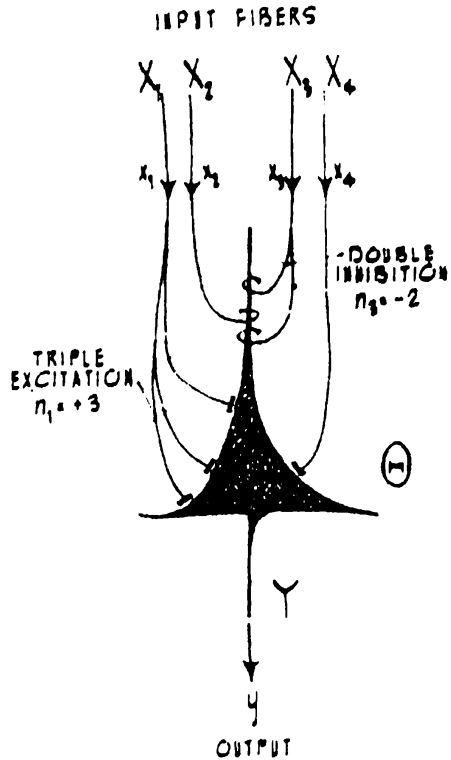

Fig. 9. Symbolic representation of a McCulloch formal neuron.

viii) The neuron requires a time interval of Δt to complete its output.

We shall briefly elucidate these rules with the aid of fig. 9 which is a symbolic representation of this element. The neuron proper is the triangular figure, symbolizing the perikaryon, with a vertical extension upward receiving inhibitory fibers, each loop representing a single inhibitory synaptic junction. Excitatory junctions are symbolized as terminal buttons attached to the perikaryon. In fig. 9 the number of input fibers is:

$$N = 4 ,$$

with the following values of facilitatory and inhibitory synaptic junctions:

$$n_1 = +3 , \qquad n_2 = -1 ,$$
$$n_3 = -2 , \qquad n_4 = +1 .$$

Consider for the moment all input fibers in the

ON state and the threshold at zero:

$$x_1 = x_2 = x_3 = x_4 = 1 , \qquad \theta = 0 . \qquad (16)$$

The internal state Z is, according to eq. (16) given by

$$Z = 1 \cdot (+3) + 1 \cdot (-1) + 1 \cdot (-2) + 1 \cdot (+1) = +1 = -\epsilon ;$$

hence, according to rule (vii) eq. (16), we have:

$$y = \phi(1) = 1 ,$$

so the element "fires"; its output is ON. Raising the threshold one unit, $\theta = 1$, still keeps the element in its ON state, because its internal state does not fall below zero. From this we can conclude that a completely disconnected element with zero threshold always has its output in the ON state.

If in fig. 9 the threshold is raised to +2, the simultaneous excitation of all fibers will not activate the element. Furthermore, it is easily seen that with threshold +5 this element will never fire, whatever the input configuration; with threshold -4 it will always fire.

The two-valuedness of all variables involved, as well as the possibility of negation (inhibition)

Fig. 10 Threshold defining the function computed by a McCulloch formal neuron.

and affirmation (excitation), make this element an ideal component for computing logical functions in the calculus of propositions where the ON or OFF state of each input fiber represents the truth or falsity of a proposition X_i, and where the ON or OFF state of the output fiber Y represents the truth value of the logical function ϕ computed by the element.

Let us explicate this important representation with a simple example of an element with two input fibers only ($N = 2$), each attached to the element with only a single facilitatory junction ($n_1 = n_2 = +1$). We follow classical usage and call our input fibers A and B, – rather than X_1 and X_2 (which pays off only if many input fibers are involved and one runs out of letters of the alphabet). Fig. 10 illustrates the situation. First, we tabulate all input configurations – all "input states" that are possible with two input fibers when each may be independently ON or OFF. We have four cases: A and B both ON or both OFF, and A ON and B OFF, and A OFF and B ON, as indicated in the left double column in fig. 10.

In passing, we may point out that with N input fibers, two choices for each, we have in general

$$\mathcal{H}_{in} = 2^N \qquad (17)$$

possible input states.

Returning to our example, we now tabulate, for a particular threshold value θ, the output $\phi_\theta(A, B)$ which has been computed by this element for all input states, i.e., all combinations of A, B being ON or OFF. For zero threshold ($\theta = 0$), this element will always be in its ON state, hence, in the column $\theta = 0$ we insert "one" only. We raise the threshold one unit ($\theta = 1$) and observe that our element will fire only if either A or B, or both A and B are ON. We proceed in raising the threshold to higher values until a further increase in threshold will produce no changes in the output functions, all being always OFF.

In order to see that for each threshold value this element has indeed calculated a logical function on the propositions A and B, one has only to interpret the "zeros" and "ones" as "false" and "true" respectively, and the truth values in each θ-column, in conjunction with the double column representing the input states, become a table called – after Wittgenstein (1956) – the "truth table" for the particular logical function. In the example of fig. 10, the column $\theta = 0$ represents "Tautology" because $\phi_0(A, B)$ is always true (1), independent of whether or not A or B are true: "A or not-A, and B or not-B". For $\theta = 1$ the logical function "A or B" is computed; it is
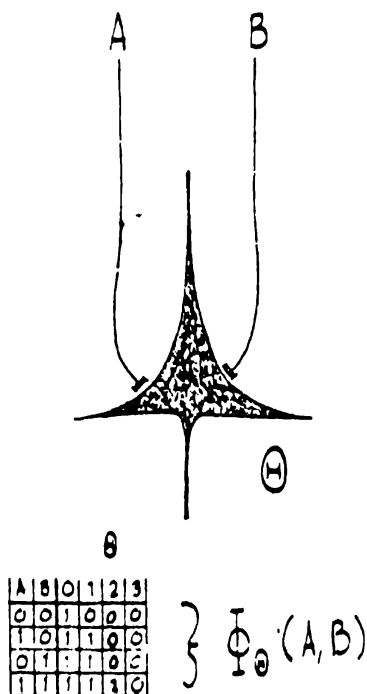
false (0) only if both A and B are false. $\theta = 2$ gives "A and B" which, of course, is only true if both A and B are true, etc.

Today there are numerous notations in use, all denoting these various logical functions, but based on different reasons for generating the appropriate representations, which all have their advantages or disadvantages.

The representation we have just employed is that of Wittgenstein's truth table. This representation permits us to compute at once the number of different logical functions that are possible with N propositions (arguments). Since we know that N two-valued arguments produce n_{in} different states, each of which again has two values, true or false, the total number of logical functions is, with eq. (17):

$$n_{LF} = 2^{n_{IN}} = 2^{2^{N}} . \tag{18}$$

For two arguments $(N = 2)$ we have precisely 16 logical functions.

Another symbolism in use is that proposed by Russell and Whitehead (1925), and Carnap (1925) who employ the signs "•", "v", "→", "-" for the logical "and", "or", "implies", "non" respectively. It can be shown that all other logical func-

tions can be represented by a combination of these functions.

Finally, we wish to mention still another form for representing logical functions, with the aid of a formalized Venn diagram. Venn, in 1881, proposed to show the relation of classes by overlapping areas whose various sections indicate joint or disjoint properties of these classes (fig. 11). McCulloch and Pitts (1943) dropped the outer contours of these areas, using only the center cross as lines of separation. Jots in the four spaces can represent all 16 logical functions. Some examples for single jots are given in fig. 11. Expressions with two or more jots have to be interpreted as the expressions with single jots connected by "or". Hence,

$$\overset{\cdot\cdot}{\times} = \text{"(neither } A \text{ nor } B) \text{ or } (A \text{ and } B)\text{"} ,$$

which, of course, represents the proposition "A is equivalent to B". The similarity of this symbol with the greek letter chi suggested the name "chiastan" symbol. The advantage of this notation is that it can be extended to accommodate logical functions of more than two arguments (Blum, 1962).

In table 1 we show that, with two exceptions all logical functions with two input lines can be computed by a single McCulloch formal neuron, if full use is made of the flexibility of this element by using various thresholds and synaptic junctions. For convenience, this table lists, in six different "languages", all logical functions for two arguments. The first column gives a digital representation of Wittgenstein's truth function (second column), taken as a binary digit number to be read downwards. The third column gives the appropriate chiastan symbol, and the fourth column shows the corresponding element with its synaptic junctions and its appropriate threshold value. The two functions which cannot be computed by a single element require a network of three elements. These are given in fig. 12 and are referred to in the appropriate entries. The fifth column shows the same functions in Russell-Carnap symbols, while the sixth column translates it into English.

The seventh, and last, column lists the "logical strength" Q of each function. According to Carnap (1938) it is possible to assign to each logical function a value which expresses the intuitive feeling for its strength as a logical function. We intuitively consider a logical function to be weak if it is true in most of the cases, irrespective of whether its arguments are true or false. The tautology, which is always true, tells
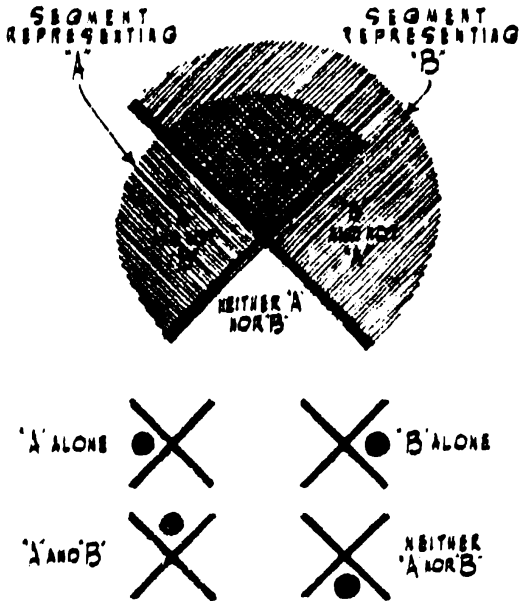

Fig. 11. Development of the Chiastan symbol for logical functions from Venn's diagrams.

Table I

1	2	3	4	5	6	7
#	A B	GESTALT SYMBOL	NEURAL NET A B θ=THRESHOLD	SYMBOLIC LOGIC	NAME	Q
0		✕	θ=3	$(A \cdot \bar{A})(B \cdot \bar{B})$	CONTRADICTION (ALWAYS FALSE)	4
1		✕	θ=0	$\bar{A} \cdot \bar{B}$	NEITHER A NOR B	3
2		✕	θ=2	$A \cdot \bar{B}$	A ONLY	3
3		✕	θ=1	$\bar{B}$	NOT B	2
4		✕	θ=2	$\bar{A} \cdot B$	B ONLY	3
5		✕	θ=1	$\bar{A}$	NOT A	2
6		✕	FIG 12b	$(A \cdot \bar{B})(B \cdot \bar{A})$	EITHER A OR B (EXCLUSIVE OR)	2
7		✕	θ=1	$\bar{A} + \bar{B}$	NOT BOTH, A AND B	1
8		✕	θ=2	$A \cdot B$	A AND B	3
9		✕	FIG 12a	$(A \cdot B) + (\bar{A} \cdot \bar{B})$ $A \rightleftarrows B$	A IS EQUIVALENT TO B	2
10		✕	θ=1	A	A	2
11		✕	θ=0	$B \rightarrow A$	B IMPLIES A	1
12		✕	θ=1	B	B	2
13		✕	θ=0	$A \rightarrow B$	A IMPLIES B	1
14		✕	θ=1	$A + B$	A OR B (INCLUSIVE OR)	1
15		✕	θ=0	$(A + \bar{A})(B + \bar{B})$	TAUTOLOGY (ALWAYS TRUE)	0
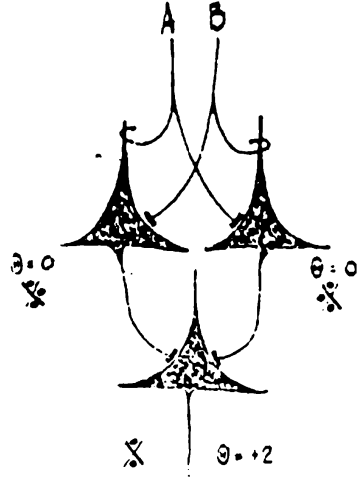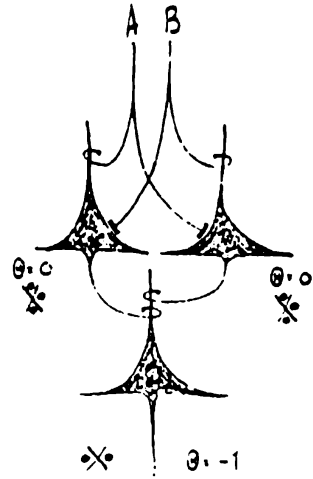
Fig. 12. Nets of McCulloch formal neurons computing the logical functions (a) "A is equivalent to B" and (b) "either A or else B".

us nothing about the truth values of its arguments. On the other hand, a function which is true only when all its arguments are true we consider to be a strong logical function. Consequently, as a measure of strength one may take the number of ways in which the logical function is false. In other words, counting the number of zeros in Wittgenstein's truth table gives the logical strength of the function. Inspection of table 1 shows an interesting relationship between threshold and strength, because for a given synaptic distribution the logical strength increases with increasing threshold. This observation will be of importance in our discussion of

adaptive nets, because by just raising the thresh-
old to an appropriate level, the elements will be
constrained to those functions which "education"
accepts as "proper".

Since we have shown that all logical functions
with two variables can be represented by McCul-
loch formal neurons, and since in drawing net-
works composed of elements that compute logical
functions it is in many cases of no importance to
refer to the detailed synaptic distributions or
threshold values, we may replace the whole
gamut by a single box with appropriate inputs
and outputs, keeping in mind, however, that the
box may contain a complex network of elements
operating as McCulloch formal neurons. This
box represents a universal logical element, and
the function it computes may be indicated by at-
taching to it any one of the many available sym-
bolic representations.

We shall make use of this simplified formal-
ism by introducing an element that varies the
functions it computes, not by manipulation of its
thresholds but according to what output state was
produced, say, one computational step earlier.
Without specifying the particular functions this
element computes, we may ask what we can ex-
pect from such an element, from an operational
point of view. The mathematical formalism that
represents the behavior of such an element will
easily show its salient features. Let $X(t)$ be the
N-tuple $(x_1, x_2, x_3, \ldots, x_N)$ representing the input
state at time t for N input fibers, and $Y(t)$ the
M-tuple $(y_1, y_2, \ldots, y_M)$ representing its output
state at time t. Call Y' its output state at $t - \Delta t$.
Hence,

$$Y = \Phi(X, Y') . \qquad (19)$$

In order to solve this expression we have to know
the previous output state Y' which, of course, is
given by the same relation only one step earlier
in time. Call X' the previous input state, then:

$$Y' = \Phi(X', Y") .$$

and so on. If we insert these expressions into
eq. (19), we obtain a telescopic equation for Y in
terms of its past experience $X', X", X'", \ldots$ and
Y_0, the birth state of our element:

$$Y = \Phi(X, X', X", X'", \ldots, Y_0) .$$

In other words, this element keeps track of
its past and adjusts its *modus operandi* accord-
ing to previous events. This is doubtless a form
of "memory" (or another way of adaptation)
where a particular function from a reservoir of
available functions is chosen. A minimal element
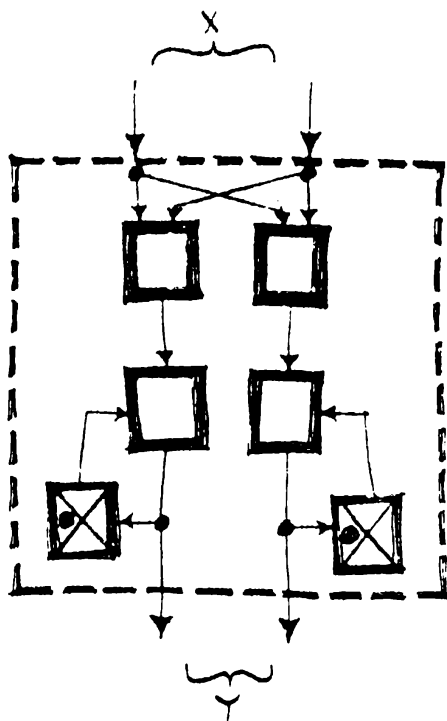that is sufficient for the development of cumula-

Fig. 13. Ashby element computing recursive functions.

tively adaptive systems has been worked out by
Ashby (see Fitzhugh, 1963) (see fig. 13). It is
composed of at least one, at most three, McCul-
loch formal neurons, depending upon the func-
tions to be computed in the unspecified logical
elements. We shall call such a minimal element
an "Ashby element". The mathematical machin-
ery that goes along with such elements is called
recursive function theory, hence elements of this
general form may be called recursive elements.

We have as yet discussed elements with two
inputs only. However, it is easily seen that Mc-
Culloch's concepts can be extended to neurons
with many inputs as fig. 9 may remind us. How-
ever, the number of logical functions that cannot
be computed by using only a single neuron in-
creases rapidly — 2 out of 16 for two inputs, and
152 out of 256 possible functions for three inputs
(Verbeek, 1962) — and networks composed of sev-
eral elements have to be constructed. These
networks will be discussed in the following chap-
ter.

In the preceding discussion of the operations

of a McCulloch formal neuron a tacit assumption was made, namely, that the information carried on each fiber is simply its ON or OFF state. These states have to be simultaneously presented to the element, otherwise its output is meaningless with respect to these states. A term like "input strength" is alien to this calculus; a proposition is either true or false. This requires all components in these networks to operate synchronously, i.e., all volleys have not only to be fired at the same frequency, they have also to be always in phase. Although there are indications that coherency of pulse activity is favored in localized areas – otherwise an E.E.G. may show only noise – as long as we cannot propose a mechanism that synchronizes pulse activity, we have to consider synchronism as a very special case. Since this article is not the place to argue this out, we propose to investigate what happens as pulses, with various frequencies, pass over various fibers.

3.1.2. Asynchronism

Assume that along each fiber X_i travels a periodic chain of rectangular pulses with universal pulse width Δt (see fig. 14), but with time intervals τ_i varying from fiber to fiber. The probability that fiber X_i activates its synaptic junctions at an arbitrary instant of time is clearly

$$p_i = \frac{\text{duration of pulse}}{\text{duration of pulse interval}} = \frac{\Delta t}{\tau_i} , \qquad (20)$$

Fig. 14 Schematic of pulse width, pulse interval and refractory period.

or, replacing the periodic pulse interval on fiber X_i by the frequency f_i, we have

$$p_i = f_i \, \Delta t , \qquad (21)$$

for the probability that X_i's synaptic junctions are activated, and the probability

$$q_i = 1 - p_i \qquad (22)$$

that they are inactivated. The probability of a particular input state $X(x_1, x_2, \ldots, x_N)$ which is characterized by the distribution of "ones" and "zeros" of the input values x_i, and which may be represented by an N-digit binary number

$$0 \leqslant X \leqslant 2^N - 1 , \qquad (23)$$

$$X = \sum_1^N x_i \, 2^{i-1} , \qquad x_i = 0, 1 , \qquad (24)$$

is given by the Bernoulli product

$$P_X = \prod_1^N p_i^{x_i} \, (1 - p_i)^{(1 - x_i)} \qquad (25)$$

with

$$\sum_{X=0}^{X=2^N-1} P_X = 1 , \qquad (26)$$

which simply arises from the consideration that the simultaneous presence or absence of various events with probabilities p_i or $(1-p_i)$ respectively is just the product of these probabilities. The presence or absence of events is governed by the exponents x_i and $(1 - x_i)$ in the Bernoulli product which are 1 and $(1 - 1) = 0$ in presence, and 0 and $(1 - 0) = 1$ in absence of the event of interest, namely the activation of the synaptic junctions of the ith fiber.

Having established the probability of a particular input state, we have simply to find out under which conditions the element fires in order to establish its probability of firing. This, however, we know from our earlier considerations (eqs. (15), (16)) which define those input states that activate the output fiber. As we may recall, an activated output $(y = 1)$ is obtained when the internal state Z equals, or exceeds, zero:

$$Z = \sum n_i x_i - \theta \, \cdot -\epsilon$$

with n_i representing the number of (positive or negative) synaptic junctions of the ith fiber and θ being, of course, the threshold. Hence, for a given threshold and a certain input state $X(x_1, x_2, \ldots, x_N)$ output state $y_{\theta, X}$ is defined by

$$y_{\theta,x} = \begin{cases} 1 \text{ for } \Sigma n_i \, r_i - \theta \cdots - \epsilon \,, \\ 0 \text{ for } \Sigma n_i \, r_i - \theta \quad - \epsilon \,. \end{cases} \tag{26}$$

Since whenever the output is activated y will assume a value of unity, the probability p of its activation is the sum of all probabilities of those input states that give y a value of "one":

$$p = \sum_{X=0}^{X=2^N-1} y_{\theta,x} \, P_X \,. \tag{27}$$

Since all terms in the above expression will automatically disappear whenever an input state is present that fails to activate the output ($y = 0$), eq. (27) represents indeed the activation probability of the output fiber. If we again assume that the ON state of the output fiber conforms with the universal pulse duration Δt, which holds for all pulses traveling along the input fibers, we are in a position to associate with the probability of output excitation a frequency f according to eqs. (26), (27):

$$f = \frac{1}{\Delta t} \sum y_{\theta,X} \, P_X \,, \tag{28}$$

which we will call — for reasons to be given in a moment — the "internal frequency".

Let us demonstrate this mathematical apparatus in the simple example of fig. 10. With letters A and B for the names of fibers X_1 and X_2 we have, of course, $A(x_1 = 0, 1)$ and $B(x_2 = 0, 1)$ and the four possible input states are again:

x_1	x_2
0	0
1	0
0	1
1	1

The four corresponding Bernoulli products are after eq. (25):

$$(1 - p_1)(1 - p_2)$$
$$p_1 \cdot (1 - p_2)$$
$$(1 - p_1) \cdot p_2$$
$$p_1 \cdot p_2$$

In order to find which of these products will contribute to our sum (eq. (27)) that defines the desired output activation probability, we simply consult the truth table for the function that is computed after a specific threshold has been selected, and add those terms to this sum for which the truth tables give a "one" ($y = 1$). The

following table lists for the four values of θ as chosen in fig. 10 the resulting probabilities of output excitation according to eq. (27).

Table 2

θ	Sum of Bernoulli products = p
0	$(1-p_1)(1-p_2)+(1-p_1)p_2+p_1(1-p_2)+p_1 p_2 = 1$
1	$p_1(1-p_2)+(1-p_1)p_2+p_1 p_2 = p_1+p_2-p_1 p_2$
2	$p_1 p_2 = p_1 p_2$
3	$\cdots = 0$

With eq. (28) we have at once what we called the internal frequency of the element when its threshold is set to compute a particular logical function. Using the notation as suggested in column 5 of table 1 for denoting logical functions as subscript, and denoting with T tautology and with C contradiction, the computer frequencies representing the various logical functions of the arguments f_1 and f_2, again represented as frequencies, are as follows.

Table 3

$\theta = 0$	$f(T) = 1/\Delta t$
$\theta = 1$	$f(v) = f_1 + f_2 - \Delta t f_1 f_2$
$\theta = 2$	$f(\bullet) = \Delta t f_1 f_2$
$\theta = 3$	$f(C) = 0$

This table indicates an interesting relationship that exists between the calculus of propositions and the calculus of probabilities (Landahl et al., 1943). Furthermore, it may be worthwhile to draw attention to the fact — which may be shown to hold in general — that low threshold values, resulting in "weak" logical functions, give this element essentially a linear characteristic; it simply adds the various stimuli f_i. This is particularly true, if the stimuli are weak and their cross-products can be neglected. For higher threshold values the element is transformed into a highly non-linear device, taking more and more cross-products of intensities into consideration until for the strongest logical function being short of contradiction — the logical "AND" — the single cross-product of all stimuli is computed.

We have carefully avoided associating with "p" or with "f" the actual output frequency. We called f the "internal frequency". The reason will become obvious in a moment. It is well known that after the production of each pulse a physiological neuron requires a certain moment

of time Δt_R - the so-called "refractory period" - to recover from its effort and to be ready for another pulse (see also fig. 14). But in table 3 the resulting frequency for zero threshold is the reciprocal of the pulse duration Δt, which, of course, is unmanageably high for a neuron whose refractory period is clearly much longer than the duration of its pulse

$$\Delta t_R \quad \Delta t . \qquad (29)$$

It is very easy indeed to accommodate this difficulty in our calculations, if we only realize that the actual output frequency f_0 of the element is the frequency f at which it "wants" to fire, reduced by the relative time span in which it cannot fire:

$$f_0 = f(1 - f_0 \Delta t_R) . \qquad (30)$$

Solving for the output frequency in terms of internal frequency and refractory period we have:

$$f_0 = \frac{f}{1 + f \Delta t_R} . \qquad (31)$$

From this is easily seen that the ultimate frequency at which an element can fire is asymptotically approached for $f \to 1/\Delta t$, and is given by:

$$f_{0\,max} = \frac{1}{\Delta t + \Delta t_R} \quad \frac{1}{\Delta t} . \qquad (32)$$

which is in perfect agreement with our concept of the physiological behavior of this element. The actual values for pulse duration and refractory period may be taken from appropriate sources (Eccles, 1952; Katz, 1959).

3.2. The neuron as an "integrating element"

The element discussed in the previous paragraphs is an "All or Nothing Device" par excel-

lence, and in the two subtitles "Synchronism" and "Asynchronism" we investigated only how this element behaves when subjected to stimuli for which we changed our interpretation of what is a meaningful signal. While in the synchronous case a string of OFFs and ONs was the signal, in the asynchronous case pulse frequencies carried the information. However, in both cases the element always operated on a "pulse by pulse" basis, thus reflecting an important feature of a physiological neuron.

In this paragraph we wish to define an element that incorporates some of the concepts which are associated with neurons as they were first postulated by Sherrington (1952), and which are best described in the words of Eccles (1952, p. 191): "All these concepts share the important postulate that excitatory and inhibitory effects are exerted by convergence on to a common locus, the neuron, and there integrated. It is evident that such integration would be possible only if the frequency signalling of the nervous system were transmuted at such loci to graded intensities of excitation and inhibition".

In order to obtain a simple phenomenological description of this process we suggest that each pulse arriving at a facilitatory or inhibitory junction releases, or neutralizes, a certain amount q_0 of a hypothetical agent, which, left alone, decays with a time constant $1/\lambda$. We further suggest that the element fires whenever a critical amount q^* of this agent has been accumulated (see fig. 15).

Again we assume N fibers X_i attached to the element, each having n_i facilitatory (+) or inhibitory (-) synaptic junction. Each fiber operates with a frequency f_i. The number of synaptic activations per unit time clearly is the algebraic sum:
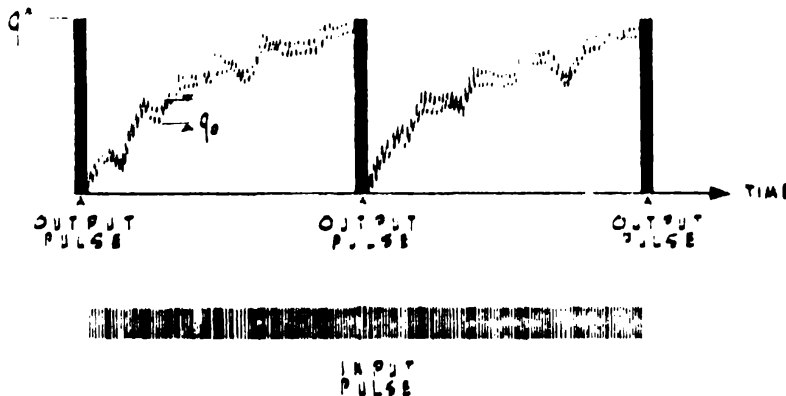
$$S = \sum_1^N n_i f_i . \qquad (33)$$

Hence, the differential equation that describes the rate of change in the amount of the agent q as a consequence of stimulus activity and decay is

$$\frac{dq}{dt} = Sq_0 - \lambda q , \qquad (34)$$

whose solution for q as a function of time t is:

$$q = \frac{q_0 S}{\lambda} (1 - e^{-\lambda t}) , \qquad (35)$$

if at time $t = 0$ we also have $q = 0$.

Let Δt^* be the time required to accumulate the amount q^* at the element, and let the activity S change during many such time intervals. Clearly, the "internal frequency" of this element is

$$f = 1/\Delta t^* . \qquad (36)$$

Inserting these into eq. (35):

$$q^* = \frac{q_0 S}{\lambda} (1 - e^{-\lambda/f}) , \qquad (37)$$

we have an expression that relates the frequency f with the stimulus activity S. For convenience we introduce new variables x and y which represent normalized input and output activity respectively and which are defined by:

$$x = Sq_0/\lambda q^* , \qquad y = f/\lambda . \qquad (38)$$

With these, eq. (37) can be rewritten:

$$\frac{1}{x} = 1 - e^{-1/y} , \qquad (39)$$

or, solved for y:

$$y = 1/[\ln x - \ln(x-1)] . \qquad (40)$$

This relation shows two interesting features. First, it establishes a threshold for excitation:

$$x_\theta = 1$$

or

$$S_\theta = \lambda q^*/q_0 .$$

Because for the logarithmic function to be real its argument must be positive, or $x > 1$. It may be noted that the threshold frequency S_θ is given only in terms of the element's intrinsic properties λ, q_0 and q^*.

The second feature of the transfer function of this element is that for large values of x it becomes a linear element. This is easily seen if we use the approximations

$$\frac{1}{1-\epsilon} \approx 1 + \epsilon$$

and

$$\ln(1+\epsilon) \approx \epsilon$$

for

$$\epsilon = 1/x \ll 1 .$$

Under these conditions eq. (40) becomes simply:

$$y = x$$

or

$$f = Sq_0/q^* .$$

This "nice" feature of our element is, of course, spoiled by the considerations which were presented earlier, namely, that the activation frequency S, which is the sum-total of the impinging frequencies (see eq. (33)) might be too high to be handled by a physiological neuron. In order to adjust for the frequency limit that is expected from our element, we proceed in precisely the same way as was suggested in eq. (30): we introduce into eq. (41) a formal interaction term that reduces its potential activity x commensurate with its actual activity:

$$y_0 = (1 - \mu y_0) , \qquad (42)$$

where the factor μ will become evident in a moment. Solving for y_0

$$y_0 = \frac{x}{1 + \mu x} \qquad (43)$$

and for $x \to \infty$ we have

$$y_{0\,max} = 1/\mu . \qquad (44)$$

Denormalization according to (38) and comparison with (32) gives

$$f_{0\,max} = \frac{\lambda}{\mu} = \frac{1}{\Delta t + \Delta t_R} , \qquad (45)$$

or

$$\mu = \lambda(\Delta t + \Delta t_R) . \qquad (46)$$

In other words, the parameter μ expresses all neuronic delays in units of the agent's decay constant.

In order to make the high frequency correction applicable for the whole operational range of our element, we simply replace x in eq. (40) by $x/(1 + \mu x)$ of eq. (43) which is the adjusted equivalent to the unadjusted eq. (41).

With this adjustment we have the output frequency of our element defined by

$$y_0 = \frac{1}{\ln \frac{x}{x(1-\mu)-1}} . \qquad (47)$$

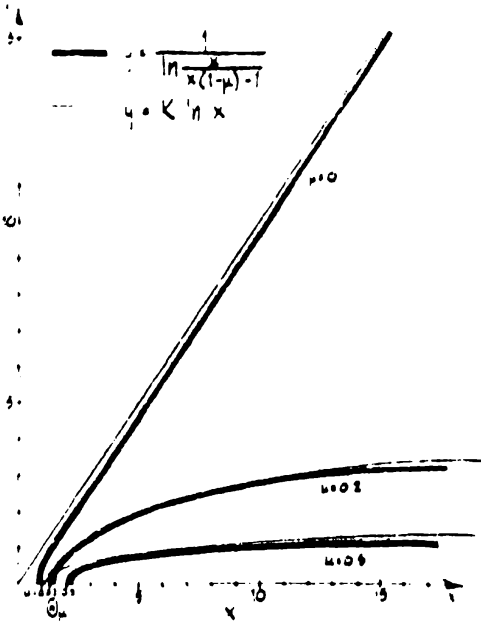A graphical representation of the input-output relationship of this element according to eq. (47)

Fig. 10. Transfer function of an integrating element under various operating conditions (μ) (bold line), and approximation by a logarithmic function (thin line).

is given in fig. 16 for three different values of μ (0; 0.2; 0.5). For small μ (μ 0.05) the element is linear with threshold $x_0 = 1$. For large μ (~ 0.5) the element is an almost perfect "All or Nothing" device with threshold

$$x_0 = \frac{1}{1 - \mu} .$$ (48)

For values of μ in the range 0.1 — 0.5 the element displays a logarithmic transfer function. Indeed, it can easily be shown that in the "vicinity" of

$$x \approx \frac{1}{\mu(1 - \mu)} .$$ (49)

Eq. (47) can be approximated by

$$y_0 = K \ln x ,$$ (50)

$$K = \frac{1}{4} \frac{\mu}{(1 - \mu)[\ln(1 - \mu)]^2} .$$ (51)

It may be noted that this "vicinity" extends over an appreciable range as can be seen by the approximation according to (50) which in fig. 16

is superimposed in thin line over the exact curves given in bold line.

With the choice of the parameter μ we are in a position to change the transfer function of this element considerably. We suggest the following nomenclature for the elements that arise for different values of μ:

$\mu = 0$: "Sherrington" element (linear characteristic)

$\mu \approx 0.2$: "Weber-Fechner" element (logarithmic characteristic)

$\mu \to 1$: "All or Nothing" element (step-function characteristic).

With these and the other elements as specified in the previous paragraph, the McCulloch element and the Ashby element, we are now prepared to discuss the behavior of networks that incorporate these elements as their basic computer components.

4. SOME PROPERTIES OF COMPUTING NETWORKS

Of the myriad networks that are not only theoretically possible but are indeed incorporated into the neural architecture of living organisms, space and ignorance will permit only a small glimpse into their vast richness. In addition, the large variety of solutions that evolution has provided in different species for their specific cognitive problems makes it difficult to present this topic from a single ordering point of view, except that here we are dealing with networks. However, in the last decades a number of general principles have been carved out from this large complex of problems and in the following an attempt will be made to do justice to some of them by briefly suggesting their conceptual framework and by giving some examples to illustrate the underlying ideas.

We shall open our discussion with two paragraphs which represent extremes in the spectrum that goes from the concrete to the abstract. The first paragraph shows the possibility of orderly behavior in a "mixed" net, the neuromuscular net in the sea urchin, where within elements of their own kind no interaction takes place, but where each kind uses the other for integrated action. The second paragraph touches briefly the McCulloch-Pitts theorem which, in a sense, ends or starts all discussions about networks.

The next paragraphs discuss the development of cognitive networks, first, in which cellular

identity is recognized, while the subsequent con-
siderations are based solely on the localizability
of groups of cells, but their individuality is lost.
The chapter concludes with a brief account of
stability and immunology of neural networks and
with some remarks on adaptive nets and how
they store information.

4.1. A neuro-muscular net

Fulton (1943) opens his comprehensive trea-
tise on neurophysiology with a brief account of
the early evolutionary stages in the development
of the nervous system. Rightly so, because the
appreciation of these early stages leaves no
doubt as to the ultimate purpose of this system,
namely, to serve as a computer that links detec-
tion with appropriate action. Following Parker
(1943) we give in fig. 17 schematically the three
decisive steps which are the foundation for the
emergence of neural systems with the complexity
of a mammalian brain. Fig. 17a shows symbo-
lically the "independent effector" (muscle cell in
a sponge) that translates directly a general
"stimulus" into action – contraction in most
cases. The first ste . rom detection to discrimi-
nation is accomplished by separating detection
and action and localizing these functions in dif-
ferent elements (fig. 17b). This permits the de-
velopment of specific sensors responsive to cer-
tain stimuli only (light, chemistry, touch, etc.).
The final step in preparing the tripartite archi-
tectural organization of the nervous system –
detector, computer, effector – is suggested in
fig. 17c, where an intermediate ganglion cell
acts as a primordial nucleus for what is to be-
come the information processing interface be-
tween detection and action.

Although an array of such simply organized
units as in fig. 17b appears not to have the prop-
erties which we would expect from a neural net,
for there is no direct connection from neuron to
neuron, these systems still deserve to be called
interaction nets from a general point of view,
because a particular state in one unit – say a
contraction – may influence the state of its neigh-
bors via the mechanical properties of the medi-
um in which they are embedded. That such
"mixed nets" are capable of highly organized be-
havior may be illustrated with the beautiful ob-
servations of Kinosita (1941) on the kinetics of
the spines of the sea-urchin.

When a localized stimulus is given to the body
surface of a sea-urchin, the spines around the
stimulated spot respond so as to lean towards
the stimulated spot, and this response dimin-
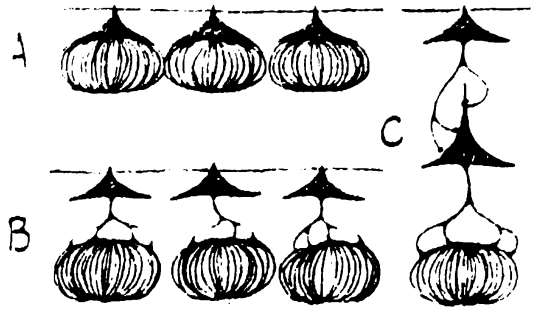ishes rapidly with distance from the locus of

Fig. 17. Primitive nerve nets: (a) independent effector;
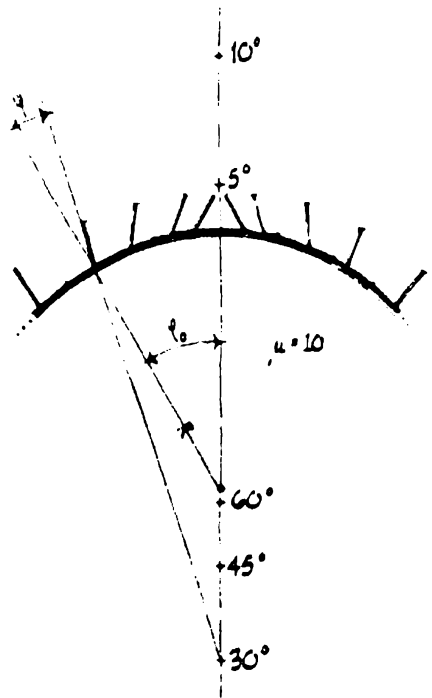(b) receptor-effector system; (c) receptor-computer-
effector system.

Fig. 18. Position of spines in the sea-urchin after
stimulation at the North pole.

stimulation (fig. 18). The first thought that comes
to mind is to assume an anastomosing plexus of
interacting nerve cells which transmit the infor-
mation of this perturbation over an appropriate
region to cause contraction of the muscle fibers
attached to the spines (Uxküll, 1896). However,
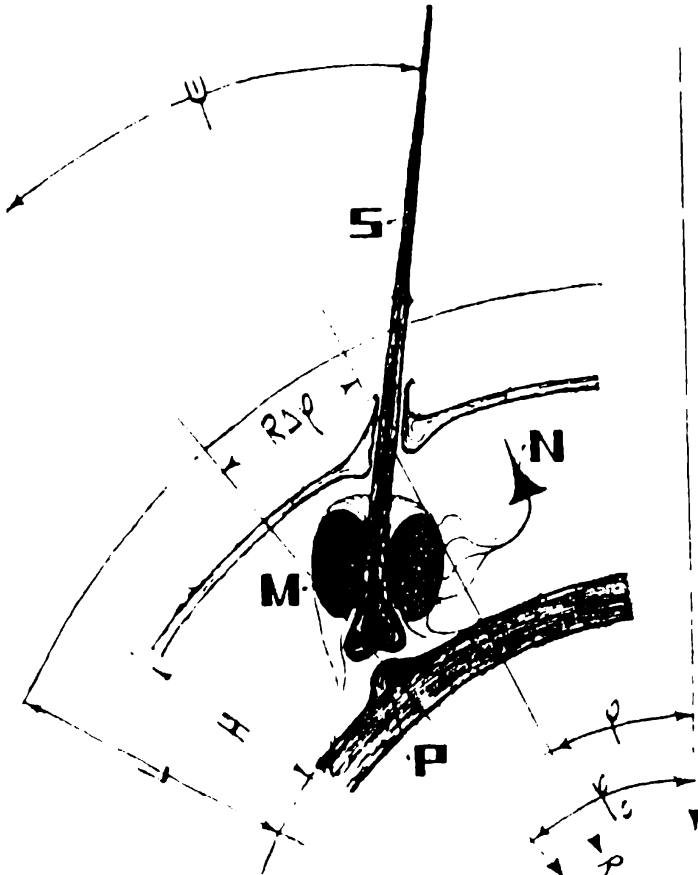Kinosita was able to demonstrate that there are

Fig. 19. Schematic of anatomy and geometry in the vicinity of a sea-urchin spine

no fiber to fiber connections, only proximate fiber-muscle connections, hence each receptor pair has to operate according to local information of the deformation of its surroundings caused by deformations of the more distant regions.

Local anatomy and geometry in the neighborhood of a spine is schematically given in fig. 19 which exaggerates certain proportions for purposes of clarity. The spine S, centered on pivot P which is attached to a fixed shell with radius R, can bend in all directions. Muscle fibers M contract when stimulated by neuron N which will respond to an extension (stretch) of the integument. If somewhere at the surface a muscle bundle contracts, it causes the integument to follow, which produces a slight local stretch that is sensed by the local neuron which, in turn, causes its associated muscle to contract, and so on. Consider a spine localized at angle ϕ_0. When

bent at angle ψ from its radial rest position ($\psi = 0$) it will shift the integument surface from ϕ_0 to ϕ. Assume that a stimulus is applied at the North Pole ($\phi_0 = 0$), then the shift $\Delta\phi = \phi_0 - \phi$ at angle ϕ_0 is the result of the summation of differential contractions $-d(\Delta\phi)/d\phi_0$ of intermediate muscles. These, in turn, contract according to the efferent stimulus of their associated neurons which fire in proportion to the local perturbation, i.e., the difference between the extension of the relaxed integument L and the stretched integument H. Hence, the differential equation that governs the *local* receptor-effector system is:

$$\frac{d(\Delta\phi)}{d\phi_0} = -k(L - H) , \qquad (52)$$

where the proportionality constant k represents the combined transfer functions for neuron and

muscle fiber. From inspection of fig. 19 we have the simple geometric relations:

$$H = L \cos \psi$$

and

$$R\Delta\phi = L \sin \psi . \qquad (53)$$

Introducing a dimensionless paremeter μ which combines physiological and anatomical constants:

$$\mu = kR ,$$

the differential equation (52) can be rewritten with the aid of (53) to read:

$$\frac{d \sin \psi}{d\phi_0} = -\mu(1 - \cos \psi) , \qquad (54)$$

which can readily be solved to yield a transcendental equation in ψ:

$$\cot(\tfrac{1}{2}\psi) + \psi = \mu\phi_0 + \cot(\tfrac{1}{2}\psi_0) + \psi_0 , \qquad (55)$$

where ψ_0 denotes the original perturbation at $\phi_0 = 0$. For a chosen value of $\mu = 10$ this equation was numerically evaluated and served as a basis to conxtruct fig. 18. The entries along the axis of symmetry indicate the focal points of spines located at corresponding angles ϕ_0. The original perturbation is assumed to be strong ($\psi_0 = 45°$).

The sole purpose of this somewhat detailed account of a relatively insignificant network was to suggest that organized behavior that seems to be governed by a central control that operates according to an "action at a distance" principle can very well arise from a localized point function that permanently links elements in an infinitesimal neighborhood. Notice how the behavior can be expressed in differential eqs. (52) or (54) which contain local properties only. The whole system swings into action whenever the "boundary value" – i.e., the stimulus – changes. The operational principle here is "action by contagion". We shall later discuss this principle in greater detail in connection with interaction networks.

4.2. *The McCulloch-Pitts theorem*

A network composed of McCulloch elements we shall call a "McCulloch formal network". The central issue of the McCulloch-Pitts (1943) theorem is the synthesis of such networks, which compute any one of the 2^{2^n} logical functions that can be defined by n propositions. In other words, any behavior that can be defined at all logically, strictly, and unambiguously in a finite number of words can be realized by such a formal network. Since in my opinion this theorem not only is one

of the most significant contributions to the epistemology of the 20th century, but also gives important clues as to the analysis of physiological neural nets, it is impossible in an article about nerve nets not, at least, to touch upon the basic ideas and consequences that are associated with this theorem. Its significance has best been appraised in the words of the late John Von Neumann (1951):

"It has often been claimed that the activities and functions of the human nervous system are so complicated that no ordinary mechanism could possibly perform them. It has also been attempted to name specific functions which by their nature exhibit this limitation. It has been attempted to show that such specific functions, logically completely described, are *per se* unable of mechanical neural realization. The McCulloch-Pitts result puts an end to this. It proves that anything that can be exhaustively and unambiguously described, anything that can be completely and unambiguously put into words, is *ipso facto* realizable by a suitable finite neural network".

We shall give now a brief summary of the essential points of this theorem. As already mentioned, the McCulloch-Pitts theorem shows that to any logical function of an arbitrary number of propositions (variables) a network composed of McCulloch elements can be synthesized that is equivalent to any one of these logical functions. By "equivalence" is meant that it functions so as to compute the desired logical function. This can be accomplished by singling out input fibers of some of its elements and output fibers of some other elements and then defining what original stimuli on the former are to cause what ultimate responses of the latter.

Since the basic element of the networks to be discussed – the McCulloch formal neuron – is capable of computing only some of all logical functions which can be constructed from precisely two variables, and since an arbitrary set of propositions may contain temporal relationships, we require three more steps to reach the generality claimed by the theorem. The first step involves purely logical argument and shows (a) by using substitution and the principle of induction the possibility of constructing n-variable expressions from two-variable expressions, and (b) the possibility of expressing uniquely any logical function of n-variables in a certain normal form. The second step introduces an operator S that takes care of a single synaptic delay and thus permits the representation of temporal relationships, while the third step utilizes some formal properties of this operator to obtain nor-

mal form expressions that are immediately translatable into network language.

First step:

(a) Consider a logical function of the two variables A_1 and B_1

$$[A_1, B_1] . \qquad (56)$$

Let B_1 be a logical function of the two variables A_2 and B_2:

$$B_1 = [A_2, B_2] ,$$

and, in general,

$$B_i = [A_{i+1}, B_{i+1}] . \qquad (57)$$

Iterative substitution of (57) into (56) gives:

$$[A_1, A_2, A_3, \ldots , A_n, B_n] ,$$

which, by induction, holds for all .

(b) It can be shown (Hilbert and Ackermann, 1928) that any logical function of n arguments can be represented by a partial conjunction ($\cdot$) of disjunctions (v) that contain each variable X_i either affirmed ($x_i = 1$) or negated ($x_i = 0$). Let each disjunction be represented by D_Z, where Z is the decimal representation of the binary number

$$0 \leq Z = \sum_1^n 2^{i-1} x_i \leq 2^n - 1 , \qquad (58)$$

and let K_p represent the (partial) conjunction of those disjunctions present ($D_Z = 1$, otherwise $D_Z = 0$) in the logical function, where P is the decimal representation of the binary number

$$0 \leq P = \sum_0^{2^n-1} 2^Z D_Z \leq 2^{2^n} - 1 . \qquad (59)$$

The terms K_p are called "Schroeder's constituents". Consequently, there are 2^{2^n} different conjunctions possible with these constituents. Each of these conjunctions represents uniquely one logical function.

Expanding these conjunctions by virtue of the distributivity of ($\cdot$) and (v) into a disjunction of conjunctions

$$C_1 \text{ v } C_2 \text{ v } C_3 \ldots , \qquad (60)$$

one arrives, after cancellation of contradictory terms ($X_i \cdot \bar{X}_i$), at Hilbert's disjunctive normal form which, derived this way, is again a unique representation of one of the 2^{2^n} logical functions. This form will be used in the synthesis of networks.

As an example of this procedure take the two-variable logical function expressing the equiva-

lence of A and B. Let X_1 and X_2 be A and B respectively. The four Schroeder constituents are with Z from 0 to 3:

$$D_0 = (\bar{A} \text{ v } \bar{B}) ,$$
$$D_1 = (A \text{ v } \bar{B}) ,$$
$$D_2 = (\bar{A} \text{ v } B) ,$$
$$D_3 = (A \text{ v } B) .$$

Since

$$A \rightarrow B = D_1 \cdot D_2 ,$$

we have

$$P = 0 \cdot 2^0 + 1 \cdot 2^1 + 1 \cdot 2^2 + 0 \cdot 2^3 = 6$$

and

$$K_6 = D_1 \cdot D_2 = (A \text{ v } \bar{B}) \cdot (\bar{A} \text{ v } B) .$$

Expanding the right hand side gives

$$(A \cdot \bar{A} \text{ v } B \cdot A) \text{ v } (\bar{A} \cdot \bar{B} \text{ v } B \cdot \bar{B}) ,$$

which after cancellation of contradictions yields the desired expression in Hilbert's disjunctive normal form:

$$(B \cdot A) \text{ v } (\bar{A} \cdot \bar{B}) = A \rightarrow B .$$

The second step considers the synaptic delay Δt at each McCulloch element. Let $N_i(t)$ denote the action performed by the ith element at time t, or for short

$$N_i = N_i(t) . \qquad (61)$$

In order to facilitate expressions that consider n synaptic delays earlier, a recursive operator S is introduced and defined as

$$N_i(t - n\Delta t) \equiv S^n N_i , \qquad (62)$$

its iteration represented by its power. Clearly, this operator is applicable to propositions as well.

The third step establishes distributivity of the operator S with respect to conjunction and disjunction:

$$S(N_i \cdot N_j) = SN_i \cdot SN_j ,$$
$$S(N_i \text{ v } N_j) = SN_i \text{ v } SN_j . \qquad (63)$$

Since each function of temporal propositions can be expressed in terms of Hilbert's disjunctive normal form, application of the recursive operator S permits each proposition to appear in the form $S^k X_i$ and thus can be translated into the corresponding neural expression (62), which localizes each element in the network and defines its function.

We shall illustrate this procedure with the

same simple example that was chosen by the
authors of this theorem. It is known as the "illu-
sion of heat and cold".

"If a cold object is held to the skin for a mo-
ment and removed, a sensation of heat will be
felt; If it is applied for a longer time, the sensa-
tion will be only of cold, with no preliminary
warmth, however transient. It is known that one
cutaneous receptor is affected by heat and an-
other by cold".

We may now denote by N_1 and N_2 the proposi-
tions "heat is applied" and "cold is applied" re-
spectively, but interchangeably we may denote
by N_1 and N_2 the activity of the receptors "heat
receptor active" and "cold receptor active".
Similarly, we shall denote by N_3 and N_4 the
propositions "heat is felt" and "cold is felt" re-
spectively which can be translated into the ac-
tivity of the elements producing the appropriate
sensations *mutatis mutandis*.

The temporal propositional expression for the
two observations can now be written:

Input	Output
$SN_1 \vee S^3 N_2 \cdot S^2 \bar{N}_2$	N_3
$S^2 N_2 \cdot SN_2$	N_4

where the required persistence in the sensation
of cold (N_4) is assumed to be two synaptic de-
lays, while only one delay is required for sensa-
tion of heat (N_3).

We utilize distributivity of S

$$S(N_1 \vee S(SN_2 \cdot \bar{N}_2)) = N_3 ,$$

$$S((SN_2) \cdot N_2) = N_4 ,$$

and develop the whole net in individual steps of
nets for two variables, working our way from

inside out of the brackets. We first approach the
expression for N_4 and construct a net for

$$SN_2 = N_A . \qquad \text{(fig. 20.1)}$$

We complete the relation for N_4 by drawing the
net (bold line):

$$N_4 = S(N_A \cdot N_2) , \qquad \text{(fig. 20.2)}$$

which is, of course,

$$N_4 = S(SN_2 \cdot N_2) .$$

We approach now the expression for N_3 and draw
(bold line):

$$S(N_A \cdot \bar{N}_2) = N_B , \qquad \text{(fig. 20.3)}$$

and complete the whole net by drawing (bold
line):

$$N_3 = S(N_1 \vee N_B) . \qquad \text{(fig. 20.4)}$$

Although this simple example does not do jus-
tice to the profoundness of the McCulloch-Pitts
theorem, it emphasizes not only the important
relationship between formal networks and formal
logic but also the minimal structural necessity
to accommodate functional requirements.

4.3. Interaction networks of discrete, linear elements

We now turn our attention to networks which
are composed of "linear elements", i.e., of Mc-
Culloch elements operating with low thresholds
($\theta \approx 1$) and weak signals ($f_i \ll 1/(\Delta t + \Delta t_R)$) in an
asynchronous network, or of Sherrington ele-
ments ($\mu \to 0$) which perform algebraic summa-
tion on their inputs.

The formalism which handles the situation of
an arbitrary number of interacting elements has
completely been worked out by Hartline (1959)
who showed in a series of brilliant experiments
the mutual inhibitory action of proximate fibers
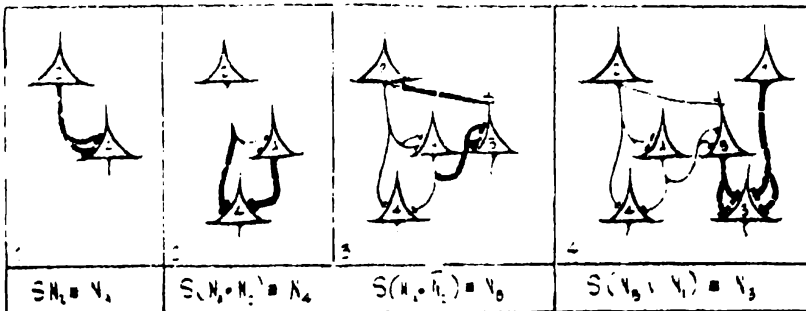
Fig. 20. Stepwise development of a McCulloch-Pitts network that computes the "illusion of heat and cold". Bold lines
represent added network elements ($\theta = 2$, everywhere).

in the optic stalk of the horseshoe crab by illuminating various neighbors of a particular ommatidium in the crab's compound eye.

Consider n linear elements e_i, each of which is actively connected to all others and to itself. We have a perfect connection matrix, all rows and columns being non-zero. Let $\rho(i)$ and $\sigma(i)$ represent response and external stimulus of element e_i respectively and permit a certain fraction a_{ij} of the response of element e_i to contribute to the stimulus of element e_j. The response of element e_1 is under these circumstances clearly

$$\rho(1) = \sigma(1) + a_{11}\rho(1) + a_{21}(2) + \ldots + a_{n1}\rho(n) , \quad (64)$$

or in general for the jth element:

$$\rho(j) = \sigma(j) + \sum_{i=1}^{n} a_{ij}\rho(i) , \quad (65)$$

where for simplicity σ and ρ are expressed in the same arbitrary units and where the coefficients a_{ij} form again a square matrix which we will call the numerical interaction matrix

$$A_n = \|a_{ij}\|_n . \quad (66)$$

In order to obtain a solution for the n unknowns $\rho(j)$ in terms of all stimuli $\sigma(1), \sigma(2), \ldots, \sigma(n)$, we first express all stimuli $\sigma(j)$ in terms of the various responses $\rho(1), \rho(2), \ldots, \rho(n)$. From (64) we have for $\sigma(1)$

$$\sigma(1) = (1 - a_{11})\rho(1) - a_{21}\rho(2) - \sigma_{31}\rho(3) \ldots$$

or in general for the jth element

$$\sigma(j) = \sum_{i=1}^{n} s_{ij}\rho(i) , \quad (67)$$

where the s_{ij} form again a square matrix S_n which we will call the stimulus matrix

$$S_n = \|s_{ij}\|_n \quad (68)$$

with

$$s_{ij} = \begin{cases} -a_{ij} & \text{for } i \neq j , \\ (1 - a_{ii}) & \text{for } i = j . \end{cases} \quad (69)$$

In the formalism of matrix algebra the n values for σ as well as for ρ represent n-dimensional vectors (column matrices) and (67) can be formally represented by:

$$\sigma_n = S_n\rho_n . \quad (70)$$

In order to find ρ_n expressed in terms of σ_n one "simply" inverts the matrix S_n and obtains

$$\rho_n = S_n^{-1}\sigma_n . \quad (71)$$

which implies solving the n equations in (67) for the n unknowns $\rho(1), \rho(2), \ldots, \rho(n)$. We introduce the response matrix R_n, defined by

$$R_n = \|r_{ij}\|_n = S_n^{-1} . \quad (72)$$

Let $|D|$ denote the characteristic determinant $|s_{ij}|$, and S_{ij} the product of $(-1)^{i+j}$ with the determinant obtained from $|s_{ij}|$ by striking out the ith row and jth column, then the response matrix elements r_{ij} are given by

$$r_{ij} = \frac{S_{ji}}{D} , \quad (73)$$

and we have the solution for the responses:

$$\rho_n = R_n\sigma_n . \quad (74)$$

Clearly, a solution for ρ_n can be obtained only if the characteristic determinant D does not vanish, otherwise all responses approach infinity, which implies that the system of interacting elements is unstable. It is important to note that stability is by no means guaranteed if the interactions are inhibitory, for the inhibition of an inhibition is, of course, facilitation.

The actual calculation of a response matrix, given the numerical interaction matrix, is an extremely cumbersome procedure that requires the calculation of n^2 matrices, each of which demands the calculation of $n!$ products consisting of n factors each, that is $n^3 \cdot n!$ operations all together. Under these circumstances it is clear that manual computation can be carried out for only the most simple cases, while slightly more sophisticated situations must be handled by high speed digital computers; even they prove insufficient if the number of elements goes beyond about, say, 50. However, the horseshoe crab performs these operations in a couple of milliseconds by simultaneous parallel computation in the fibers of the optic tract. The *Limulus*' eye is — so to say — made for matrix inversion.

In order to clarify procedures we give a simple example of four elements e_1, e_2, e_3, e_4 which are thought to be placed at the four corners of a square labeled clockwise with e_1 in the NW corner. We assume mutual interaction to take place between neighbors only, and all coefficients to be alike at a. The connection matrix for this configuration is

		e_j			
		1	2	3	4
	1	0	1	0	1
e_i	2	1	0	1	0
	3	0	1	0	1
	4	1	0	1	0

49

and the numerical interaction matrix

$$A_4 = \begin{Vmatrix} 0 & a & 0 & a \\ a & 0 & a & 0 \\ 0 & a & 0 & a \\ a & 0 & a & 0 \end{Vmatrix} .$$

From this we obtain with eq. (69) the stimulus matrix

$$S_4 = \begin{Vmatrix} 1 & -a & 0 & -a \\ -a & 1 & -a & 0 \\ 0 & -a & 1 & -a \\ -a & 0 & -a & 1 \end{Vmatrix} ,$$

which inverted gives the following response matrix:

$$R_4 = \frac{1}{D} \begin{Vmatrix} 1-2a^2 & a & 2a^2 & a \\ a & 1-2a^2 & a & 2a^2 \\ 2a^2 & a & 1-2a^2 & a \\ a & 2a^2 & a & 1-2a^2 \end{Vmatrix} .$$

The characteristic determinant is

$$D = 1 - 4a^2$$

with the two roots $a = \pm\frac{1}{2}$. Hence, the system becomes unstable, whenever the interaction coefficient a approaches $+\frac{1}{2}$ (facilitation) or $-\frac{1}{2}$ (inhibition).

We are now in a position to write all responses $\rho(j)$ in terms of the stimuli $\sigma(i)$. For the response of the first element we have

$$D \cdot \rho(1) = (1-2a^2)\sigma_1 + a\sigma_2 + 2a^2\sigma_3 + a\sigma_4 ,$$

the others are obtained by cyclic rotation of indices or directly from R_n.

For uniform stimulation of all elements, $\sigma(i) = \sigma_0$ for all i, the uniform response is

$$\rho_0 = \frac{1}{1-2a}\sigma_0 ,$$

which clearly depends upon the sign of the interaction coefficient, giving increased or decreased responses for facilitation or inhibition respectively.

If uniform stimulation is maintained for e_2, e_3, e_4 ($\sigma_2 = \sigma_3 = \sigma_4 = \sigma_0$), but element e_1 is stimulated by a ($\pm$) superposition of σ^* ($\sigma_1 = \sigma_0 + \sigma^*$), and all stimuli and responses are expressed in terms of uniform stimulus and response we have

$$\frac{\rho(i)}{\rho_0} = 1 + b_i \frac{\sigma^*}{\sigma_0}$$

$$b_1 = \frac{1-2a^2}{1+2a} , \qquad b_2 = b_4 = \frac{a}{1+2a} , \qquad b_3 = \frac{2a^2}{1-2a} .$$

The quadratic terms for a in the numerator of b_1 and b_3 show clearly the effect of "double negation" by "inhibition of inhibition", for these terms are independent of the sign of a.

As a final example we show the far-reaching influence of a local perturbation in a mixed net which consists of a linear array of quadrupoles as above with an inhibitory interaction coefficient of $a = -0.3$, and where each quadrupole is actively connected to its neighbor on one side, but passively connected to its neighbor on the other side (see fig. 21). Elements which actively connect quadrupoles transmit their full response to their neighbors. A unit stimulus is applied to element e_1 in the first quadrupole only. The resulting response is plotted next to each element, the length of bars representing intensity.

This example may again be taken as an instance of the principle "action by contagion"
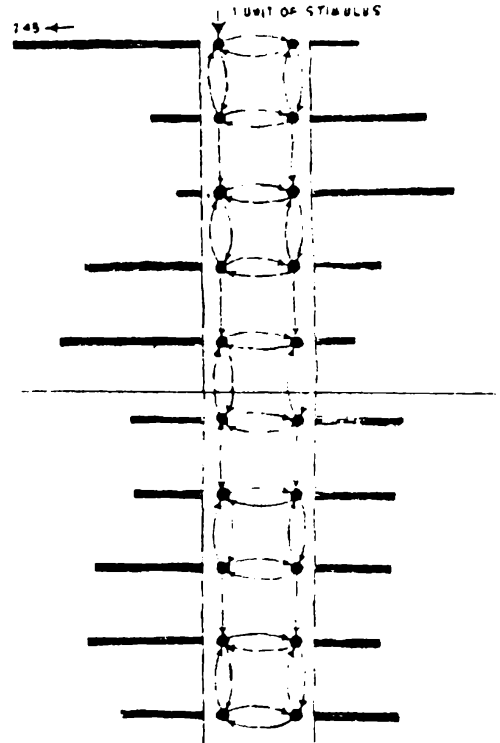
Fig. 21. Responses in a mixed action-interaction network after a single stimulus is applied to the NW element in the first interacting quadrupole.

which we met earlier in a mixed interaction net (sea-urchin). In this case, however, the local perturbation spreads in the form of a decaying oscillation.

The discussion of interaction in nets composed of discrete linear elements has shown thus far two serious deficiencies. The first deficiency is clearly the insurmountable difficulty in handling efficiently even simple net configurations. We shall see later that this difficulty can be circumvented at once, if the individuality of elements is dropped and only the activity of elements associated with an infinitesimal region in space is taken into consideration. The powerful apparatus developed in the theory of integral equations will take most of the burden in establishing the response function, given a stimulus function and an interaction function.

The second deficiency becomes obvious if we have to answer the question of where we localize the operation that transmits "a certain fraction a_{ij}" of the activity of element e_i to element e_j. Is this a property of the transmitting or of the receiving element? Clearly, our simple model of elements does not yet take care of this possibility. However, at this stage of the development it is irrelevant to decide whether we make transmitter or receiver responsible for the regulation of the amount of the transmitted agent (see fig. 22a). However, it is necessary to bestow on at least one of them the capacity to regulate the transmitted agent. We decide to make the receiver responsible for this operation, justifying this decision by the possibility of interpreting this regulatory operation — at least for neural synaptic contacts — as the number of facilitatory or inhibitory synaptic junctions of a fiber that synapses element e_i onto e_j. Hence, for our present purposes we adopt a representation of our linear elements (fig. 22b) which modifies the transmitted signal at their inputs according to the numerical value of the active connection coefficient a_{ij}. Later, however, we shall see that both, transmitter and receiver define this coefficient.

4.4. *Active networks of discrete, linear elements*

We draw directly from our definitions for action nets and for the various operational modalities of elements as discussed in earlier paragraphs; however, we shall introduce in this paragraph for the first time some constraints on the spatial distribution of elements. These constraints will be tightened considerably while we proceed.

4.4.1. Linear elements

Consider a set of $n = 2m$ linear elements, half of which are general receptors and the other half general effectors. A weak geometrical constraint, which does not affect the generality of some of the following theorems but facilitates description, is to assume spatial separation of receptors and effectors. The locus of all general receptors, e_{1i}, we shall call "transmitting layer" L_1, and the locus of all general effectors, e_{2j}, $(i, j = 1, 2, 3, \ldots, m)$ the "collecting layer" L_2, regardless of the dimensionality of these loci, i.e., whether these elements are arranged in a one-dimensional array, on a two-dimensional surface, or in a specifiable volume.

We consider the fraction of activity in element e_{1i} that is passed on to an element e_{2i} as its partial stimulus a_{ij} (i). The action coefficients for the m^2 pairs define the numerical action matrix

$$A_m = \| a_{ij} \|_m . \qquad (75)$$

With linear elements in the collector layer their responses are the algebraic sum of their partial stimuli:

$$\rho(j) = \sum_i^m a_{ij} \, \sigma(i) , \qquad (76)$$

or in a matrix notation

$$\rho_m = A_m \, \sigma_m . \qquad (77)$$

Since this result is in complete analogy to interaction nets (eq. (74)) where the response matrix R_n establishes the stimulus-response relationship, we have the following theorem:

Any stable interaction network composed of m elements can be represented by a functionally equivalent action network composed of $2m$ elements (see fig. 23):

$$\| r_{ij} \|_m = \| a_{ij} \|_m .$$

This result is of significance insofar as it shows that two entirely different structures have precisely the same stimulus-response characteristic. For example, Hartline's observation of inhibitory *interaction* amongst the fibers in the optic stalk of the horseshoe crab can be explained equally well by an appropriate post-ommatidial *action* net. It is only the anatomical evidence of the absence of such nets which forces us to assume that interaction processes are responsible for the observed phenomena.

The converse of the above theorem "for each action net there exists a functionally equivalent interaction net" is true only if the characteristic determinant of the inverse of A_m does not vanish.
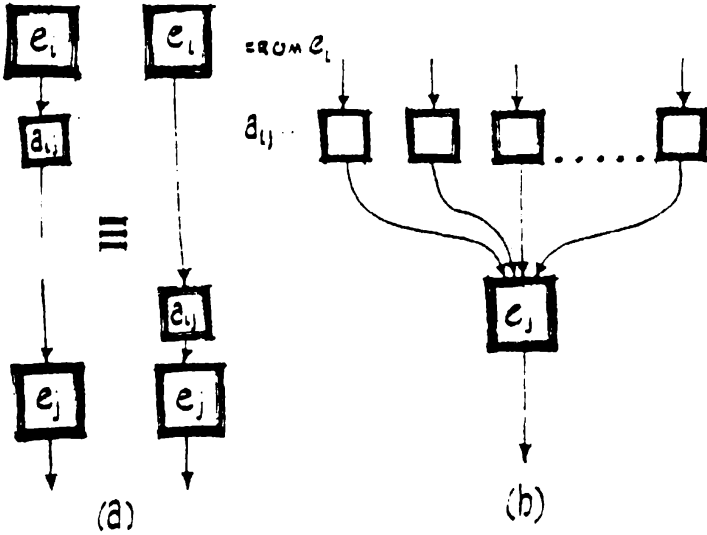
51

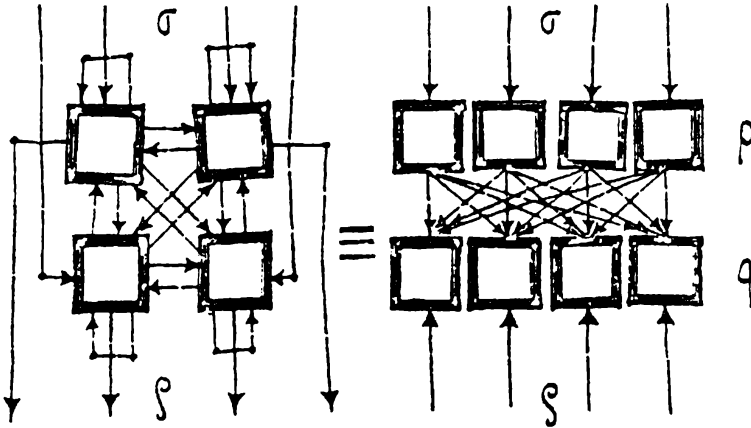Fig. 22. Formal equivalence of localization of operations.

Fig. 23. Equivalence of action network with interaction network.

We consider $k+1$ cascaded layers L_i ($i = 0, 1, \ldots, k$) with the transmitting layer L_0 the locus of receptors proper, and with all elements in layer L_{i-1} acting upon all elements in layer L_i, their actions defined by an action matrix A_{mi}. The action performed by the receptors e_{0j} on the ultimate effectors e_{k1} is again (see eq. (3)) defined by the matrix product of all A_{mi}

$$\text{Cas}(A_{m1}, A_{m2}, \ldots, A_{mk}) = \prod_1^k A_{mi}. \quad (78)$$

Hence, we have the following theorem:

Any cascaded network of a finite number of layers, each acting upon its follower with an arbitrary action matrix can be replaced by a functionally equivalent single action net with an action matrix

$$\| a_{ij}^{(k)} \|_m = \prod_{l=1}^k \| a_{ij}^{(l)} \|_m. \quad (79)$$

Again, gross structural differences may lead to indistinguishable performances.

We generalize our observation in Chapter 1, eq. (8) concerning the invariance of certain action nets to cascading. An action matrix with all rows alike

$$a^\bullet_{ij} = a^\bullet_{kj} ,$$

and for which

$$\sum_{j=1}^{m} a^\bullet_{ij} = 1$$

is invariant to being cascaded. Let

$$A^\bullet_m = \|a_{ij}\|_m ,$$

we have

$$[A^\bullet_m]^k \, A^\bullet_m . \tag{80}$$

4.5. Action networks of discrete, localized elements

After these general remarks we introduce more stringent geometrical constraints. We now assume that elements e_{hi} are not only traceable to a certain layer L_h, but also that each element within one layer can be localized as to its precise position in this layer (see fig. 24). We first consider only action phenomena between elements of two adjacent layers L_i and L_{i+1}, which, for simplicity, we may call L_p and L_q. We subdivide each layer into lattice elements with appropriate dimensions Δx, Δy, Δz, and $\Delta \xi$, $\Delta \eta$, $\Delta \zeta$ in L_p and L_q respectively, so as to be able

to accommodate in each lattice element precisely one element e_{pi} and e_{qj} respectively. For simplicity we shall assume the corresponding lattice dimensions in both layers to be alike $\Delta x = \Delta \xi$, etc., because first, it does not infringe on the generality of the remarks we wish to make and, if there is indication to the contrary, it is not difficult to match the metric of the two layers L_p and L_q by appropriate transformations.

We are now in a position to label each element in layers L_p and L_q according to the coordinates of the lattice element in which it resides. Let p and q represent the coordinate triple that locates the lattice elements in the respective layers. We have:

$$\begin{aligned} p[x, y, z] &= [x \cdot \Delta x, y \cdot \Delta y, z \cdot \Delta z] , \\ q[u, v, w] &= [u \cdot \Delta \xi, v \cdot \Delta \eta, w \cdot \Delta \zeta] , \end{aligned} \tag{81}$$

where x, y, z, u, v, w are integers $0, \pm 1, \pm 1, \ldots$. It is clear that for one-dimensional or two-dimensional layers the definitions for p and q boil down to $p[x]$, $q[u]$, and $p[x, y]$, $q[u, v]$ respectively. And it is also clear that we may now drop the second index in the labeling of elements e_{pi} and e_{pj}, for e_p and e_q suffices to identify each element, since p and q define the locus of its position.

Again we wish to express the action exerted by element e_p to element e_q. To this end we define an "action function" $K(p, q)$ which specifies
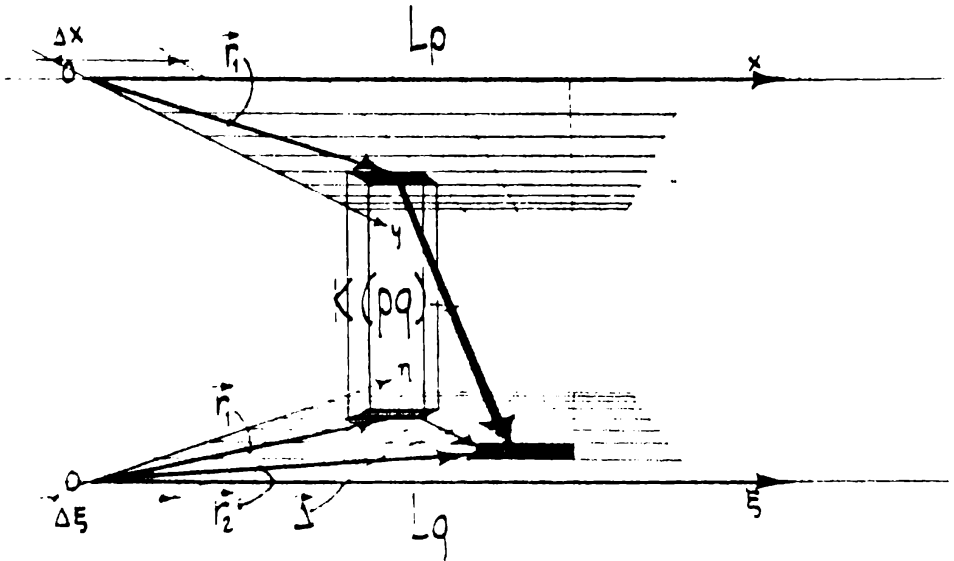
Fig. 24 Geometry in an action network.

for each pair $\{e_p, e_q\}$ the fraction of activity in e_p that is transmitted to e_q. Of course, this action function may again be represented by an action matrix K_m. However, with our knowledge of the, position of each element in both layers we may be able to associate the transmitted amount of activity with certain geometrical relationships which exist between elements of the two layers. In other words, the m^2 entries k_{pq} in the action matrix K_m may be all considered to be functions of the loci of the elements with which these entries are associated:

$$k_{pq} = k_{pq}(x, y, z; u, v, w) . \qquad (82)$$

The assumed dependency of the transmitted activity on geometrical relationships justifies the term action *function*. On the other hand, this is precisely the kind of relationship which is alluded to, if reference is made, say, to "cortical organization" or "organization of neural interaction". It is, to a certain extent, the genetic program that produces anatomical – read "geometrical" – constraints which prohibit, within certain limits, arbitrary developments of conceivable structures. The noteworthy feature of eq. (82) is that it links activity with geometry, in other words, function with structure.

Written out explicitly in terms of stimulus and response, the action function appears under the triple sum taken over all elements in the transmission layer L_p:

$$\rho(u, v, w) = \sum_x \sum_y \sum_z K(xyz, uvw)\, \sigma(xyz) . \qquad (83)$$

A discussion of the general properties of $K(p, q)$ goes beyond the scope of this article (Inselberg and Von Foerster, 1962; Von Foerster, 1962; Taylor, 1962). However, there is no need to go to extremes if the simple assumptions about prevailing geometry will suffice to show the significance of these concepts. Consequently, we are going to introduce further geometrical constraints.

Periodicity in structure is, as was suggested earlier, a ubiquitous feature in organic nets. We define a periodic action function with orders x_0, y_0, z_0 to be an action function which is invariant to translations of whole multiples of these periods:

$$K(x + ix_0;\ y + jy_0;\ z + kz_0;\ u + ix_0;\ v + jx_0;\ w + kz_0)$$

$$= K(xyz, uvw) . \qquad (84)$$

$$i, j, k = 0, \pm 1, \pm 2, \dots .$$

Clearly, a network with such a periodic action function produces outputs in L_q that are invariant to any stimulus distribution which is translated with same periodicity x_0, y_0, z_0.

In order to have response invariance to stimulus translation *everywhere* along the receptor set L_q, we must have:

$$x_0 = \Delta x , \qquad y_0 = \Delta y , \qquad z_0 = \Delta z . \qquad (85)$$

From this we may draw several interesting conclusions. First, the action functions so generated are independent of position, for the smallest interval over which they can be shifted is precisely the order of their period. Second, under the conditions of translatory invariance the action functions reduce to sole functions of the difference of the coordinates which localize the two connected elements in their respective layers:

$$K(p, q) = k(\Delta) , \qquad (86)$$

where Δ is a vector with components

$$\Delta = [(x - \xi), (y - \eta), (z - \zeta)] . \qquad (87)$$

We introduce symmetric, anti-symmetric and spherically symmetric action functions which have the following properties respectively:

$$K_s(-\Delta) = K_s(\Delta) . \qquad (88)$$

$$K_a(-\Delta) = -K_a(\Delta) , \qquad (89)$$

$$K_r(\Delta) = K_r(\Delta) . \qquad (90)$$

It is easy to imagine the kind of abstractions these action functions perform on the set of all stimuli which are presented to the receptor set in L_p, if we assume for a moment that both layers, L_p and L_q, are planes. Clearly, in all cases the responses in the effect of set are invariant to all translations of any stimulus distribution ("pattern") in the receptor set. Moreover, K_s gives invariance to reversals of stimuli symmetric to axes $y = 0$ and $x = 0$ (e.g., 3 into ϵ, or M into W), while K_a gives invariance to reversals of stimuli symmetric to $y = \pm x$ (e.g., $\sim$ into S; and $>$ into V). Finally, action function K_r gives invariance to all stimulus rotations as well as translations (i.e., some reversals as above plus, e.g., N into Z). The planes of symmetry in three dimensions which correspond to the lines in two dimensions are clearly the three planes defined by the axes xy, yz, zx, in the first case, and, in the second case, the three planes defined by the six origin-centered diagonals that cut through the three pairs of opposite squares in the unit cube.

Although for analytic purposes the action function has desirable properties, from an experimental point of view it is by no means con-

venient. In order to establish in an actual case the action function of, say, element e_p^* in the receptor set, it is necessary to keep this element stimulated while searching with a microprobe though all fibers of a higher nucleus to pick those that are activated by e_p^*. Since this is obviously an almost impossible task, the procedure is usually reversed. One enters a particular fiber e_q^* of a higher nucleus, and establishes, by stimulation of elements e_p, which one of these activates e_q^*. In this way it is the receptor field which is established, rather than the action field. However, considering the geometrical constraints so far introduced, it is easy to see that, with the exception of the anti-symmetric action function, action function $K(p,q)$ and "receptor function" $G(q,p)$ are identical:

$$G_s = K_s ,$$
$$G_r = K_r ,$$
$$G_a = -K_a . \qquad (91)$$

For more relaxed geometrical constraints the expressions relating receptor function and action function may be more complex, but are always easy to establish.

It may have been noticed that a variety of structural properties of networks have been discussed without any reference to a particular action function. Although the actual computational labor involved to obtain stimulus-response relationships in action nets is far less than in inter-

action nets, the machinery is still clumsy if nets are of appreciable sophistication.

Instead of demonstrating this clumsiness in some examples, we postpone the discussion of such nets. In the next paragraph the appropriate mathematical apparatus to bypass this clumsiness will be developed. Presently, however, we will pick an extremely simple action net, and explore to a full extent the conceptual machinery so far presented with the inclusion of various examples of operation modalities of the network's constituents.

Example: Binomial action function

Fig. 25a represents our choice. It is a one-dimensional periodic action net with unit periodicity, its predominant feature being lateral inhibition. The universal action function and receptor function of this network are quickly found by inspection and are drawn in fig. 25b and c respectively. Obviously, these functions are symmetric, hence

$$G_1^*(q,p) = K_1^*(p,q) = K_1^*(x-u) = K_1^*(\Delta)$$

$$= (-1)^\Delta \binom{2}{1+\Delta} ,$$

i.e.,

$$K_1^*(\Delta) = \begin{cases} -1, & \Delta = -1, \\ -2, & \Delta = 0, \\ -1, & \Delta = +1, \end{cases}$$

everywhere else $K_1^* = 0$.

(a)

$K(pq)$ $G(qp)$ $\theta \cdot 2$

(b)   (c)   (d)   (e)
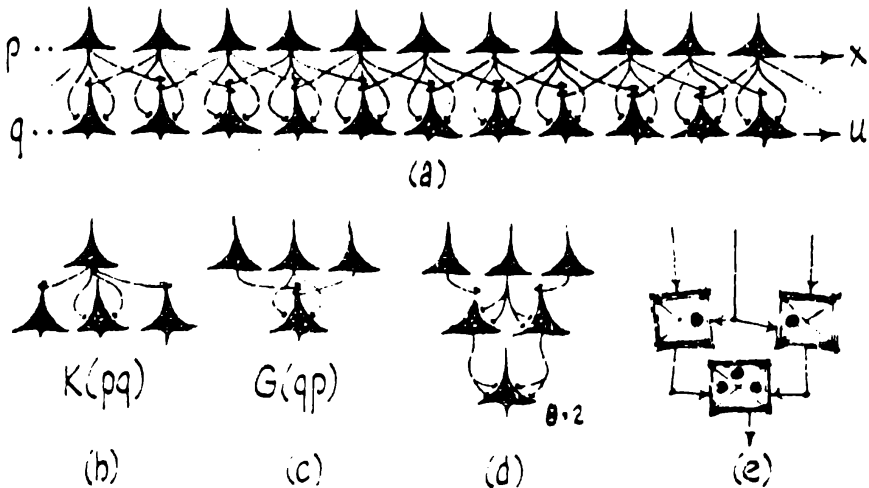
Fig. 25 One dimensional periodic action net. Network (a); action function (b); receptor function (c); equivalent two-input element network (d); symbolic representation (e).

The equivalent net with McCulloch elements having only two inputs is given in fig. 25d. In fig. 25e its equivalent is symbolized in purely logical terms. The index 1 in $G_1^\bullet$ and $K_1^\bullet$ is adopted to discriminate this receptor and action function from a general class of such functions, $K_m^\bullet$, the so-called mth order binomial action function:

$$K_m^\bullet = (-1)^\Delta \begin{pmatrix} 2m \\ m+\Delta \end{pmatrix}, \qquad (92)$$

$$\Delta = 0, \pm 1, \pm 2, \ldots, \pm m .$$

These arise from cascading binomial action nets m times, as suggested in fig. 26:

$$K_2^\bullet = (-1)^\Delta \begin{pmatrix} 4 \\ 2+\Delta \end{pmatrix} = \begin{cases} +1, & \Delta = -2, \\ -4, & \Delta = -1, \\ +6, & \Delta = 0, \\ -4, & \Delta = +1, \\ +1, & \Delta = +2. \end{cases}$$

(i) Sherrington element

In order to obtain stimulus response relationship in the network of fig. 25 we have to specify the operational modality of the elements. First we assume a strictly linear model (Sherrington element). Let $\sigma(x)$ and $\rho(u)$ be stimulus and response at points x and u respectively. We have

$$\rho(u) = 2\sigma(x) - \sigma(x - \Delta x) - \sigma(x + \Delta x) .$$

Expanding σ around x we obtain:

$$\sigma(x \pm \Delta x) = \sigma(x) \pm \frac{\partial \sigma}{\partial x} \Delta x + \frac{1}{2} \frac{\partial^2 \sigma}{\partial x^2} \Delta^2 x \\ \pm \frac{1}{6} \frac{\partial^3 \sigma}{\partial x^3} \Delta^3 x + \ldots ,$$

which, inserted above, gives

$$\rho(u) = \frac{\partial^2 \sigma}{\partial x^2} \Delta^2 x + \frac{1}{12} \frac{\partial^4 \sigma}{\partial x^4} \Delta^4 x + \ldots .$$

Neglecting fourth order and higher terms, this lateral inhibition net extracts everywhere the second derivative of the stimulus distribution. It can easily be shown that the mth binomial action functions will extract the $2m$th derivative of the stimulus. In other words, for uniform stimulus, strong or weak, stationary or oscillating, these nets will not respond. However, this could have been seen by the structure of their binomial action function, since

$$\sum_{-m}^{+m} (-1)^\Delta \begin{pmatrix} 2m \\ m+\Delta \end{pmatrix} = 0 .$$

(ii) McCulloch element; asynchronism

We change the *modus operandi* of our elements, adopt a McCulloch element with unit threshold ($\theta = 1$), and operate the net asynchronously. We ask for the output frequency of each effector element, given the stimulus distribution. For simplicity, we write our equations in terms of the ON probability p of the elements. With numbers 1, 2, 3, we label the afferent fibers in the receptor field (see fig. 25c). The truth table is easily established, giving an output ON for input states (010), (011) and (110) only. The surviving Bernoulli products (eq. (25)) are:

$$p = (1 - p_1)p_2(1 - p_3) + p_1 p_2 (1 - p_3) + (1 - p_1)p_2 p_3 ,$$

which yields:

$$p = p_2 - p_1 p_2 p_3 .$$

Uniform stimulation, i.e., $p_1 = p_2 = p_3 = p_0$, pro-
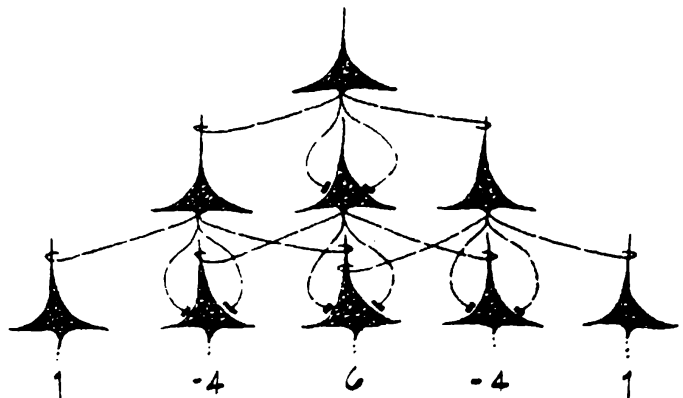
Fig. 26. Cascading of binomial action function of mth function of ($m+1$)th order.

1   -4   6   -4   1
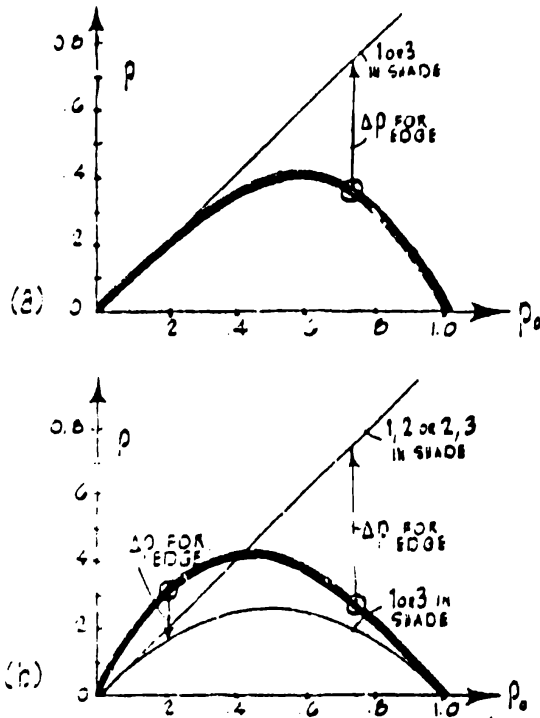
Fig. 27. Transfer function of a McCulloch formal neuron net when operated asynchronously. (a) lateral inhibiting network. (b) lateral facilitating network.

duces a "leakage" frequency

$$p_{uni} = p_0 - p_0^3 ,$$

which disappears for strong stimulation (see fig. 27a bold line). With either element (1) or (3) OFF, i.e., with an "edge" in the stimulus field, we have

$$p_{edg} = p_0 ,$$

the difference between these frequencies Δp is, of course, an indication of detection sensitivity. Inspection of fig. 27a shows that this element is a poor edge detector in the dark, but does very well in bright light.

It might be worthwhile to note that a reversed action function (1; -2; 1) with lateral facilitation has two operational modes, one of them with considerable sensitivity for low intensities (fig. 27b). A slight nystagmus with an amplitude of one element switches between these two modes, and thus represents an edge detector superior to the one with lateral inhibition.

(iii) McCulloch element; synchronism

We finally change the *modus operandi* of our net to synchronous operation with McCulloch element ($\theta = 1$). Clearly, for uniform stimulus distribution, strong or weak, steady or flicker, all effectors will be silent; the net shows no response. Nevertheless, an edge will be readily observed.

However, a net incorporating much simpler elements will suffice. A McCulloch element with only two inputs computing the logical function "either A or B" (see function No. 6, table 1) clearly computes an "edge". This function represents again a symmetric action function. Asymmetry may be introduced by choosing a McCulloch element that computes, say, function No. 2: "A only". A net incorporating this element in layer L_1 is given in fig. 28a. The result is, of course, the detection of an asymmetric stimulus property, the presence of a "right hand edge". Hence, in order to detect directionality in the stimulus field the net must mirror this directionality in the connectivity of its structure or in the operation of its elements. Utilizing synaptic delays that occur in layer L_1 we have attached a second layer L_2 that computes in D the function "C only". Consequently, layer L_2 detects right edges moving to the right. While C computes the presence of a right hand edge D will be silent, because the presence of a right edge implies a stimulated B which, simultaneous with an active C, gives an inactive D. Similarly, a left edge will leave D inactive; but C is inactive during the presence of a left edge. However, D will be active at once if we move the right hand edge of an obstruction to the right. Under these circumstances the synaptic delay in C will cause C to report still a right hand edge to D, while B is already without excitation. Of course, movements to the left remain unnoticed by this net. The equivalent net using the appropriate McCulloch formal neurons is given in fig. 28b.

Thanks to the remarkable advances in experimental neurophysiology, in numerous cases the existence of abstracting cascaded action networks in sensory pathways has been demonstrated. In their now classic paper "What the frog's eye tells the frog's brain" Letvvin et al. (1959) summed up their findings: "The output from the retina of the frog is a set of four distributed operations on the visual image. These operations are independent of the level of general illumination and express the image in terms of: (1) local sharp edges and contrast; (2) the curvature of edge of a dark object; (3) the movement of edges; and (4) the local dimmings produced by
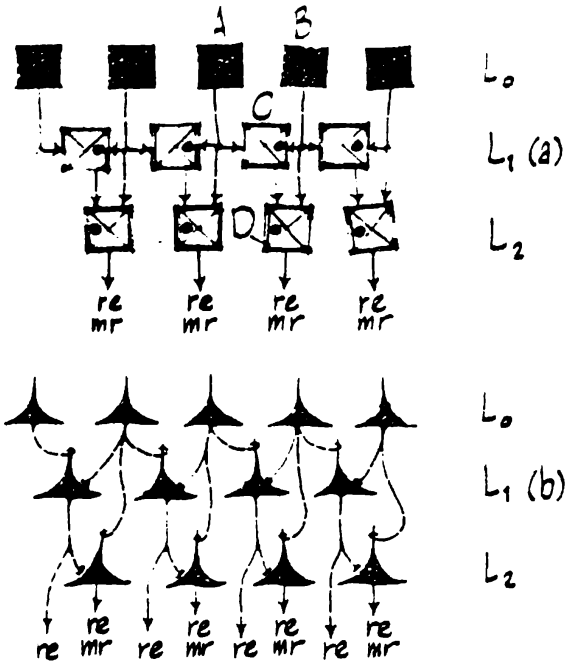
Fig. 28. Net detecting right hand edges, and right hand edges moving to the right (a). Formal presentation. (b) Simplest equivalent neural net.

movement or rapid general darkening".

To these properties Maturana (1962) adds a few more in the eye of the pigeon. Here anti-symmetric action functions produce strong directionalities. There are fibers that report horizontal edges, but only when they move. A vertical edge is detected whether it is moving or not. A careful analysis of the receptor function $G(q,p)$ in the visual system in cats and monkeys has been carried out by Hubel (1962); Mountcastle et al. (1962) explored the complicated transformations of multilayer mixed action-interaction networks as they occur at the thalamic relay nucleus with the somatic system as input.

3.7. Action networks of cell assemblies

Sholl (1956) estimates the mean density $\bar{v}$ of neurons in the human cortex at about 10^7 neurons per cubic centimeter, although this number may vary considerably from region to region. This density implies a mean distance $\bar{l}$ from neuron to neuron.

$$\bar{l} = (\bar{v})^{-\frac{1}{3}} = 5 \times 10^{-3} \text{ cm}$$

or approximately 50 μ. This distance corre-

sponds, of course, to the lattice constants Δx, Δy, Δz, etc. In the previous paragraph and gives the elementary lattice cell a volume of $\Delta x \Delta y \Delta z \approx 10^{-7}$ cm^3. Even with the best equipment available today we cannot reproducibly attain a given point in the brain within this range. Consequently, our cell by cell approach describes a highly idealistic situation, and the question arises as to whether or not some of the earlier concepts can be saved if we wish to apply them to a much more realistic situation.

Let us assume optimistically that one is able to locate a certain point in the living cortex within, say 0.5 mm = 500 μ. This defines a volume of uncertainty which contains approximately a thousand neurons. This number, on the other hand, is large enough to give negligible fluctuations in the total activity, so we are justified in translating our previous concepts, which apply to individual elements e_i, e_j as, e.g., stimulus $\sigma(i)$, response $\sigma(j)$, into a formalism that permits us to deal with assemblies of elements rather than with individuals. Moreover, as long as these elements connect with other elements over a distance appreciably larger than the uncertainty of its determination, and there is a considerable fraction of cortical neurons fulfilling this condition, we are still able to utilize the geometrical concepts as before. To this end we drop the cellular individuality and refer only to the activity of cell assemblies localizable within a certain volume.

In analogy to the concept of "number density" of neurons, i.e., the number of neurons per unit volume at a certain point (xyz) in the brain, we define "stimulus density" $\sigma(xyz)$ in terms of activity per unit volume as the total activity S measured in a certain volume, when this volume shrinks around the point (xyz) to "arbitrary" small dimensions:

$$\sigma(xyz) = \lim_{V \to 0} \frac{S}{V} = \frac{dS}{dV}. \tag{93}$$

Similarly, we have for the response density at (uvw):

$$\rho(uvw) = \lim_{V \to 0} \frac{R}{V} = \frac{dR}{dV}, \tag{94}$$

if R stands for the total response activity in a macroscopic region.

We wish to express the action exerted by the stimulus activity around some point in a transmitting "layer" L_1 on to a point in a receiving layer L_2. In analogy to our previous considerations we may formally introduce a "distributed

action function" $K(xyz, uvw)$ which defines the incremental contribution to the response density $d\rho(uvw)$ from the stimulus activity that prevails in an incremental volume dV_1 around a point (xyz) in the transmitting layer. This activity is, with our definition of stimulus density $\sigma(xyz)$ dV_1. Consequently

$$d\rho(uvw) = K(xyz, uvw)\, \sigma(xyz)\, dV_1 . \qquad (95)$$

In other words, K expresses the fraction per unit volume of the activity around point (xyz) that contributes to the response at (uvw). The total response elicited at point (uvw) from all regions in layer L_1 is clearly the summation of all incremental contributions, if we assume that all cells around (uvw) are linear elements. Hence, we have

$$\rho(uvw) = \int_{V_1} K(xyz, uvw)\, \sigma(xyz)\, dV_1 , \qquad (96)$$

where V_1, the subscript to the integral sign, indicates that the integration has to be carried out over the whole volume V_1 representing the extension of layer L_1.

With this expression we have arrived at the desired relation that gives the response density at any point in L_2 for any stimulus density distribution in layer L_1, if the distributed action function K is specified.

In order to make any suggestions as to the

form of this distributed action function, it is necessary to enliven the formalism used so far with physiologically tangible concepts. This we shall do presently. At the moment we adopt some simplifying notations. First, we may in various instances refer to cell assemblies distributed along surfaces (A) or along lines (D). In these cases we shall not change symbols for σ and ρ, although all densities refer in these cases to units of length. This may be permissible because the units will be clear from context. Second, we shall adopt for the discussion of generalities vector representation for the localization of our points of interest and introduce the point vector r. Discrimination of layers will be done by subscripts. We have the following correspondences:

transmitting "layer": x, y, z; r_1; D_1; A_1; V_2;

collecting "layer": u, v, w; r_2; D_2; A_2; V_2.

The physiological significance of the distrib-

Fig. 29 Departure of fibers in the transmitting layer of an action network of cell assemblies
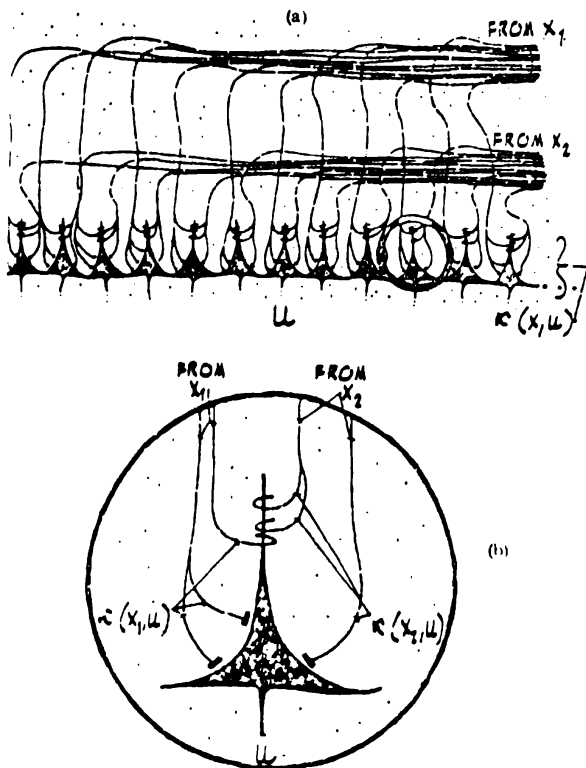
Fig. 30 Arrival of fibers at the target layer of an action network of cell assemblies.

uted action function $K(r_1, r_2)$ will become evident with the aid of figs. 29 and 30. Fig. 29a – or 29b – shows a linear array of neurons in a small interval of length dx about a point x in layer L_1. These neurons give rise to a number of axons N_x, some of which, say, $N_x(u_1)$, are destined to contact in the collecting layer L_2 with elements located in the vicinity of u_1; others, say, $N_x(u_2)$, will make contact with elements located at u_2, and so on:
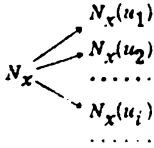
We shall define a "distribution function" $k(x, u_i)$ which is simply that fraction of all the fibers that emerge from x and terminate at u_i:

$$k(x, u_i) = \frac{N_x(u_i)}{N_x}$$

or

$$N_x(u_i) = k(x, u_i) N_x . \tag{97}$$

Clearly, if we take the summation over all targets u_i reached by fibers emerging from x, we must obtain all fibers emerging from x:

$$N_x = \sum_{\text{all } i} N_x(u_i) = \sum_{\text{all } i} k(x, u_i)N_x = N_x \sum_{\text{all } i} k(x, u_i) ,$$

or, after cancellation of N_x on both sides:

$$1 = \sum_{\text{all } i} k(x, u_i) .$$

If we consider now infinitesimal targets of length du, the above summation takes on the form of an integral

$$\int_{-\infty}^{+\infty} k(x, u) \, du = 1 . \tag{98}$$

This suggests that $k(x,u)du$ may be interpreted as the probability for a fiber which originates at x will terminate within an interval of length du in the vicinity of u. Consequently, one interpretation of $k(x, u)$ is a "probability density function".

With this observation we may derive an expression for the contribution of region x to the fiber number density of region u. The corresponding fiber number densities are clearly defined by

$$n(x) = \frac{dN}{dx} \quad \text{and} \quad n(u) = \frac{dN}{du} .$$

Since $n(x)dx$ fibers all together emerge from an interval of length dx in the vicinity of x, their contribution to the number of fibers in the interval du at u is:

$$d^2N (u) = [k(x, u)du][n(x)dx] \tag{99}$$

with d^2 indicating that this is an infinitesimal expression of second order (an infinitesimal amount of an infinitesimal amount. Compare with eq. (97), its finite counterpart). Dividing in eq. (99) both sides by du we note that

$$\frac{d^2 N_x(u)}{du} = d\left(\frac{dN}{du}\right)_x = dn_x(u) \tag{100}$$

is the contribution of x to the number density of fibers at point u. Using eqs. (99), (100), we can now express the desired relation between source and target densities by

$$dn(u) = k(x, u)n(x)dx . \tag{101}$$

From this point of view, k represents a mapping function that defines the amount of convergence or divergence of fiber bundles leaving the vicinity of point x and destined to arrive in the vicinity of point u. Clearly, k represents an important structural property of the network.

For the present discussion it is irrelevant whether certain neurons around x are the donors for elements around u, or whether we assume that after axonal bifurcation some branches are destined to contact elements around u. In both cases we obtain the same expression for the fractional contribution from x and u (compare figs. 29a and b).

If we pass over each fiber an average amount $\bar{s}$ of activity, we obtain the stimulus which is funneled from x to u by multiplying eq. (101) with this amount:

$$\bar{s} \, dn(u) = d\sigma(u) = k(x, u)\sigma(x)dx , \tag{102}$$

because

$$\bar{s}n(x) = \sigma(x) , \qquad \bar{s}n(u) = \sigma(u) . \tag{103}$$

If the fractional stimulus density $d\sigma(u)$ at the target were translated directly into response density, we would have

$$\sigma(u) = \rho(u) .$$

However, this is not true, for the arriving fibers will synapse with the target neurons in a variety of ways (fig. 30). Consequently, the resulting re-

sponse will depend upon the kind and strength of these synaptic junctions which again may be a function of source and target points. To accommodate this observation we introduce a local transfer function $\kappa(x, u)$, that relates arriving stimulus with local response

$$d\rho(u) = \kappa(x, u) \, d\sigma(u) . \qquad (104)$$

With the aid of eq. (102) we are now in a position to relate stimulus density in the source area to response density in the target area:

$$d\rho(u) = \kappa(x, u) \, k(x, u) \, \sigma(x) \, dx . \qquad (105)$$

Clearly, the product of the two functions κ and k can be combined to define one "action function"

$$K(x, u) = \kappa(x, u) \, k(x, u) , \qquad (106)$$

and eq. (105) reduces simply to

$$d\rho(u) = K(x, u) \, \sigma(x) \, dx . \qquad (107)$$

Comparison of this equation with our earlier expression for the stimulus-response relationship (eq. (95)) shows an exact correspondence, eq. (107) representing the x-portion of the volume representation in eq. (95).

With this analysis we have gained the important insight that action functions — and clearly also interaction functions — are composed of two parts. A structural part $k(r_1, r_2)$ defines the geometry of connecting pathways, and a functional part $\kappa(r_1, r_2)$ defines the operational modalities of the elements involved. The possibility of subdividing the action function into two clearly separable parts introduces a welcome constraint into an otherwise unmanageable number of possibilities.

We shall demonstrate the workings of the mathematical and conceptual machinery so far developed on three simple, but perhaps not trivial, examples.

(i) Ideal one-to-one mapping

Assume two widely extending, but closely spaced, parallel surfaces, representing layers L_p and L_q. Perpendicular to L_p emerge parallel fibers which synapse with their corresponding elements in L_q without error and without deviation. In this case the mapping function k is simply Dirac's Delta Function *

$$k(r_1, r_2) = \delta^2(r_1 - r_2) ,$$

---

* The exponent in δ indicates the dimensionality of the manifold considered. Here it is two-dimensional, hence δ^2

$$\delta(x - x_0) = \begin{cases} 0 & \text{for } x \neq x_0 \\ \infty & \text{for } x = x_0 \end{cases}$$

and

$$\int_{-\infty}^{+\infty} \delta(x - x_0) \, dx = 1 .$$

For simplicity, let us assume that the transfer function κ is a constant a. Hence

$$K(r_1, r_2) = a\delta^2(r_1 - r_2) ,$$

and, after eq. (96):

$$\rho(r_2) = a \int_{L_p} \delta^2(r_1 - r_2) \, \sigma(r_1) \, dA_1 = a\sigma(r_1) .$$

As was to be expected, in this simple case the response is a precise replica of the stimulus, multiplied by some proportionality constant.

(ii) Ideal mapping with perturbation

Assume we have the same layers as before, with the same growth program for fiber descending upon L_q, but this time the layers are thought to be much further apart. Consequently, we may expect the fibers to be affected by random perturbations, and a fiber bundle leaving at r_1 and destined for $r_2 = r_1$ will be scattered according to a normal (Gaussian) distribution. Hence, we have for the mapping function

$$k(r_1, r_2) = 1/2\pi h^2 \exp(-\Delta^2/2h^2)$$

with

$$\Delta^2 = |r_1 - r_2|^2$$

and h representing the variance of the distribution.

Assume furthermore that the probability distribution for facilitatory and inhibitory contacts are not alike, and let κ be a constant for either kind. The action function is now:

$$K(r_1, r_2) = K(\Delta^2)$$

$$= a_1 \exp(-\Delta^2/2h_1^2) - a_2 \exp(-\Delta^2/2h_2^2) .$$

The stimulus-response relation is

$$\rho(r_2) = \int_{L_1} K(\Delta^2) \, \sigma(r_1) \, dA_1 .$$

For a uniform stimulus distribution

$$\sigma = \sigma_0 ,$$

$$\rho(r_2) = \text{constant} = 2\pi\sigma_0(a_1 h_1^2 - a_2 h_2^2) ,$$

and if

$$a_1 h_1^2 = a_2 h_2^2 ,$$
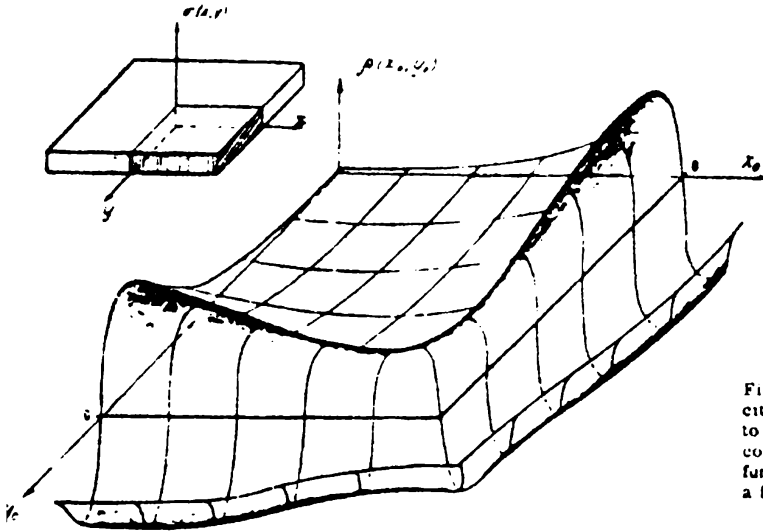
Fig. 31. Response distribution eli-cited by a uniform stimulus confined to a square. Contour detection is the consequence of a distributed action function obtained by superposition of a facilitatory and inhibitory Gaussian distribution.

then

$$\rho = 0 .$$

As one may recall, the interesting feature of giving zero-response to finite stimuli was obtained earlier for discrete action functions of the binomial form. A similar result for the Gaussian distribution should therefore not be surprising, if one realizes that the continuous Gaussian distribution emerges from a limit operation on the binomial distribution. What are the abstracting properties of this net with a random normal fiber distribution?

Fig. 31 represents the response activity for a stimulus in the form of a uniformly illuminated square. Clearly, this network operates as a computer of contours. It may be noted that the structure of this useful network arose from random perturbation. However, there was an original growth program, namely, to grow in parallel bundles. Genetically, this is not too difficult to achieve; it says: "Repeat". However, such a net is of little value, as we have just seen. It is only when noise is introduced into the program that the net acquires its useful properties. It may be mentioned that this net works best when the zero-response condition is fulfilled. This, how-ever, requires adaptation.

(iii) Mapping into a perpendicular plane

The previous two examples considered action functions with spherical symmetry. We shall now explore the properties of an action function of type K_s with lateral symmetry. Such an action
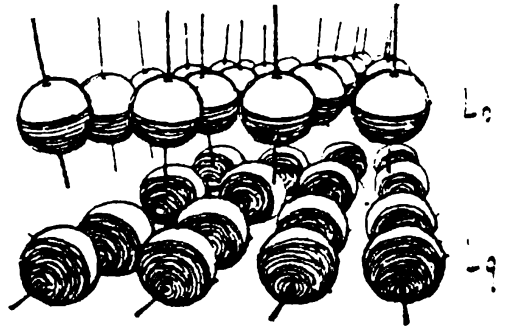
Fig. 32. Geometrical relationship between layers of neurons with perpendicular orientation of their axes.

function may arise in the following way (see fig. 32):

Assume again two surfaces layers, L_p, L_q where in L_p all neurons are aligned in parallel with their axis of symmetry perpendicular to the layer's surface, while in L_q they are also in parallel, but with their axis of symmetry lying in the surface of L_q.

We consider pyramidal neurons and represent them by spheres. We let the North pole (-) coincide with basal axonal departure and the South pole (-) with the upper branchings of apical dendrites. The perikaryon is central. This spherical dipole assumes physiological significance if we associate with the neuron's structural difference when seen from north or from south different

probabilities for the establishment of facilitatory or inhibitory synapses. For simplicity we assume that for an afferent fiber the probability of making facilitatory or inhibitory connection is directly proportional to the projected areas of northern and southern hemisphere respectively, seen by this fiber when approaching the neuron. Hence, a fiber approaching along the equatorial plane has a 50-50 chance to either inhibit or facilitate. A fiber descending upon the South Pole inhibits with certainty. Let there be two kinds of fibers descending from L_p to L_q. The first kind maps with a Dirac delta function the activity of l_p into L_q.

$$k_1(\Delta) = \delta(\Delta) .$$

Let k_2, the mapping function of the second kind, be any spherical symmetric function that converges:

$$\int_{L_1} k_2(\Delta^2)dA_1 = \text{constant} ,$$

for instance, a normal distribution function. The associated local transfer function κ_2, however, is not spherical symmetric because of the lateral symmetry of the probability of (ι) connections. Simple geometrical considerations (fig. 33) show that κ_2 is of the form

$$\kappa_2(r_1, r_2) = a \cos \phi ,$$

where ϕ is the angle between Δ and the N-S axis of elements.

The action function of the network is

$$K(\Delta) = a\delta(\Delta) \, k_2(\Delta^2) \cos \phi ,$$

and the response density for a given stimulus:

$$\rho(r_2) = a \int_{l_t} \delta(\Delta) \, k_2(\Delta^2) \cos \phi \, \sigma(r_1) \, dA_1 .$$

What does this system compute?

Its usefulness becomes obvious if we assume that the stimulus of L_q is a contour that has been computed in L_p from a preceding network say, L_o, L_p. On behalf of the δ-function this contour, and nothing else, maps from L_p into L_q. Take, for instance, a straight line with uniform intensity σ_0 to be the stimulus for L_q (see fig. 34). Since two points symmetrical to any point on this line contribute

$$ak_2\sigma_0(\cos \phi + \cos(180 - \phi)) \Delta A_1 = 0 ,$$

because

$$\cos \phi = -\cos(180 - \phi) .$$
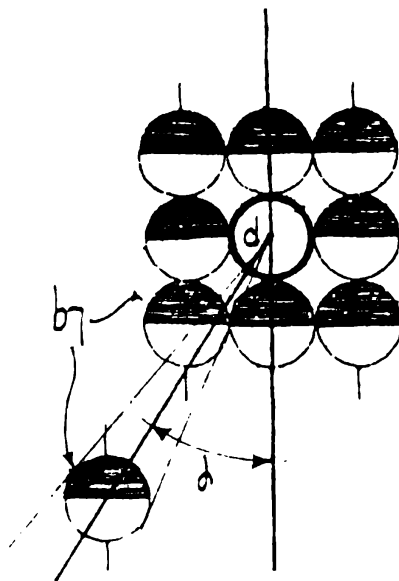
Fig. 33. Geometrical relationship between elements of two layers. View from the top.
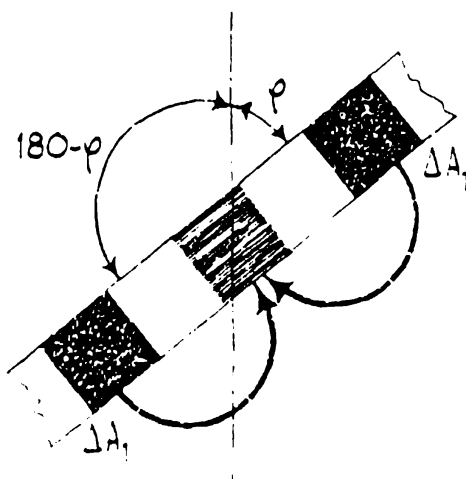
Fig. 34 Insensitivity of antisymmetrical action net to straight lines.

a straight line gives no response. Curvature, however, is reported. The total net-response

$$\bar{R} = \int_{L_2} \rho(r_2)\, dA_2$$

vanishes for bilateral symmetrical figures with their axis of symmetry parallel to the orientation of elements in L_q. However, when turned away from this position, $\bar{R} \neq 0$ (see fig. 35).
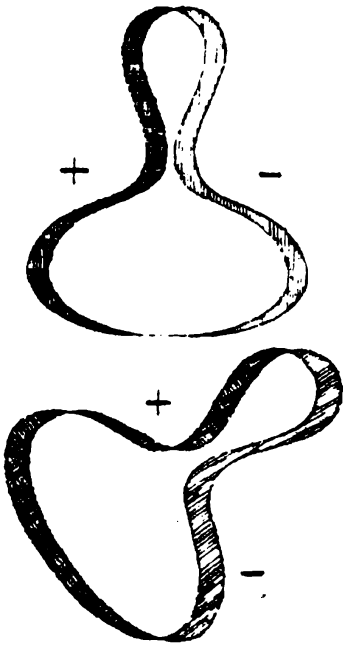
Fig. 35. Responses of antisymmetrical action net to figures with bilateral symmetry.

4.8. *Interaction networks of cell assemblies*

We consider the case where connections between all elements in the network appear freely and a separation into layers of purely-transmitting and purely-receiving elements is impossible. When all return connections between interconnected elements are cut, the system reduces to an action network. Clearly, we are dealing here with the more general case and consequently have to be prepared for results that do not yield as easily as in the previous case. However, the methods and concepts developed in the previous section are immediately applicable to our present situation.
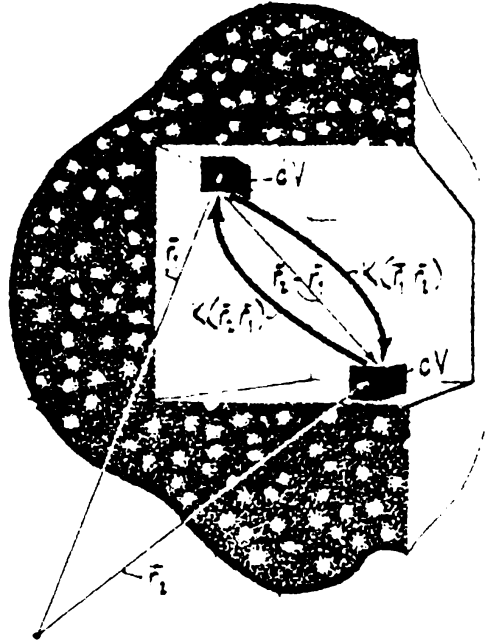
Fig. 36. Geometry in an interaction network of cell assemblies.

Fig. 36 sketches a large interaction network, confined to volume V, in which a certain portion has been cut away in order to make clear the geometrical situation. We fix our attention on elements in the vicinity of point r_2 and determine the contribution to their stimulation from other regions of the network. We consider in particular the contribution to r_2 from the activity around point r_1. We proceed precisely as before (see eq. (95)) and write

$$d\rho(r_2) = K(r_1, r_2)\, \rho(r_1)\, dV, \qquad (103)$$

where K, the distributed interaction function, defines again the fraction per unit volume of the activity prevailing around point r_1 that is transmitted to point r_2, and $\rho(r_1)dV$ is clearly the activity of elements in the vicinity of point r_1.

Of course, the same physiological interpretation that has been given to the action function is applicable to the interaction function, except that properties of symmetry refer to symmetry of exchanges of stimulation between two points. In other words, $K(r_1, r_2)$ and $K(r_2, r_1)$ describe the proportions that are transmitted from r_1 to r_2, and back from r_2 to r_1 respectively. These proportions may not necessarily be the same.

In addition to the stimuli contributed by elements of its own network, each element may, or may not, receive stimulation from fibers descending upon this network from other systems that do not receive fibers from the network under consideration. We denote the elementary stimulation so contributed to r_2 by $d\sigma(r_2)$. It is, of course, no restriction to assume that these fibers stem from another network, say V_0, that functions as action network on to our system. In this case $d\sigma(r_2)$ may be directly replaced by $d\mu(r_2)$ of eq. (95), noting, however, that the action function in this expression has to be changed into, say, $A(r_0, r_2)$, where r_0 indicates positions of elements in this donor system. With $d\sigma(r_2)$ representing a stimulus from external sources to elements around r_2 of our network we have for the total elementary stimulus at r_2:

$$d\mu(r_2) = d\sigma(r_2) + K(r_1, r_2)\, \rho(r_1)\, dV , \qquad (109)$$

which summed over the entire volume V gives the desired stimulus-response relationship for any point in this volume:

$$\rho(r_2) = \sigma(r_2) + \int_V K(r_1, r_2)\, \rho(r_1)\, dV . \qquad (110)$$

This equation cannot be readily solved by integration, unlike the case for action networks, because here the unknown quantity ρ appears not only explicitly on the left-hand side of this equation, but also implicitly within the integral. Expressions of this type are called integral equations and (110) above belongs to the class of integral equations of the second kind. The function $K(r_1, r_2)$ is usually referred to as the "kernel", and methods of solution are known, if the kernel possesses certain properties.

It is fortunate that a general solution for eq. (110) can be obtained (Inselberg and Von Foerster, 1962, p. 32) if the kernel K is a function of only the distance between points r_1 and r_2:

$$K(r_1, r_2) = K(r_1 - r_2) = K(\Delta) , \qquad (111)$$

where Δ stands again (see eq. (86)) for the vector expressing this distance. These kernels represent precisely the kind of interaction function we wish to consider, for it is this property that makes the computations in the network invariant to stimulus translations (see eq. (85)).

The general solution for response, given explicitly in terms of stimulus and interaction function, is, for the x-component:

$$\rho(x) = 1/\sqrt{2\pi} \int_{-\infty}^{+\infty} \frac{F_s(u)}{1 - \sqrt{2\pi}\, F_k(u)}\, e^{-\mathrm{i}xu}\, du , \qquad (112)$$

where F_s and F_k are the Fourier transforms of stimulus distribution σ and interaction function K respectively:

$$F_s(u) = 1/\sqrt{2\pi} \int_{-\infty}^{+\infty} \sigma(x)\, e^{\mathrm{i}xu}\, dx , \qquad (113a)$$

$$F_k(u) = 1/\sqrt{2\pi} \int_{-\infty}^{+\infty} K(t)\, e^{\mathrm{i}tu}\, dt , \qquad (113b)$$

with t representing the x component of the distance and i the imaginary unit

$$t = \Delta_x , \qquad \mathrm{i} = \sqrt{-1} . \qquad (114)$$

The expressions are valid for the other components, mutatis mutandis.

With respect to these results two comments are in order. First, one should observe the analogy of eq. (110) with the result obtained in the case of interaction of individually distinguishable elements (eq. (65)).

$$\rho(j) = \sigma(j) + \sum_{i=1}^{n} a_{ij}\, \rho(i) , \qquad (65)$$

where summation over the activity of individual elements and the interaction coefficients a_{ij} correspond to integration and interaction function in (110) respectively. However, the cumbersome matrix inversions as suggested in eqs. (66) to (74) disappear, because the Fourier transforms in eqs. (112) and (113) perform these inversions — so to say — in one stroke. Thus, a general study of network structures will have to proceed along the lines suggested here, otherwise sheer manipulatory efforts may attenuate the enthusiasm for exploring some worthwhile possibilities.

The other comment refers to our earlier observation of the functional equivalence of discrete action and interaction nets (see fig. 23). The question arises whether or not an action network can be found that has precisely the same stimulus-response characteristic as a given interaction network of cell assemblies. It is not insignificant that this question can be answered in the affirmative. Indeed, it can be shown that a functional equivalent action net with action function $A(t)$ can be generated from a given interaction net with interaction function $K(t)$ by the Fourier transform

$$A(t) = 1/.2\pi \int_{-\infty}^{+\infty} \frac{1}{1 - \sqrt{2\pi}\, F_k(u)}\, e^{-\mathrm{i}tu}\, du . \qquad (115)$$

These transforms are extensively tabulated (Magnus and Oberhettinger, 1949) and permit one

to establish quickly the desired relationships.

Since the same performance can be produced by two entirely different structural systems one may wonder what is Nature's preferred way of accomplishing these performances: by action or by interaction networks? This question can, however, be answered only from an ontogenetic point of view. Since the "easy" way to solve a particular problem is to use most of what is already available, during evolution the development of complex net structures of either kind may have arisen out of a primitive nucleus that had a slight preference for developing in one of these directions. Nevertheless, there are the two principles of "action at a distance" and "action by contagion", where the former may be employed when it comes to highly specified, localized activity, while the latter is effective for alerting a whole system and swinging it into action. The appropriate networks which easily accommodate these functions are obvious, although the equivalence principle may reverse the situation.

Examples

(i) Gaussian, lateral inhibition

We give as a simple example an interaction net that produces highly localized responses for not-well-defined stimuli. Consider a linear network with purely inhibitory interaction in the form of a normal distribution:

$$K(\Delta) = K(x_1 - x_2) \sim \exp[-p(x_1 - x_2)^2] .$$

As a physiological example one may suggest the mutually inhibiting action in the nerve net attached to the basilar membrane which is assumed to be responsible for the sharp localization of frequencies on it.

Suppose the stimulus – in this case the displacement of the basilar membrane as a function of distance x from its basal end – is expressed in terms of a Fourier series

$$\sigma(x) = a_0 + \Sigma a_i \sin(2\pi i x/\lambda) + \Sigma b_i \cos(2\pi i x/\lambda) ,$$

$$i = 0, 1, 2, \dots ,$$

with coefficients a_i, b_i and fundamental frequency λ. It can be shown from eq. (110) that the response will be also a periodic function which can be expressed as a Fourier series with coefficients a_i^* and b_i^*. These have the following relation to the stimulus coefficients:

$$\frac{a_i^*}{a_i} = \frac{b_i^*}{b_i} = \frac{p}{p + \exp[-4\pi(i/\lambda)^2]} .$$

Since this ratio goes up with higher mode num-

bers i, the higher modes are always enhanced, which shows that indeed an interaction function with inhibitory normal distribution produces considerable sharpening of the original stimulus.

(ii) Antisymmetric interaction

As a final example of a distributed interaction function which sets the whole system into action when stimulated only by that most local stimulus, the Dirac delta function, we suggest an antisymmetric one-dimensional interaction function

$$K(\Delta_x) = \frac{\sin^2 \Delta_x}{\Delta_x} .$$

This function inhibits to the left ($\Delta < 0$) and facilitates to the right ($\Delta_x > 0$). It is, in a sense, a close relative to the one-directional action-interaction function of the quadrupole chain (fig. 21) discussed earlier. We apply to this network at one point x a strong stimulus:

$$(x) = \delta(x) .$$

The response is of the form

$$\rho(x) = \delta(x) - a \frac{\sin x}{x} \cos(x - \alpha) ,$$

a and α being constants.

Before concluding this highly eclectic chapter on some properties of computing networks it is to be pointed out that a general theory of networks that compute invariances on the set of all stimuli has been developed by Pitts and McCulloch (1947). Their work has to be consulted for further expansion and deeper penetration of the cases presented here.

5. SOME PROPERTIES OF NETWORK ASSEMBLIES

In this approach to networks we first considered nets composed of distinguishable elements. We realized that in most practical situations the individual cell cannot be identified and we developed the notions of acting and interacting cell assemblies whose identity was associated only with geometrical concepts. The next logical step is to drop even the distinguishability of individual nets and to consider the behavior of assemblies of nets. Since talking about the behavior of such systems makes sense only if they are permitted to interact with other systems – usually called the "environment" – this topic does not properly belong to an article confined to networks and, hence, has to be studied elsewhere

(Pask, 1966). Nevertheless, a few points may be made, from the network point of view, which illuminate the gross behavior of large systems of networks in general.

We shall confine ourselves to three interrelated points that bear on the question of stability of network assemblies. Stability of network structures can be understood in essentially three different ways. First, in the sense, of a constant or periodic response density within the system despite various input perturbations (Dynamic Stability); second, in terms of performance, i.e., the system's integrity of computation despite perturbations of structure or function of its constituents (Logical Stability); third, to reach stabilities in the two former senses despite permanent changes in the system's environment (Adaptation). We shall briefly touch upon these points.

5.1. *Dynamic stability*

Beurle (1962) in England and Farley and Clark (1962) at MIT were probably the first to consider seriously the behavior of nets of randomly connected elements with transfer functions comparable to eq. (47). Both investigated the behavior of about a thousand elements in a planar topology and a neighborhood connection scheme. Beurle used his network to study computation with distributed memory. To this end, elements were constructed in such a way that each activation caused a slight threshold reduction at the site of activity and made the element more prone to fire for subsequent stimuli. The system as a whole shows remarkable tendencies to stabilize itself, and it develops dynamic "engrams" in the form of pulsating patterns. Farley and Clark's work is carried out by network simulation on the Lincoln Laboratory's TX-a computer; and the dynamic behavior resulting from defined stimuli applied to selected elements is recorded with a motion picture camera. Since elements light up when activated, and the calculation of the next state in the network takes TX-2 about 0.5 seconds, the film can be presented at normal speed and one can get a "feeling" for the remarkable variety of patterns that are caused by variations of the parameters in the network. However, these "feelings" are at the moment our best clues to determine our next steps in the approach to these complicated structures.

Networks composed of approximately one thousand Ashby elements (see fig. 13) were studied by Fitzhugh (1963) who made the significant observation that slowly adding connections to the element defines with reproducible accuracy, a "connectedness" by which the system swings from almost zero activity to full operation, with a relatively small region of intermediate activity. This is an important corollary to an observation made by Ashby et al. (1962), who showed that networks composed of randomly connected McCulloch elements with facilitatory inputs only, but controlled by a fixed threshold, show no stability for intermediate activity, only fit or coma, unless threshold is regulated by the activity of elements.

In all these examples, the transfer function of the elements is varied in some way or another in order to stabilize the behavior of the system. This, however, implies that in order to maintain dynamic stability one has to sacrifice logical stability, for as we have seen in numerous examples (e.g., fig. 10) variation in threshold changes the function computed by the element. Hence, to achieve both dynamic and logical stability it is necessary to consider logically stable networks that are immune to threshold variation.

5.2. *Logical stability*

McCulloch (1958) and later Blum (1962), Verbeek (1962) and Cowan (1962) were probably the first to consider the distinction between proper computation based on erroneous arguments (calculus of probability) and erroneous calculation based on correct arguments (probabilistic logic). It is precisely the latter situation one encounters if threshold in a system is subjected to variations. On the other hand, it is known that living organisms are reasonably immune to considerable variations of threshold changes produced by, say, chemical agents. Clearly, this can only be accomplished by incorporating into the neural network structure nets that possess logical stability.

The theory developed by the authors mentioned above permits the construction of reliable networks from unreliable components by arranging the various components so that, if threshold changes occur in the net, an erroneous computation at one point will be compensated by the computation at another point. However, the theory is further developed for independent changes of threshold everywhere, and nets can be developed for which the probability of malfunctioning can be kept below an arbitrarily small value, if sufficient components are added to the system. This, of course, increases its redundancy. However, this method requires substantially fewer elements to achieve a certain degree of reliability than usual "multiplexing" requires in order to operate with equal reliability. Due to a multivaluedness in the structure of these nets,

an additional bonus offered by this theory is im-
munity against perturbation of connections.

All these comments seem to imply that varia-
bility of function tends to increase the stability
of systems to such a degree that they will be-
come too rigid when secular environmental
changes demand flexibility. On the contrary,
their very complexity adds a store of potentially
available functions that enables these networks
to adapt.

5.3. *Adaptation*

This ability may be demonstrated on an extra-
ordinarily simple network composed of McCul-
loch elements operating in synchrony (fig. 37).
It consists of two arrays of elements, black and
white denoting sensory and computer elements
respectively. Each computer element possesses
two inputs proper that originate in two neigh-
boring sensory elements A, B. The output of
each computer element leads to a nucleus Σ that
takes the sum of the outputs of all computer
elements. The logical function computed by these
is not specified. Instead it is proposed that each
element is capable — in principle — of computing
all 16 logical functions. These functions are to
change from functions of lowest logical strength
Q (see table 1 or fig. 38) to functions of higher
logical strength in response to a command given
by an "improper input", whose activity is de-
fined by Σ that operates on the functions — the
inner structure — of these elements and not, in a
direct sense, on their outputs.

Consider this net exposed to a variety of
stimuli which consist of shadows of one-dimen-
sional objects that are in the "visual field" of
the sensors. With all computing elements oper-
ating on functions with low Q the feedback loop
from Σ is highly active and shifts all functions
to higher Q's. It is not difficult to see that this
process will go on with decreasing activity in
the loop until function No. 6 or even functions
No. 4, or 2 of table 1 are reached, at which in-
stant the net is ready to compute the presence of
edges in the stimulus field — as has been shown
in fig. 25 — and the loop activity is reduced to the
small amount that remains when objects happen
to be in the visual field. Since it is clear that the
output of the whole system represents the num-
ber of edges present at any moment, it repre-
sents at the same time a count of the number of
objects, regardless of their size and position,
and independent of the strength of illumination.

Consider this system as being in contact with
an environment which has the peculiar property
of being populated by a fixed number of objects
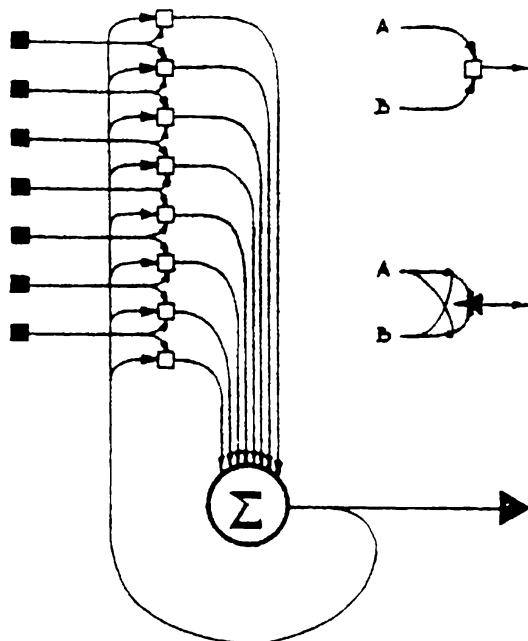
Fig. 37. Adaptive network.
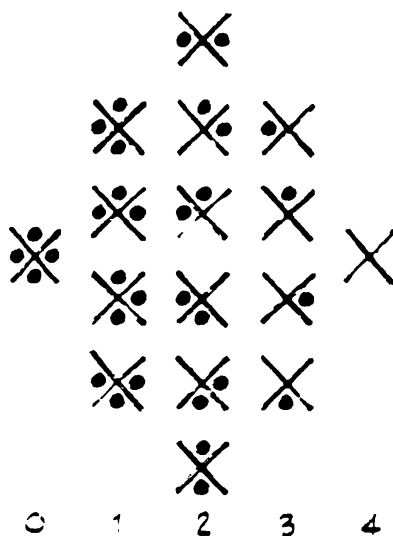
Fig. 38. Logical functions ordered according to in-
creasing logical strength Q.

that freely move about so that the limited visual field consisting of, say, N receptors perceives only a fraction of the number of these objects. In the long run, our system will count objects normally distributed around a mean value with a standard deviation of, say, σ.

The amount of information (entropy) H_{OUT} as reported by the system about its environment is (Shannon and Weaver, 1949):

$$H_{OUT} = \ln \sqrt{2\pi e}$$

with

$$e = 2.71828\ldots$$

On the other hand, with N binary receptors its input information is

$$H_{IN} = N \gg H_{OUT} .$$

This represents a drastic reduction in information – or reduction in uncertainty – which is performed by this network and one may wonder why and how this is accomplished. That such a reduction is to be expected may have been suggested by the system's indifference to a variety of environmental particulars as, e.g., size and location of objects, strength of illumination, etc. But this indifference is due to the network's abstracting powers, which it owes to its structure and the functioning of its constituents.

5.4. Information storage in network structures

Let us make a rough estimate of the information stored by the choice of a particular function computed by the network elements. As we have seen, these abstractions are computed by sets of neighbor elements that act upon one computer element. Let n_s and n_h be the number of neighbors of the kth order in a two-dimensional body-centered square lattice and hexagonal lattice respectively (see fig. 39):

$$n_s = (2k)^2 , \qquad n_h = 3k(k+1) . \tag{116}$$

The number of logical functions with n inputs is (eq. (18))

$$N = 2^{2^n} ,$$

and the amount of information necessary to define a particular function is:

$$H_F = \log_2 N = 2^n \tag{117}$$

On the other hand, the input information on a sensory organ with N binary receptors is

$$H_{IN} = N .$$

A sensory network is properly matched to its structure if

$$H_{IN} = H_F , \tag{119}$$

$$k = 1 \qquad * \qquad n = 4$$
$$k = 2 \qquad \blacktriangle, * \qquad n = 16$$
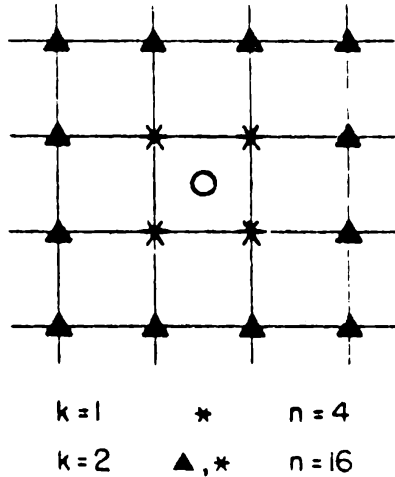
Fig. 39. First and second order neighbors in a body centered cubic lattice (two dimensional "cube").

or, in other words, if its input information corresponds to its computation capacity stored in its structure. We have with eqs. (116), (117), (119):

$$N = \begin{cases} 2^{4k^2} & \text{square lattice} \\ 2^{3k(k+1)} & \text{hexagonal lattice} \end{cases}$$

The following table relates the size of the sensory organ that is properly matched to its computing network which utilizes kth order neighbors, constituting a receptor field of n elements:

N	k_s	k_h	$n_s = n_h$
10^2	1.29	1.07	6.63
10^4	1.82	1.68	13.2
10^5	2.04	1.91	16.7
10^6	2.24	2.12	20.0
10^7	2.41	2.31	23.2
10^8	2.58	2.52	26.5

This table indicates that in an eye of, say, 10^6 receptor elements a receptor field of more than 20 elements is very unlikely to occur.

In conclusion it may be pointed out that the evolution of abstracting network structures as a consequence of interactions with an environment gives rise to new concepts of "memory" which do not require the faithful recording of data. In fact, it can be shown that a theory of memory that is based on mechanisms that store events is not only uneconomical bordering on the impossi-

ble, but also is incapable of explaining the most primitive types of behavior in living organisms that show one or another form of retention (Von Foerster, 1965; Von Foerster et al., 1966).

ACKNOWLEDGEMENTS

REFERENCES

Ashby, W. R., 1956. *An introduction to cybernetics* (Chapman and Hall, London).

Ashby, W. R., H. Von Foerster and C. C. Walker, 1962, Instability of pulse activity in a net with threshold, Nature 196, 561.

Bar-Hillel, Y., 1955. Semantic information and its measure, in: *Cybernetics*, eds. H. Von Foerster, M. Mead and H. L. Teuber (Josiah Macy Jr. Foundation, New York) p. 33.

Beurle, R. L., 1962. Functional organizations in random networks, in: *Principles of self-organization*, eds. H. Von Foerster and G. W. Zopf Jr. (Pergamon Press, New York) p. 291.

Blum, M., 1962, Properties of a neuron with many inputs, in: *Principles of self-organization*, eds. H. Von Foerster and G. W. Zopf Jr. (Pergamon Press, New York) p. 98.

Carnap, R., 1938, *The logical syntax of language* (Harcourt, Brace and Co., New York).

Cowan, J., 1962, Many-valued logics and reliable automata, in: *Principles of self-organization*, eds. H. Von Foerster and G. W. Zopf Jr. (Pergamon Press, New York) p. 135.

Eccles, J. C., 1952, *The neurophysiological basis of mind* (Clarendon Press, Oxford).

Farley, B., 1962, Some similarities between the behavior of a neural network model and electrophysiological experiments, in: *Self-organizing systems*, eds. M. C Yovits et al. (Spartan Books, Washington, D. C., 1962) p. 207.

Fitzhugh II, H. S., 1963, Some considerations of polystable systems, *Bionics Symposium*, Dayton, Ohio, USAF Aeronautical Systems Division, Aerospace Medical Division, 19-21 March 1963.

Fulton, J. F., 1943, *Physiology of the nervous system* (Oxford University Press, New York).

Hartline, H. K., 1959, Receptor mechanisms and the integration of sensory information in the eye, *Biophysical science*, ed. J. L. Oncley (John Wiley and Sons, New York) p. 515.

Hilbert, D. and W. Ackermann, 1928, *Grundzüge der theoretischen Logik* (Springer, Berlin).

Hubel, D. H., 1962, Transformation of information in the cat's visual system, in: *Information processing in the nervous system*, eds. R. W. Gerard and J. W. Duyff (Excerpta Medica Foundation, Amsterdam) p. 160.

Inselberg, A. and H. Von Foerster, 1962, *Property extraction in linear networks*, NSF Grant 17414, Tech. Rep. No. 2, (Electrical Engineering Research Laboratory, Engineering Experiment Station, University of Illinois, Urbana).

Katz, B., 1959, Mechanisms of synaptic transmission, in: *Biophysical science*, ed. J. L. Oncley (John Wiley and Sons, New York) p. 524.

Kinosita, H., 1941, Conduction of impulses in superficial nervous systems of sea urchin, Jap. J. Zool. 9, 221.

Landahl, H. D., W. S. McCulloch and W. Pitts, 1943. A statistical consequence of the logical calculus of nervous nets. Bull. Math. Biophys. 5, 135.

Lettvin, J. Y., H. R. Maturana, W. S. McCulloch and W. Pitts, What the frog's eye tells the frog's brain, Proc. Inst. Radio Engrs. 47, 1940.

Magnus, W. and F. Oberhettinger, 1949, *Special functions of mathematical physics* (Chelsea, New York) p. 115.

Maturana, H. R., 1962, Functional organization of the pigeon retina, in: *Information processing in the nervous system*, eds. R. W. Gerard and J. W. Duyff (Excerpta Medica Foundation, Amsterdam) p. 170.

McCulloch, W. S., 1958, The stability of biological systems, in: *Homeostatic mechanisms*, ed. H. Quastler (Brookhaven Natl. Lab. Upton) p. 207.

McCulloch, W. S., 1962, Symbolic representation of the neuron as an unreliable function, in: *Principles of self-organization*, eds. H. Von Foerster and G. W. Zopf Jr. (Pergamon Press, New York) p. 91.

McCulloch, W. S. and W. Pitts, A logical calculus of the ideas immanent in nervous activity, Bull. Math. Biophys. 5, 115.

Mountcastle, V. B., G. F. Poggio and G. Werner, 1962, The neural transformation of the sensory stimulus at the cortical input level of the somatic afferent system, in: *Information processing in the nervous system*, eds. R. W. Gerard and J. W. Duyff (Excerpta Medica Foundation, Amsterdam) p. 196.

Parker, G. H., 1943, *The elementary nervous system* (Lippincott, Philadelphia).

Pask, G., 1966, Comments on the cybernetics of ethical, sociological and psychological systems. In: *Progress in biocybernetics*, vol. 3, eds. N. Wiener

and J P Schadé (Elsevier, Amsterdam) p. 158.

Pitts, W. and W S McCulloch, 1947, How we know universals, Bull. Math. Biophys. 9, 127.

Russell, B. and A N Whitehead, 1925, *Principia mathematica* (Cambridge University Press, Cambridge).

Shannon, C. E and W. Weaver, 1949, *The mathematical theory of communication* (University of Illinois Press, Urbana) p. 56.

Sherrington, C. S., 1906, *Integrative action of the nervous system* (Yale University Press, New Haven).

Sherrington, C. S., 1952, Remarks on some aspects of reflex inhibition, Proc. Roy. Soc. B 96, 519.

Sholl, D. A., 1956, *The organization of the cerebral cortex* (Methuen and Co., London).

Stevens, S. S., 1957, Psych. Rev. 64, 153.

Taylor, D. H., 1962, *Discrete two-dimensional property detectors*, Tech. Rep. No. 5, NSF G-17414, (Electrical Engineering Research Laboratory, Engineering Experiment Station, University of Illinois, Urbana.)

Uxkull, J V., 1896, Über Reflexe bei den Seeigeln, Z. Biol 34, 298.

Verbeek, L., 1962, On error minimizing neural nets, in: *Principles of self-organization*, eds. H. Von Foerster and G. W. Zopf Jr. (Pergamon Press, New York) p. 122.

Von Foerster, H., 1962, Circuitry of clues to platonic ideation, in: *Artificial intelligence*, ed. C. A. Muses (Plenum Press, New York) p. 43.

Von Foerster, H., 1965, Memory without record, in: *The anatomy of memory*, ed. D. P. Kimble (Science and Behavior Books Inc., Palo Alto) p. 388.

Von Foerster, H., A. Inselberg and P. Weston, 1966, Memory and inductive inference, in: *Bionics Symposium 1966*, ed. H. Oestreicher (John Wiley and Sons, New York) in press.

Von Neumann, J., 1951, The general and logical theory of automata, in: *Cerebral mechanisms in behavior*, ed. L. A. Jeffres (John Wiley and Sons, New York) p. 1.

Wittgenstein, L., 1956, *Tractatus logico-philosophicus* (Humanities Publ., New York).
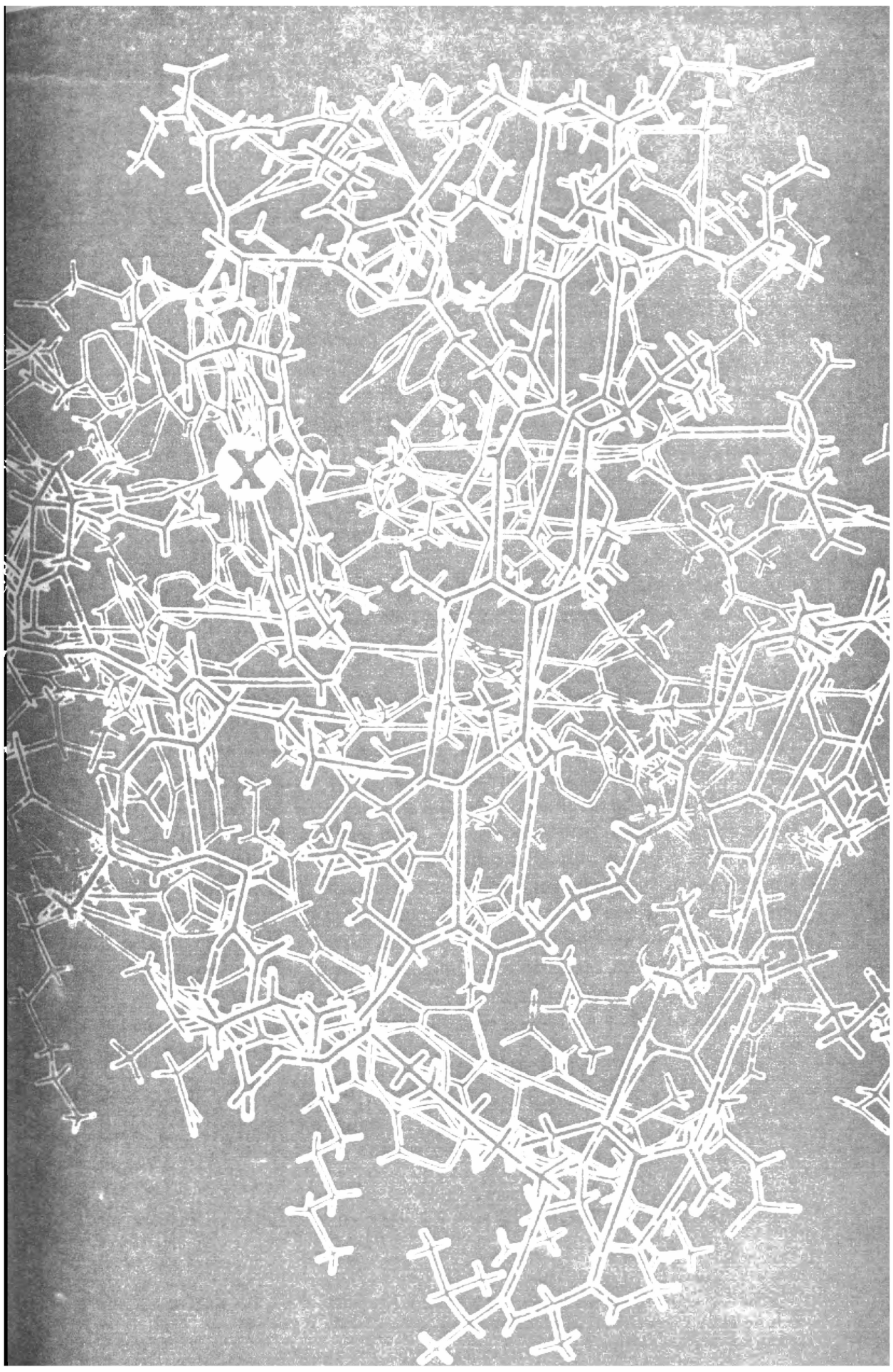
MOLECULAR BIONICS*

INTRODUCTION

The notion that the mysteries of life have their ultimate roots in the microcosm is certainly not new, and a history of this notion may take a whole lecture. Consequently, I hope that I will be forgiven if I mention in the long string of authors only Erwin Schroedinger[1] who was — to my knowledge — the first who stressed the peculiar way in which energy is coupled with entropy in those molecular processes that are crucial in the preservation of life. Since such an energy-order relationship is at the core of my presentation, let me open my remarks with an example which has the advantage of being generally known, but has the disadvantage of being taken from the inanimate world and representing — so to say — a limiting case of what I am going to discuss later.

The example I have in mind is the peculiar way in which in optical masers an "active medium" provides an ordering agent for incoherent electromagnetic waves.

Let me briefly recapitulate the basic principle of this ingenious device[2]. Fig. 1 sketches some energy levels of the sparsely interspersed chromium atoms (about one in thousand) in the ruby, an aluminum oxide in which some of the aluminum atoms are replaced by chromium. These energy levels consist of two broad absorption bands in the green and in the yellow and of a sharp metastable energy niveau corresponding to a wave length of 6,943 Å. If incoherent visible light is shown through a ruby crystal, absorption "pumps" the chromium atoms from the ground state into the excited states (Fig. 1a), from which they quickly return to the metastable state, transmitting the energy difference to lattice vibrations of the crystal (Fig. 1b). Under normal conditions the atoms would remain in this metastable state for a couple of milliseconds before they drop again to the ground state, emitting photons with the afore mentioned wave length of 6,943 Å. The remarkable feature of the maser action, however, is that a photon of precisely the same wave length may trigger this transition which results in a wave train that has not only the same wave length, but is also coherent with the trigger wave. Under certain geometrical conditions which permit the light to bounce back and forth in the crystal, cascading of these events can be achieved.

Since in coherent radiation not the individual energies E_i, but the ampli-

---
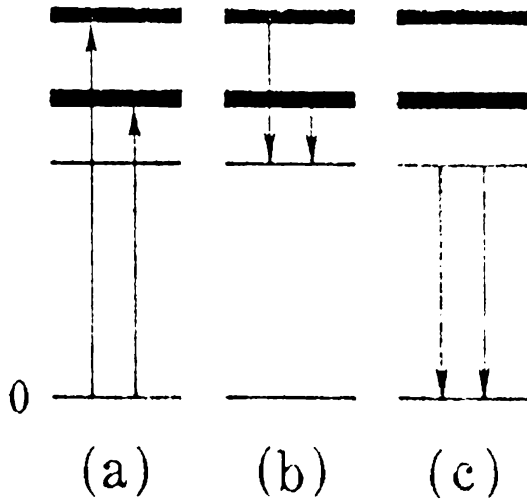
FIGURE 1. Three stages of energy and order in the operation of an optical maser. (a) Excitation by absorption. (b) Ordering. (c) Emission of coherent light.
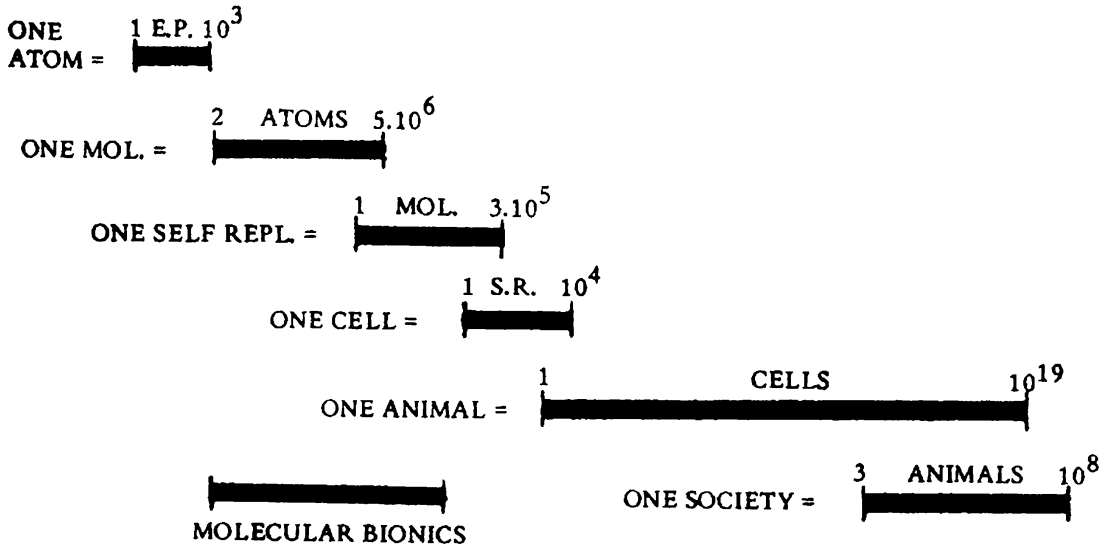
FIGURE 2. Hierarchy of biological structures. The lengths of individual bars correspond to orders of magnitude of the numbers of components which constitute, in turn, the component on the next higher level.

tudes $A_i \sim \sqrt{E_i}$ are additive, it is clear that this ordering process represents from an energy point of view a most effective super-additive composition rule[3] with total output for coherency*

$$E_C = [\, \Sigma \sqrt{E_i} \,]^2 \sim [\, \Sigma A_i]^2,$$

while for incoherent radiation we merely have

$$E_I = \Sigma E_i = \Sigma (A_i)^2$$

with all the crossterms $2A_i A_j$ missing. The salient feature in this process is that disordered energy (incoherent light) after interaction with *an agent with certain intrinsic structural properties* (the chromium atoms in the ruby) becomes a highly ordered affair (coherent light) with little penalty paid in the ordering process (energy loss to the lattice).

I hope I shall be able to demonstrate that in biological processes the order that goes hand in hand with energy is structural in kind rather than coherency of radiation. My main concern will be with those "intrinsic structural properties" of the agent that provides the nucleus for the transformation of disordered energy into organized structures of considerable potential energy. Since these nuclei will be found in the macro-molecular level, and since an understanding of their intrinsic structural properties may eventually permit the synthesis of such macro-molecules, I chose "Molecular Bionics" as the title of my paper. However, in the hierarchy of biological structures molecular bionics considers not only the structural properties of complex molecules, but also the interaction of these molecules in larger biological units, e.g., on the mitochondrial and chromosomal level (self-replicating systems) as suggested in Fig. 2. The inclusion of molecular systems is crucial to my presentation because, as we shall see later, it is the interaction of molecules within these systems which accounts for a transfer of energy to the site of its utilization. But a transfer of energy to a particular site requires tagging this energy parcel with an appropriate address or — in other words — coupling energy with order. It is clear that the efficacy of such a transfer is high if the energy parcel is large and the address code is precise. On the other hand, this is a perverse way in which energy is coupled with order — as far as things go in our universe where energies usually end in a structureless heat pot. In order to escape this fate our molecules have to incorporate into their structures considerable sophistication.

In the following I shall touch upon three features of macro-molecular structures which hopefully will provide us with some clues as to the fascinating

---

*Since in classical electro-dynamics amplitude and energy of a wave are time-averages, this is in no violation of the first law of thermodynamics. The consequence is that incoherent radiation is emitted over a long period while coherent radiation is emitted in an extraordinary small time interval.

behavior of aggregates of such molecules in living matter. These features are:

(1) Storage of information.
(2) Manipulation of information (computation).
(3) Manipulation of information *associated* with energy transfer.

STORAGE OF INFORMATION IN MOLECULES

The most pedestrian way to look at the potentialities of a complex molecule is to look at it as an information storage device.[4] This possibility offers itself readily by the large number of excitable states that go hand in hand with the large number of atoms that constitute such molecules. Consequently the chances are enhanced for the occurrence of metastable states which owe their existence to quantum mechanically "forbidden" transitions[5]. Since being in such a state is the result of a particular energy transaction, selective "read-out" that triggers the transition to the groundstate — as in the optical maser — permits retrieval of the information stored in the excited states.

There is, however, another way to allow for information storage in macromolecules where the "read-out" is defined by structural matching (templet). It is obvious that, m, the number of ways (isomeres) in which n atoms with V valences can form a molecule Z_n will increase with the number of atoms constituting the molecule as suggested in Fig. 3, where the number of valences is assumed to be V = 3. (nV/2 is the number of bonds in the molecule). Estimates of the lower and upper bonds of the number of isomeres are

$$\underline{m} \approx \frac{5}{8} n$$

$$\overline{m} \approx \left(\frac{nV}{2p(V)} \right)^{p(V)}$$

where p(N) is the number of unrestricted partitions of the positive integer N.

Since each different configuration of the same chemical compound Z_n is associated with a different potential energy, the fine-structure of this molecule may not only represent a single energy transaction that has taken place in the past but may represent a segment of the history of events in which this particular molecule has evolved. This consideration brings us immediately to the next point I would like to make, namely a complex molecule's capability to manipulate information.
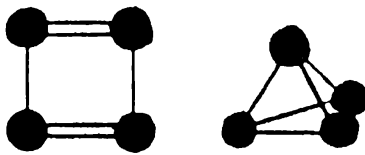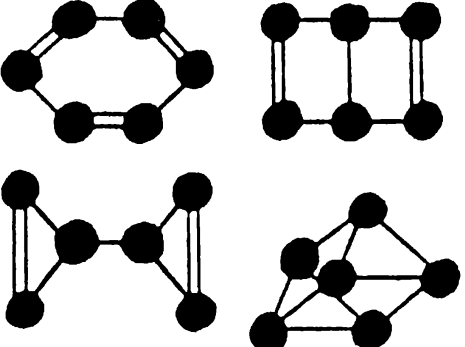
n	$nV/2$	Z_n	CONFIGURATION	m
2	3	Z_2		1
3	9/2			0
4	6	Z_4		2
5	15/2			0
6	9	Z_6		4

FIGURE 3.
Isomeric configurations
of compounds Z_n of n
atoms with valence 3.

MANIPULATION OF INFORMATION IN MOLECULES (MOLECULAR COMPUTATION)

Among the many different ways in which a single macro-molecule may be looked upon as an elementary computing element[6], the most delightful example has been given by Pattee[7]. Assume an ample supply of two kinds of building blocks A, B, which float around in a large pool. The shape of these building blocks is the same for both kinds and is depicted as the wedge shaped element denoted X_{n+1} in Fig. 4. Assume furthermore that a macro-molecule of helical configuration is growing from these building blocks according to a selection rule that is determined by the kind of building blocks that are adjacent to the spaces provided for the addition of the next element. Since 7 blocks define a complete turn, the "selector blocks" are labeled X_n and X_{n-6}. Consider now the following selection rule:

    (a)   Building blocks A are added if the selector blocks are alike (A, A or B, B):
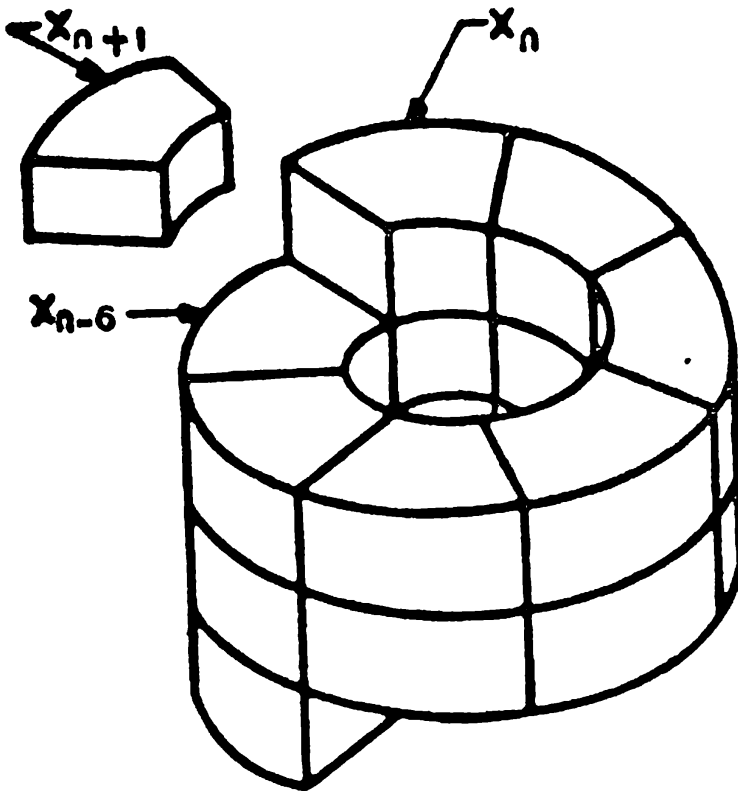
FIGURE 4.
Macromolecular sequence
computer. (Reproduced
with kind permission from
H.H. Pattee, Ref. 7.)

$$x_n = x_{n-6} \longrightarrow x_{n+1} = A$$

(b) Building blocks B are added if the selector blocks are un-
like (A, B or B, A):

$$x_n \neq x_{n-6} \longrightarrow x_{n+1} = B$$

The surprising result of the operation of this simple mechanism is not only that it keeps the helical molecule growing, it also generates a precisely defined periodic sequence of A's and B's with a period of $2^7 - 1 = 127$ symbols. More-over, this sequence is independent of the initial conditions, because any one of the 2^7 state-configurations of a single loop is contained in this sequence. This leads to the important consequence that if the helix is broken at any point or points, the pieces will resume growth of precisely the same sequence as was grown into the mother helix.

These and many other interesting features of this system can be easily derived if one realizes — as Pattee has shown — that this system is logically iso-morphic with a binary feedback shift register (Fig. 5), which is an autonomous
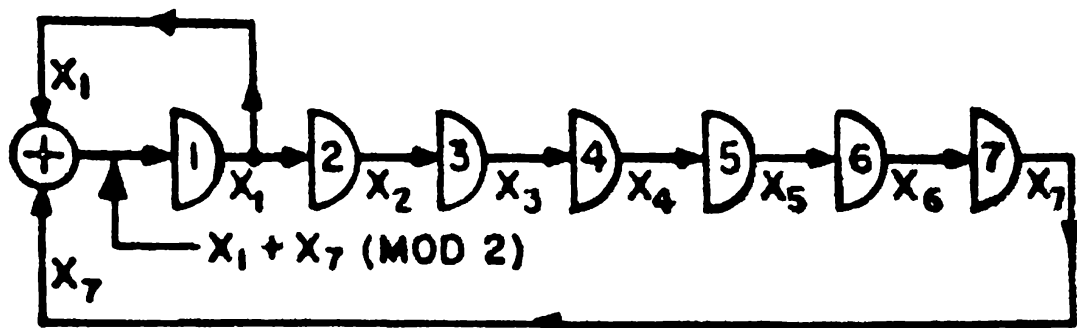
FIGURE 5. Binary feedback shift register. (Reproduced with kind permission from H.H. Pattee, Ref. 7.)

computer, operating on symbols 0, 1 (standing for A and B respectively), generating the afore mentioned maximal period sequence. The function of this shift register can be defined by five operations that in the sense of Turing[8] define the basic steps in all computations. These operations read as follows:

1. **READ** the contents of register 7 (in this case $X_7 = 0$ or 1).
2. **COMPARE** the contents of register 7 with contents of register 1.
3. **WRITE** the result (in this case $X_1 + X_2$ modulo 2) In register 1.
4. **CHANGE** state by shifting all register contents 1 register to the right.
5. **REPEAT** these five steps.

Although this is not the only way in which molecular structures may be thought of as representing elementary computer components, this point of view is representative also of other schemes insofar as in these too the computational mechanism is reduced to a set of rules that are not in contradiction with known properties of large molecules. The fruitfulness of this approach, however, is borne out by the numerous important consequences it yields and at this level of discussion there is no need for detailed account of the physics of these operations.

However, the question arises whether or not inclusion of the energetics of these operations may eventually lead to a deeper understanding of these processes which are associated with self organization and life.

MANIPULATION OF INFORMATION ASSOCIATED WITH ENERGY TRANSFER

Of the many astounding features that are associated with life I shall concern myself only with two, namely, (a) the separation of the sites of production and of utilization of the energy that drives the living organism, and (b) the remarkable efficiency in which the energy transfer between these sites is accomplished. The general method employed in this transfer is a cyclic operation that involves one or many molecular carriers that are "charged" at the site where environmental energy is absorbed and are discharged at the site of utilization. Charging and discharging is usually accomplished with chemical modifications of the basic carrier molecules. One obvious example of the directional flow of energy and the cyclic flow of matter is, of course, the complementarity of the processes of photosynthesis and respiration (Fig. 6). Light energy $h\nu$ breaks the stable bonds of anorganic oxides and transforms these into energetically charged organic molecules. These are burned up in the respiratory process, releasing the energy in form of heat $k\Delta T$ or work $p\Delta v$ at the site of utilization and return again as unorganic oxides to the site of synthesis.
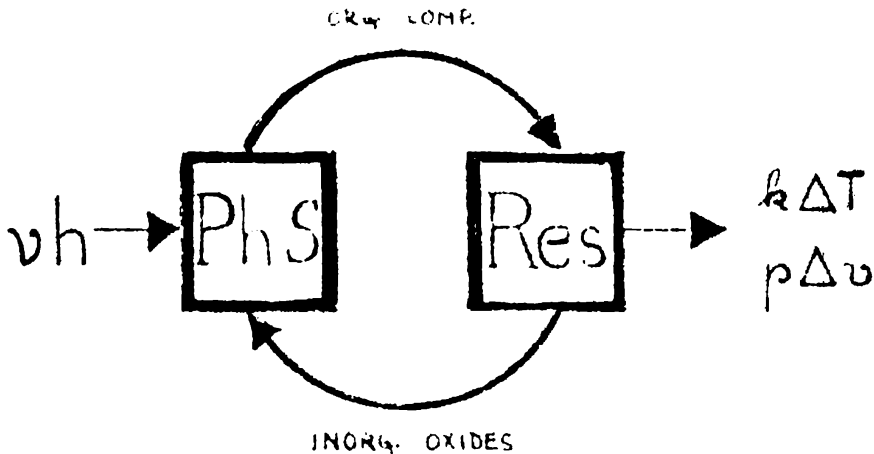
FIGURE 6.   Directional flow of energy and cyclic flow of matter in photosynthesis coupled with respiration.

Another example is the extremely involved way in which in the subcellular mitochondria the uphill reaction is accomplished which not only synthesizes adenosine triphosphate (ATP) by coupling a phosphate group to adenosine diphosphate (ADP), but also charges the ATP molecule with considerable energy which is effectively released during muscular contraction whereby ATP is converted back again into ADP by loosing the previously attached phosphate group.

I could cite many more examples which illuminate the same point, if I would draw on the wealth of information we possess on enzyme reaction. Nevertheless, I hope that these two examples show with sufficient clarity that during the phase of energy release at the site of utilization the charged carrier ejects one of its chemical constituents, maintaining, however, a structure sufficiently complex to undergo a new charging operation. The maintenance of complex structures for reutilization is doubtless a clue to the efficiency of these proceses, but in the following I would like to draw your attention to the interesting interplay of structural and energetic changes that take place in the charging and discharging operations.

To this end permit me to recapitulate briefly some elementary notions on the stability of a molecular configurations. Consider positive and negative ions, say, sodium and chlorine ions, suspended in water. Two ions of opposite charge will attract each other according to a Coulomb force that is inversely proportional to the square of their distance. However, at close distances van der Waal forces produced by the intermeshing of the electron clouds will generate a repulsion that is inversely proportional to approximately the tenth power of their distance. The upper portion of Fig. 7 sketches the force field F as a function of of distance R between two ions with opposite charges. It is clear that a state of equilibrium is obtained at that point where the force between these ions
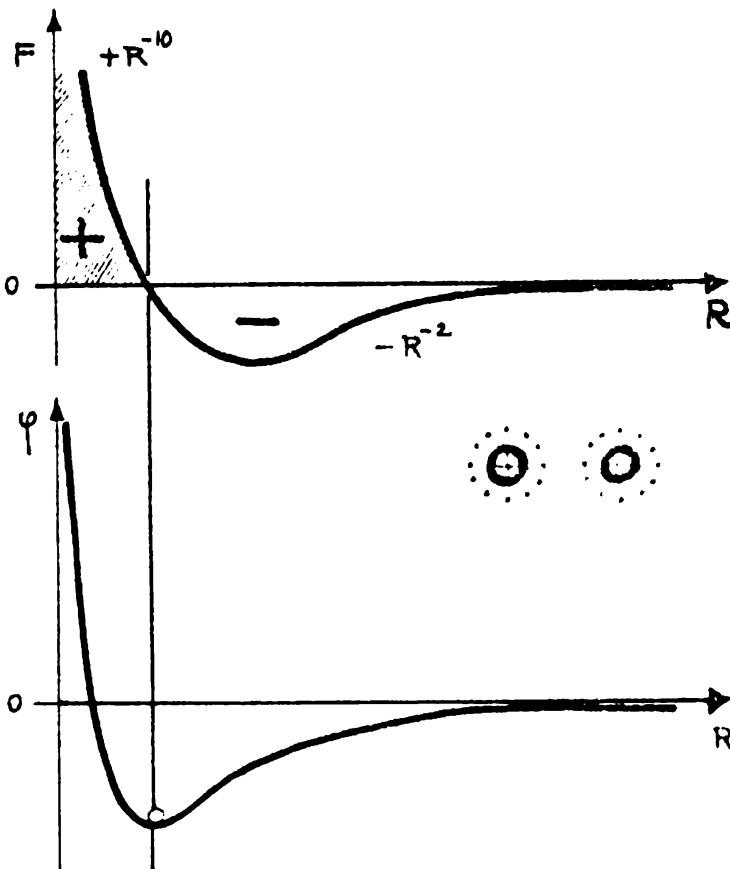
FIGURE 7.
Force field, F, and potential energy, φ, for short range repulsion and long range attraction.

vanishes (F =0). The question of whether or not this state represents a stable equilibrium is easily answered if one considers a small displacement from equilibrium. Since a small separation ($\Delta R > 0$) of an ion pair results in attraction ($\Delta F < 0$), and a small contraction ($\Delta R < 0$) results in repulsion ($\Delta F > 0$), we clearly have stable equilibrium if the conditions

$$F = 0$$

and
$$\frac{\delta F}{\delta R} < 0$$

are fulfilled. This situation is most easily visualized if the potential energy ϕ of the force field F is considered (lower portion of Fig. 7). With

$$F = -\text{grad } \phi = -\frac{\delta \phi}{\delta R}$$

the stability criteria of above become

$$\frac{\delta \phi}{\delta R} = 0$$

$$\frac{\delta^2 \phi}{\delta R^2} > 0$$

This simply states that minima of the potential energy represent stable configurations. It may be noted that owing to the long range attraction the approach of the two ions will take place spontaneously, and that owing to the decrease in potential energy during their approach some energy will be released to the environment. However, this mechanism of energy release is of no use in a system where energy is needed at a particular site. Separation of the two ions would immediately cause their mutual approach and the sites of separation and of recombination cannot be isolated.

Let us turn for a moment to another type of interaction which is characterized by a long range repulsion and a short range attraction. Fig. 8 sketches the force field and potential energy between two protons. For large distances Coulomb forces produce a repulsion according to a $1/R^2$-law, while for close distances nuclear exchange forces result in an attraction which is inverse proportional to roughly the sixth power of their distance. The change in sign of the proton-proton interaction forces defines a point of separation for which the two particles are in equilibrium (F = 0). However, this point does not represent a stable equilibrium as can easily be seen from the distribution of the potential energy which exhibits a maximum at F = 0, and not the minimum required for stability. It is easy to see that this state represents an unstable equilibrium, because a small separation of the two particles will result in repulsion and hence in further separation, and a small contraction in attraction and hence in further contraction. However, it should be noted that separation of the two particles re-
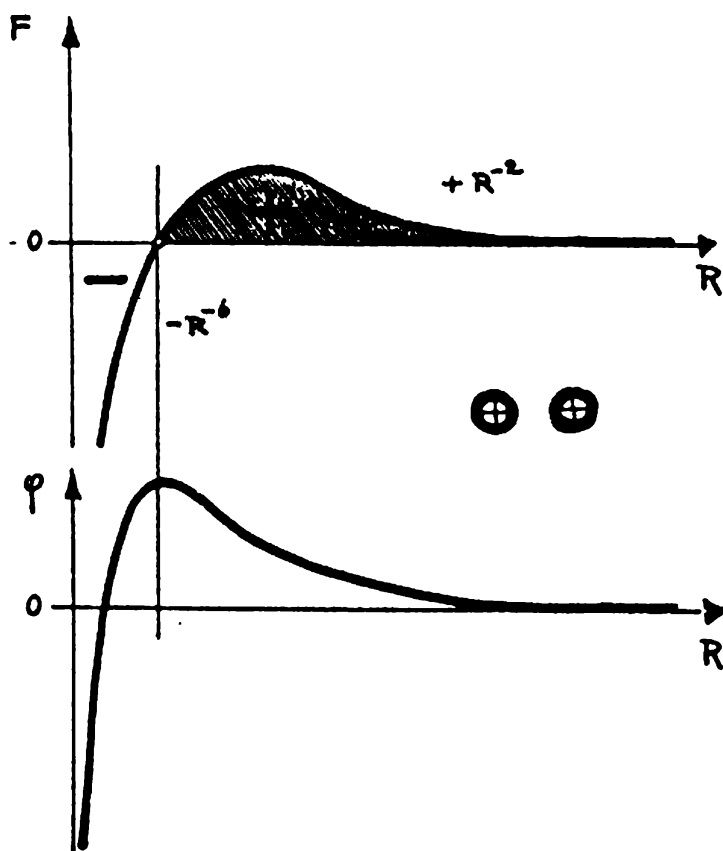
FIGURE 8.
Force field, F, and potential energy, φ, for short range attraction and long range repulsion.

sults in a considerable release of energy, if one succeeds in lifting them over the potential wall. In the proton-proton case this is a project with diminishing return; but for nuclear configurations for which the total energy is above zero and below threshold, a time and site controlled "trigger" may be applied to release what is known today as the "Big Bang."

From this anti-biological example we may obtain some suggestions for the energetics of our organic molecules. Consider a simple molecular configuration with alternating positive and negative ions at the lattice points with, say, positive ions in excess. Two such molecules will clearly repel each other. However, when pushed together, so that lattice points begin to overlap, alternating repulsion and attraction will result, depending upon the depth of mutual penetration. The force field and the potential energy for a pair of "string" molecules with only three lattice points approaching each other lengthwise is sketched in Fig. 9. The potential energy distribution indicates two equilibria, a stable and an unstable one, stability — of course — obtained when the negative centers face the positive edges. It is not difficult to imagine an external

energy source that "pumps" the system of two molecules into the elevated stable energy state by lifting them from the ground state over the potential wall. Since this configuration is stable, the charged system can be transported to an appropriate site where a trigger, that only lifts them from the elevated stable state over the potential wall, causes the release of the stored energy and simultaneously separates the components which can be reutilized for synthesis at a remote site.
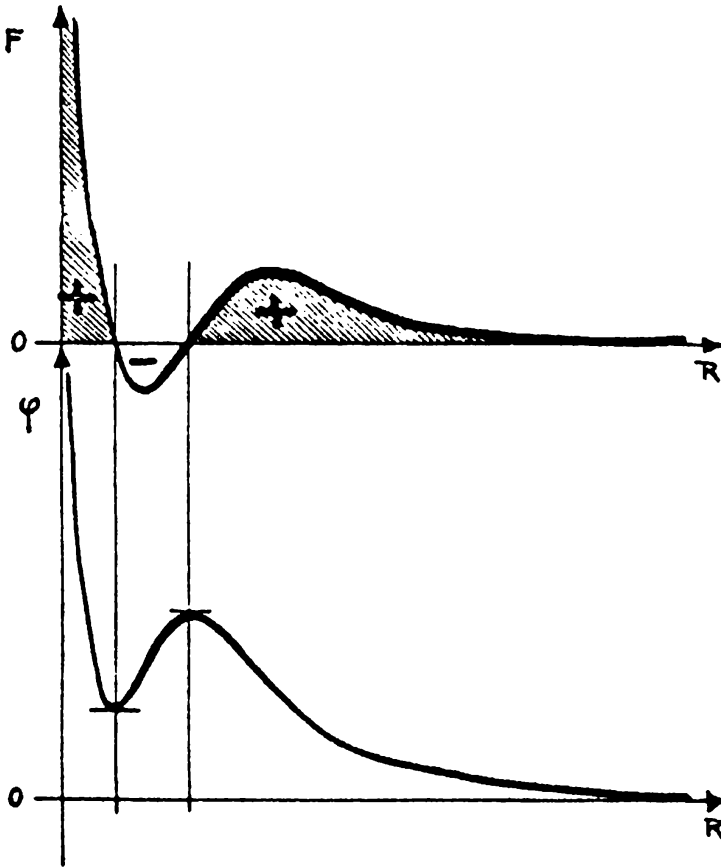
FIGURE 9.
Force field, F, and potential energy, φ, for a molecular system which releases energy by triggered fission.

Although this simple molecular model has doubtless the proper features to act as a mobile energy storage unit, I still owe you the crucial point of my thesis, namely, that this system in the charged state represents sufficient organization to encode the address of the site of utilization. In other words, the question arises how to associate with the energy state of a system the amount of organization that this state represents.

I propose to answer this question by paying attention to the quantum mechanical wave functions which are associated with all energy states of the system that are compatible with its potential energy distribution.

Consider for the moment only the one-dimensional, space dependent part of Schroedinger's wave equation for a particle with mass m in a potential field that varies with distance x according to $\phi(x)$:

$$\frac{d^2\psi}{dx^2} + \frac{2m}{\hbar^2}[E - \phi(x)]\psi = 0$$

Here $\hbar$ stands for $h/2\pi$, with h being Planck's constant, E the enrgy of the system and $\psi(x)$ the "wave-function," whose interpretation can only be given in terms of the product of ψ with its complex conjugate ψ^* representing the probability density dp/dx of finding the particle in the interval $(x) \rightarrow (x + dx)$:

$$\frac{dp}{dx} = \psi.\psi^*$$

Since for any potential distribution the system must be "at least somewhere" we have:

$$\int_{-\infty}^{+\infty} \frac{dp}{dx} \, dx = \int_{-\infty}^{+\infty} \psi.\psi^* \, dx = 1$$

Foregoing all details which refer to physical dimensions we may rewrite the wave equation in a normalized form that expresses only spatial dependence of the wave function $\psi(x)$. Using primes for expressing differentiations with respect to distance the wave equation assumes the following form:

$$\psi'' + [\lambda + \Phi(x)]\psi = 0,$$

with λ and Φ representing normalized energy and potential respectively. The important feature of this equation is that solutions for ψ for a given potential distribution $\Phi(x)$ can only be obtained if λ assumes certain characteristic values, the so-called "eigen-values" λ_i. For each of these "eigen-values" of the energy of the system a certain eigen-function ψ_i for the wave function is obtained. Consequently each eigen-energy λ_i of the system defines a probability density distribution $\left(\frac{dp}{dx}\right)_i$ for finding the system in a configuration that is associated with a distance parameter x. This quantitative association of a certain configuration (characterized by its eigen-energy) with a probability distribution is just the link which I promised to establish between energy and order of charged molecular carriers. This relation is now easily seen if one assumes for a moment that a particular configuration C_c, corresponding to an eigen value λ_c, is associated, say, with a probability distribution that is smeared all over space. Clearly such a configuration is ill defined and is of little use in serving as a highly selective key that fits only into a particular lock. We intuitively associate with this situation

concepts of disorder and uncertainty. On the other hand, if a particular configuration is associated with a probability distribution that displays a sharp peak, we may intuitively associate with this situation order and certainty, because in an overwhelming number of observations we will find our system in just this configuration which corresponds to the peak in the probability distribution. This state of affairs serves excellently for the purpose of highly selective interaction of our system with an appropriate templet.

The question arises of whether our intuitive interpretation of this situation can be translated into precise quantitative terms. Fortunately, the answer is in the affirmative. I refer to Shannon's measure of uncertainty or "entropy" for an information source with a continuous probability density function[8]. In perfect analogy to the definition of the entropy H of a set of n discrete probabilities

$$H = -\sum_{1}^{N} p_i \ln p_i \, ,$$

the entropy H for the one-dimensional continuous case is defined by

$$H = -\int_{-\infty}^{+\infty} \left(\frac{dp}{dx}\right) \ln \left(\frac{dp}{dx}\right) dx \, ,$$

with

$$\int_{-\infty}^{+\infty} \left(\frac{dp}{dx}\right) dx = 1.$$

Since the probability density function for a particular configuration C_i which corresponds to a certain eigen value λ_i can be calculated from Schroedinger's wave equation — provided the potential distribution of the system is given — we are now in the position to associate with any particular configuration C_i a measure of uncertainty H_i, via the wave function $\psi_i(x)$.
Since

$$\left(\frac{dp}{dx}\right)_i = \psi_i \psi_i^*$$

we have:

$$H_i = -\int_{-\infty}^{+\infty} \psi_i \psi_i^* \ln \psi_i \psi_i^* \, dx.$$

In general, a larger value is obtained for the entropy H of an ill defined situation if compared to the entropy of well defined situation. H vanishes for a deterministic system whose probability density function for a certain state is a Dirac delta function. It is easy to define with the aid of this measure of uncertainty a measure of relative order, R, which corresponds precisely to

Shannons measure of a redundancy [9, 10]:

$$R_i = 1 - \frac{H_i}{H_m}$$

The quantity H_m represents the measure of uncertainty of the system when it is in its most ill-defined state. In other words, H_m represents the maximum entropy of the system. Clearly the measure of relative order vanishes if the system is in its most ill defined state $H_i = H_m$, while R is unity for perfect order $H_i = 0$.

We are now in a position to watch the changes in the measure of relative order R_i, while we move through various configurations C_i of the system. Since each configuration is characterized by a particular eigen value λ_i (or eigen-energy E_i) which is compatible with the given potential distribution, my task consists now in showing that for the kind of molecules with a potential distribution that releases energy by triggered fission (Fig. 9), states of higher energy (charged system) represent indeed states with higher relative order.

However, before any such calculation can be approached it is necessary to establish the potential energy distribution Φ of the system. Since even for simple molecules consisting only of a couple of atoms the determination of the potential energy distribution presents almost unsurmountable difficulties, this is a a fortiori the case for complex organic molecules consisting of an extraordinary large number of atoms (see Fig. 2). Nevertheless, a rough guess as to the periodicity of their potential function can be made, a crude approximation of which is given in Fig. 10. The originally smoothly descending curve exhibiting minima and maxima at regular intervals, but with potential troughs that increase monotonously with increasing separation, is here approximated by a series of flat troughs with assymetric perpendicular walls to facilitate solving the wave equation for ψ (x). This can be accomplished by iterating solutions from trough to trough, and matching boundary values at the discontinuities. From these calculations approximations of the eigen values λ_i and the associated wave functions ψ_i have been obtained and two of these results, λ_n, ψ_n, and λ_m, ψ_m, are sketched in Fig. 10.

With these results the probability density distribution ψ_i^2 associated with eigen-function λ_i has been calculated, and the definite integral defining the entropy for the state characterized by λ_i was evaluated. It may be mentioned, however, that the sheer numerical labor involved in these calculations is so great that the crudest approximations have been employed in order to obtain at least
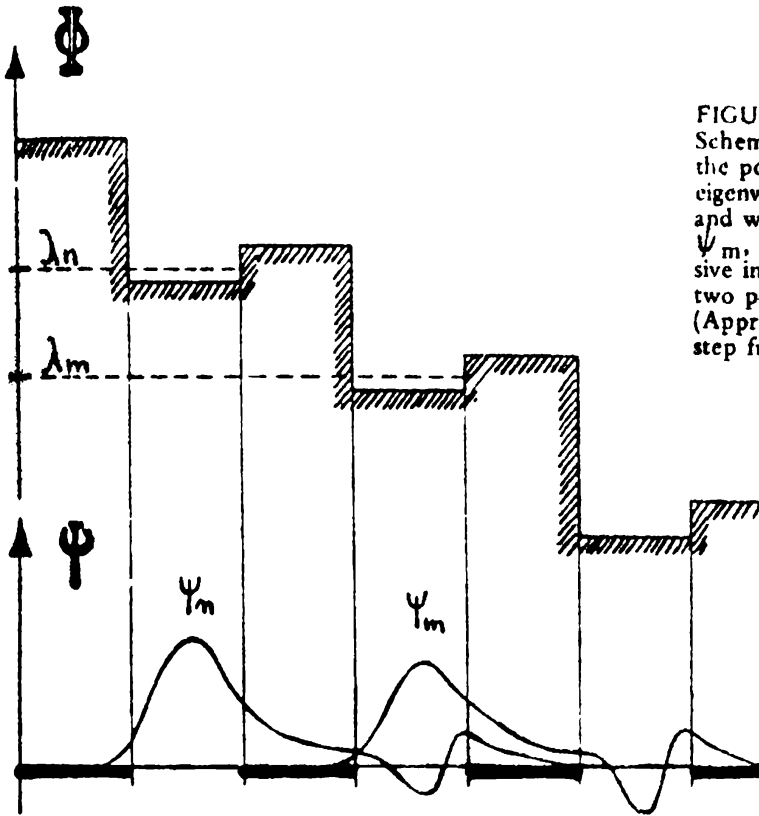
FIGURE 10.
Schematic diagram of the potential energy Φ, eigenvalues λ_m, λ_n and wave functions ψ_m, ψ_n of progressive intermeshing of t two periodic molecules. (Approximation by step functions.)

an insight into the general trend of the desired relationship in these systems.* Consequently, the following results should be taken as assertions of a qualitative argument rather than as a definite numerical evaluation of the proposed problem.

Let me present briefly the results obtained so far. Fig. 11 shows the desired relationship between energy E and relative order R in two kinds of systems. One kind is represented by the potential energy function as shown in Fig. 10, and the energies of its various states are labeled $^+$E. These are the systems which release energy by triggered fission. The other kind may be represented by a potential energy function that is a mirror image with respect to the abscissa of the previous potential function, consequently the energy is labeled $^-$E. These are the systems which release energy by spontaneous or by triggered fusion. Since flipping the potential distribution around the abscissa exchanges

---

*A far more accurate calculation establishing this relationship for a variety of smooth potential functions is presently carried out on Illiac II. The results will be published as Technical Report under the auspices of contract Af 33(657)-10659

stable troughs with instable troughs, we have stability for even maxima in the fission case, and stability for odd maxima in the fusion case. The parameter n represents by its label ascending numbers of particles forming the molecular lattices (n $\gg$ 1).
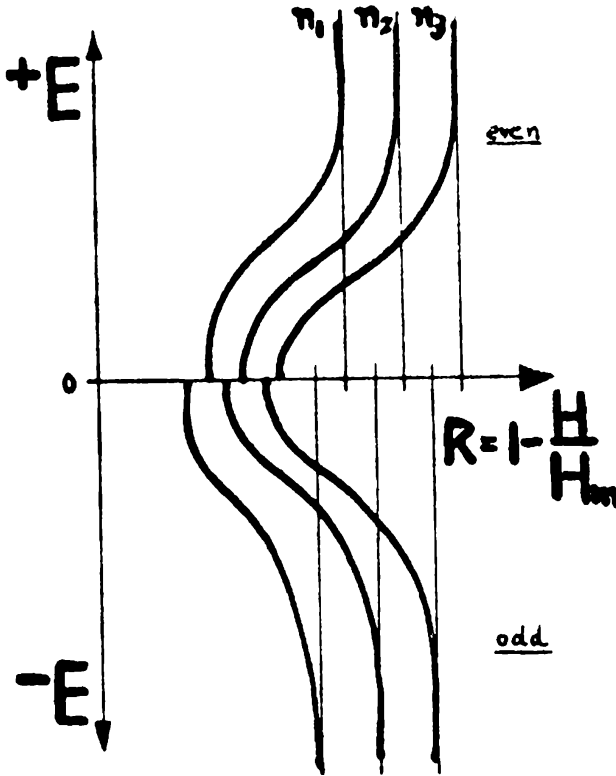
FIGURE 11.
Comparison of the relation between energy and order in molecular structures releasing energy by triggered fission (+) and molecular structures releasing energy by spontaneous or by triggered fusion (−).

Inspection of Fig. 11 clearly shows that for the same amount of energy stored in both systems the fission case ($^+$E) has a considerable edge in its measure of relative order over the fusion case ($^-$E). This gap widens slowly, but detectably, for systems incorporating more and more particles. At first glance the relatively small difference between these systems may be disappointing. However, it may be argued that it is just this small edge in increased orderliness which makes all the difference between a living and a dead system. Moreover, one should not forget that in fission systems the components can be recovered after they have done their work, while in the fusion system the components roll downhill to lower energy troughs with tighter and tighter bonds.

With these observations I hope that I have given sufficient support to my earlier argument in which I postulated a positive relationship between energy

and order in those molecules that are crucial in maintaining the life process. However, one may ask what does this observation do for us?

First, I believe, these observations help us in the clarification of the meaning of self-organizing systems. It is usually contended[11] that we have a truly self organizing system before us, if the measure of relative order increases:

$$\frac{\delta R}{\delta t} > 0.$$

There are, however, limiting cases where this relation holds, − e.g., in the formation of crystals in a super saturated solution − for which we only reluctantly apply the term "self-organization," which we feel inclined to apply for more esoteric systems as, for instance, to a learning brain, a growing ant hill, etc. With our previous observations of the coupling of order with energy, I believe, we should amend the previous definition by the provision that we have a truly self-organizing system before us, if and only if

$$\frac{\delta R}{\delta t} > 0$$

and

$$\frac{\delta R}{\delta E} > 0;$$

that is, if and only if, the measure of relative order increases with time *and* with increase of the system's energy. If one stops and thinks for a moment one may realize that fulfilling simultaneously both conditions is indeed not an easy task.

Second, if we permit our imagination to reign freely, we may think of synthesizing molecules that fulfill the above conditions. Maybe this possibility is not too farfetched if I take von Hippel's prophetic words literraly, who − years ago − coined the term "Molecular Engineering"[12]. He suggested that the time has come when engineering of molecules according to specification may not be out of reach. I can imagine a broth composed of such molecules that may not only grow for us stockings, or other useful tidbits, but may also show us some novel manifestations of life. At this level of speculation, however, everybody is his own best speculator. I only hope that the thoughts which I have presented may serve as a sufficient stimulant to make your speculations worth their while.

REFERENCES

1. Schroedinger, E.: <u>What is Life?</u> The MacMillan Company, New York, (1947).
2. J.A. Bassham: *The Path of Carbon in Photosynthesis*, Scientific American 206, No. 6, (June 1962), 88-104.
3. Von Foerster, H.: *Bio-Logic*, in <u>Biological Prototypes and Synthetic Systems, 1,</u> E.E. Bernard and M.R. Care (eds.). Plenum Press, New York, (1962); pp. 1-12.
4. Von Foerster, H.: *Das Gedaechtnis,* F. Deuticke, Vienna (1948).
5. Szent-Gyoergyi, A.: *Bioenergetics*, Academic Press, New YOrk, (1957).
6. Von Foerster H.: *Quantum Theory of Memory*, in <u>Cybernetics</u>, H. Von Foerster (ed.). Josiah Macy Jr. Foundation, New York, (1950) ; pp. 112-145.
7. Pattee, H.H.: *On the Origin of Macro-molecular Sequences*, Biophys. J., 1, (1961), 683-710.
8. Shannon, C.E. and Weaver, W.: <u>The Mathematical Theory of Communication.</u> The University of Illinois Press, Urbana (1949), p. 54.
9. Ref. (8) p. 25.
10. Von Foerster, H.: *On Self-Organizing Systems and their Environments*, in <u>Self-Organizing Systems</u>, M.C. Yovits and S. Cameron (ed.). Pergamon Press, New York (1960); pp. 31-50.
11. Ref. 10, p. 38.
12. Von Hippel, A.: *Molecular Engineering*, Science
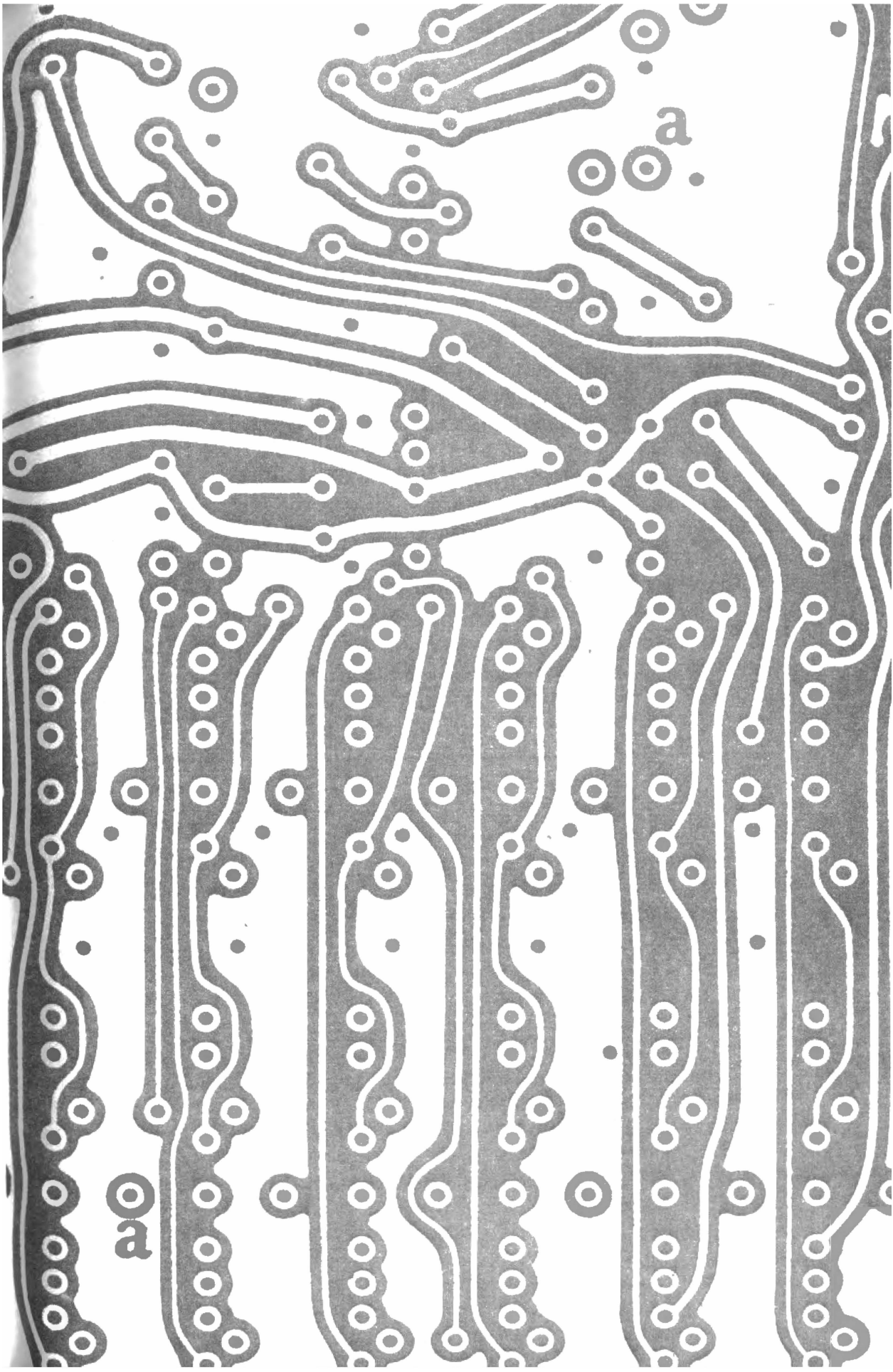
MEMORY WITHOUT RECORD

Heinz Von Foerster

VON FOERSTER:* Perhaps, I should make my position clear by opening with a metaphor. Let me confess that I am a man who is weak in properly carrying out multiplications. It takes me a long time to multiply a two or three digit number, and, moreover, when I do the same multiplication over and over again most of the time I get a different result. This is very annoying, and I wanted to settle this question once and for all by making a record of all correct results. Hence, I decided to make myself a multiplication table with two entries, one on the left (X) and one at the top (Y) for the two numbers to be multiplied, and with the product (XY) being recorded at the intersection of the appropriate rows and columns (Table 15).

TABLE 15

		Y										
X · Y		0	1	2	3	4	5	6	7	.	.	.
	0	0	0	0	0	0	0	0	0	.	.	.
	1	0	1	2	3	4	5	6	7	.	.	.
	2	0	2	4	6	8	10	12	14	.	.	.
X	3	0	3	6	9	12	15	18	21	.	.	.
	4	0	4	8	12	16	20	24	28	.	.	.
	5	0	5	10	15	20	25	30	35	.	.	.
	6	0	6	12	18	24	30	36	42	.	.	.
	7	0	7	14	21	28	35	42	49	.	.	.
	.	.	.	.	.	.	.	.	.	.	.	.
	.	.	.	.	.	.	.	.	.	.	.	.

In preparing this table I wanted to know how much paper I need to accommodate factors X, Y up to a magnitude of, say, n decimal digits. Using regular-size type for the numbers, on double-bond sheets of 8 1/2 x 11 in, the thickness D of the book containing my multiplication table for numbers up to n decimal digits turns out to be approximately

$$D = n \cdot 10^{2n-6} \text{ cm.}$$

For example, a 100 x 100 multiplication table ($100 = 10^2$; n = 2) fills a "book" with thickness

$$D = 2 \cdot 10^{4-6} = 2 \cdot 10^{-2} = 0.02 \text{ cm} = 0.2 \text{ mm.}$$

In other words, this table can be printed on a single sheet of paper.

PRIBRAM: I thought you said you couldn't multiply?

VON FOERSTER: That is true. Therefore, I manipulate only the exponents, and that requires merely addition.

Now, I propose to extend my table to multiplications of ten-digit numbers. This is a very modest request, and such a table may be handy when preparing one's Federal Income Tax. With our formula for D, we obtain for n = 10:

$$D = 10 \cdot 10^{20-6} = 10^{15} \text{ cm.}$$

In other words, this multiplication table must be accommodated on a bookshelf which is 10^{15} cm long, that is, about 100 times the distance between the sun and the earth, or about one light-day long. A librarian, moving with the velocity of light, will, on the average, require a 1/2 day to look up a single entry in the body of this table.

This appeared to me not to be a very practical way to store the information of the results of all ten-digit multiplications. But, since I needed this information very dearly, I had to look around for another way of doing this. I hit upon a gadget which is about 5 x 5 x 12 in in size, contains 20 little wheels, each with numbers from zero to nine printed on them. These wheels are sitting on an axle and are coupled to each other by teeth and pegs in an ingenious way so that, when a crank is turned an appropriate number of times, the desired result of a multiplication can be read off the wheels through a window. The whole gadget is very cheap indeed and, on the average, it will require only 50 turns of the crank to reach all desired results of a multiplication involving two ten-digit numbers.

The answer to the question of whether I should "store" the information of a 10^{10} x 10^{10} multiplication table in the form of a 8 1/2 x 11 in book 6 billion miles thick, or in the form of a small manual desk computer, is quite obvious, I think. However, it may be argued that the

computer does not "store" this information but calculates each problem in a separate set of operations. My turning of the crank does nothing but give the computer the "address" of the result, which I retrieve at once—without the "computer" doing anything—by reading off the final position of the wheels. If I can retrieve this information, it must have been put into the system before. But how? Quite obviously, the information is stored in the computer in a structural fashion. In the way in which the wheels interact, in cutting notches and attaching pegs, all the information for reaching the right number has been laid down in its construction code, or, to put it biologically, in its genetic code.

If I am now asked to construct a "brain" capable of similar, or even more complicated stunts, I would rather think in terms of a small and compact computing device instead of considering tabulation methods which tend to get out of hand quickly.

During this Conference, it has been my feeling that, in numerous examples and statements, you gentlemen have piled up considerable evidence that the nervous system operates as a computer. However, to my great bewilderment—in so far as I could comprehend some of the points of discussion—in many instances you seemed to have argued as if the brain were a storehouse of a gigantic table. To stay with my metaphor, your argument seems to have been whether the symbols in my multiplication table are printed in green or red ink, or perhaps in Braille, instead of whether digit-transfer in my desk computer is carried out by friction or by an interlocking tooth.

I have to admit that, as yet, my metaphor is very poor indeed, because my computer is a deterministic and rigid affair with all rules of operation a priori established. This system cannot learn by experience, and, hence, it should not have been brought up in a conference on "memory."

I shall expand my metaphor, then, by proposing to build a computer that has first to learn by experience the operations I require it to perform. In other words, I posed to myself the problem of constructing an adaptive computer. However, before I attempt to suggest a solution for this problem, permit me to make a few preliminary remarks.

The first point refers to the temptation to consider past experience, again, in terms of a record. This approach offers itself readily because of the great ease with which a cumulative record can be manufactured. One just keeps on recording and recording, usually completely neglecting the problems that arise when attempting to utilize these tabulations. Ignoring this ticklish point for the moment, arguments of how to record may arise. Again, to use my metaphor, one may consider whether ink should be used which fades away after a certain time if it is not reinforced, or whether valid or invalid entries should be made with attached + or - signs, or whether the print should be fat or thin to indicate the importance of the entry, etc. Questions of this sort may

come up if we have the production of the big table in mind. But the kinds of problems are, of course, of an entirely different nature, if we consider building an adaptive computer device in which the internal structure is modified as a consequence of its interaction with an environment.

I believe that many of the remarks made during this meeting addressed themselves to the problem of how to write the record, instead of how to modify the structure of a computer so that its operational modality changes with experience. The interesting thing for me, however, is that these remarks were usually made when the speaker referred to how the system "ought" to work, and not when he referred to how the system actually works. This happened, for instance, when I understood Sir John to have made the remark that, for learning, we need an increase in synaptic efficacy with use. However, as he showed in his interesting example of the reflex action of a muscle that was for a while detached from the bone, these wretched cells would just work the other way round: Efficacy increased with duration of rest. Or, if I remember Dr. Kruger's point correctly, that in order to account for forgetting we need degeneracy of neurons. From his remarks, I understand that it seems to be very hard to get these cells to die.

It is perfectly clear that the comments about what these components ought to do are suggested by the idea that they are to be used in an adaptive recording device. I propose to contemplate for a moment whether the way these components actually behave is precisely the way they ought to behave, if we use them as building blocks for an adaptive computer device. For instance, degeneracy of neurons—if it occurs —may be an important mechanism to facilitate learning, if learning is associated with repression of irrelevant responses, as Sir John pointed out earlier. On the other hand, increased efficacy of a junction, caused by a prolonged rest period, may be used as a "forgetting" mechanism when associated with some inhibitory action.

McCONNELL: I sort of hate to ruin your lovely analogy, but aren't there many cases where a table would be more efficacious? For example, prime numbers?

VON FOERSTER: Correct. But please permit me to develop my story a bit further. I shall soon tighten the constraints on my computer considerably and your comment will be taken up later. Forgive me for developing my metaphor so slowly.

McCONNELL: We're getting caught in your strand.

VON FOERSTER: You shouldn't. Just wait a little and then hit me over the head. I shall give you many occasions, I think.

My second preliminary remark with regard to the construction of an adaptive computer is concerned with the choice of a good strategy of how to approach this problem. Fortunately, in the abstract that was distributed prior to the meeting, Sir John gave us an excellent guideline: "Learning involves selectivity of response, and presumably inhi-

.tion would be significantly concerned in the repression of irrelevant
.esponse." In other words, learning involves selective operations; but,
in order to obtain the results of these operations, a computing device is
.eded to carry out these selective operations.

PRIBRAM: Did you know, Sir John, what you were saying?

ECCLES: No. (Laughter.)

VON FOERSTER: How would you do it, otherwise? How would
you select? How would you have selective manipulation?

ECCLES: How does an animal do it?

VON FOERSTER: I think an impressive array of suggestions
have come up during this meeting. I have in mind, for instance, Sir
John, your demonstration of changes in the efficacy of synaptic junctions
as a result of various stimulations. Instead of interpreting these
changes as storage points for some fleeting events, I propose to inter-
pret them as alterations in the transfer function in a computer element.
In other words, I propose to interpret these local changes as modifica-
tions—admittedly minute—of the response characteristic of the sys-
tem as a whole. In Dr. Uttley's presentation, he considered each
junction as acting as a conditional probability computer element. Dr.
Hyden's most sophisticated scheme is to compute the appropriate pro-
teins with the aid of coded DNA and RNA templets. Of course, the bio-
chemist would probably use the terms "to form" or "to synthesize,"
instead of "to compute," but in an abstract sense these terms are equiv-
alent.

Let me return to the original problem I posed, namely, the con-
struction of a computing device that changes its internal organization
as a consequence of interaction with its environment. I believe it is
quite clear that, in order to make any progress in my construction job,
I have to eliminate two questions: First, what is this environment to
which my computer is coupled? Second, what is my computer sup-
posed to learn from this environment?

As long as I am permitted to design the rules that govern the
events in this environment and the task my computer has to master, it
might not be too difficult to design the appropriate system. Assume
for the moment that the environment is of a simple form that rewards
my computer with an appreciable amount of energy whenever it comes
up with a proper result for a multiplication problem posed to it by the
environment. Of course, I could immediately plug my old desk calcu-
lator into this environment, if it were not for one catch: The number
system in which the environment poses its questions is not specified
a priori. It may be a decimal system, a binary system, a prime-num-
ber product representation, or—if we want to be particularly nasty—
it may pose its problems in Roman numerals. Although I assume my
computer has the Platonic Idea of multiplication built in, it has to
learn the number system in order to succeed in this environment.

My task is now sufficiently specified; I know the structure of the environment, I know what my system has to learn, and I can start to think of how to solve this problem.

Instead of amusing ourselves with solving the problem of how to construct this mundane gadget, let me turn to the real problem at hand, namely, how do living organisms succeed in keeping alive in an environment that is a far cry from being simple.

The question of the environment to which our systems are coupled is now answered in so far as it is nature, with all her unpredictabilities, but also with her stringent regularities which are coded in the laws of physics or chemistry.

We are now ready to ask the second question: What do we require our organisms to learn? Perhaps this question can be answered more readily if we first ask: "Why should these organisms learn at all?" I believe that if we find a pertinent answer to this question we will have arrived at the crux of the problem which brought us here. With my suggestion of how to answer this question, I will have arrived at the central point of my presentation. I believe that the ultimate reason these systems should learn at all is that learning enables them to make inductive inferences. In other words, in order to enhance the chance of survival, the system should be able to compute future events from past experience; it should be an "inductive inference computer." On the other hand, it is clear that only a system that has memory is capable of making inductive inferences because, from the single time-slice of present events, it is impossible to infer about succeeding events unless previous states of the environment are taken into consideration.

I have now completed the specifications of my task: I wish to construct an inductive inference computer whose increase of internal organization should remove uncertainties with respect to predictions of future events in its environment.

Having reached this point, let us look back to the position where we still pitched a computing device against a recording device in order to tackle our memory problem. It is clear from the task I just described that a record of the past, as detailed and as permanent as one may wish, is of no value whatsoever. It is dead. It does not give us the slightest clue as to future events, unless we employ a demon that permanently zooms along this record, computes with lightning speed a figure of merit for each entry, compares these figures in a set of selective operations, and computes from these the probability distribution of the next future events. He must do all this between each instant of time. If we insist on making records, we transfer our problem of memory to the potentialities of this demon who now has the job of acting as an inductive inference computer. Consequently, I may as well throw away the record and consider the construction of this demon who does not need to look at the record of events, but at the events themselves.

I have now concluded the metaphorical phase of my presentation which, I hope, has in some qualitative way outlined my position and my problem. I propose now to consider the problem from a quantitative point of view. In other words, before approaching the actual construction of such an inductive inference computer, it may be wise to estimate, in some way or another, how much internal organization we expect our computer to acquire during its interaction with the environment, and how much uncertainty with respect to future events it is able to remove by its acquisition of higher states of order.

It is fortunate that two decisive concepts in my argument can be defined in precise, quantitative terms. One is the concept of uncertainty, the other, the concept of order. In both cases it is possible to define appropriate measure functions which allow the translation of my problem into a mathematical formalism. Since the whole mathematical machinery I will need is completely developed in what is known today as "Theory of Information," it will suffice to give references to some of the pertinent literature (13, 2, 4) of which, I believe, the late Henry Quastler's account (11) is the most appealing one for the biologically oriented. I have the permission of the Chairman to redefine some of the basic concepts of this theory for the benefit of those who may appreciate having their memories refreshed without consulting another source.

With my apologies to those who will miss rigor in the following shadow of an outline of Information Theory, let me quickly describe some of its basic vocabulary.

The most fundamental step in a mathematical theory of information is the development of a measure for the amount of uncertainty of a situation as a whole, or—as I shall put it—for the uncertainty of a "well-defined universe." The definition of this universe under discourse can be done on several levels. The first step in its definition is to associate with this universe a finite number of distinguishable states which are, also, all the states that the universe can assume. To use some worn-out but illustrative examples, a die with its six faces, or a coin with its two sides, may be considered as such a universe. The face or the side that comes up after the die or the coin is tossed, represents a distinguishable state in these respective "universes." Due to the distinguishability of the individual states, it is possible to label them, say, S_1; S_2; S_3; etc. In general, we may call a state S_i, where i goes through all integers from 1 to n, if our universe is defined by precisely n states. Thus, for the coin:

$$n = 2; \qquad \begin{aligned} S_1 &= \text{heads,} \\ S_2 &= \text{tails,} \end{aligned}$$

and, similary, for the die, n   6, with the names of states S_1 corresponding—for simplicity—to the eyes, i, shown by the die.

As long as we do not deal with a completely deterministic universe, that is, a universe in which for each state there exists one, and only one possible successor state (and our previous examples are certainly not of this type), we intuitively associate with such indeterministic universes a certain amount of uncertainty. For instance, in the coin situation we may say that we are unable to predict the outcome of a particular toss, but we are much less uncertain with respect to the situation as a whole if we compare it with the die situation with its superior variety of possible outcomes. The question of how much uncertainty can be associated with these various situations leads to the second step in the defintion of our universes. Since, clearly, probability considerations will determine these uncertainties, I propose to associate with each state S_i in our universe the probability p_i of its occurrence. Since our universe consists of precisely n states, and hence must be in one of these states at any instant of time, we have, of course, certainty that it is at a given instant of time in any one of its states:

$$p_1 + p_2 + \ldots + p_n = \sum_1^n p_i = 1. \tag{1}$$

In the simple situation of a universe in which all probabilities p_i are alike, say, p_u -- as it is the case for an "honest" coin or an "honest" die -- the equation above is simply

$$n p_i = 1,$$

and the probability for an individual state S_i is just the inverse of the number of states:

$$p_u = p_i = \frac{1}{n} ,$$

or, for our two examples:

$$p_{coin} = \frac{1}{2} , \text{ and } p_{die} = \frac{1}{6} .$$

If we wish now to associate with each universe a measure of uncertainty, it appears, at least, to be plausible that this measure has to take into consideration the probabilities, or the uncertainties, if you wish, of all states that define this universe. In other words, the measure of uncertainty—usually denoted by H—of a particular uni-

verse should be a function of all p_i :

$$H = H (p_1; p_2; p_3; \ldots p_i \ldots p_n).$$ (2)

Since there are infinitely many functions from which we may choose one, as, for example,

$$\sum_1^n p_i^2 \; ; \quad \sum_1^n e^{pi} \; ; \quad \prod_1^n \log p_i \; ; \quad \text{etc., etc., } \ldots .$$

we are in a position to introduce certain conditions which we intuitive-ly like to see fulfilled in our final choice for a measure of the uncertainty of a universe. It cannot be stressed strongly enough that the choice of these conditions is more or less arbitrary, their justifications being solely confined to their implications.

One of these conditions may reasonably be that the measure of these uncertainties should, in a sense, reflect our intuition of the amount of these uncertainties. In other words, more uncertainty should be represented by a higher measure of uncertainty.

Another condition may be that the measure of uncertainty should vanish ($H = 0$) for a deterministic universe, that is, for a universe in which there are no uncertainties.

Finally, we may propose that the measure of uncertainty for two independent universes, U_1 and U_2 , should be the sum of the measures of uncertainty of each universe separately. In mathematical language, this is

$$H (U_1 \& U_2) = H (U_1) + H (U_2) .$$ (3)

From this last condition we may get a hint as to the form of our measure function. Consider for a moment our two examples, the coin and the die. We wish that the measure that expresses the uncertainity of a universe composed of a coin and a die equals the sum of the measures that express the uncertainties of the coin-universe and of the die-universe. Since, in the combined universe, the number of states n_{CD} is the product of the number of states of the component universes n_C, n_D :

$$n_{CD} = n_C \cdot n_D = 2 \cdot 6 = 12;$$

and since all probabilities are equal:

$$P_{CD} = \frac{1}{n_{CD}} = \frac{1}{n_C \cdot n_D} = P_C \cdot P_D ,$$

we have with a postulated addition theorem of independent universes:

$$H (P_{CD}) = H (P_C \cdot P_D) = H (P_C) + H (P_D)$$

This relation states that the desired measure function, H, taken of the product of two factors, should equal to the sum of the measure function of the factors   There is, essentially. only one mathematical function that fulfills this condition, namely, the logarithmic function. Consequently, we put tentatively:

$$H (p) = k \cdot \log (p) ,$$

with k being an as yet undetermined constant. We verify the relation above in $H (P_{CD})$:

$$H (P_{CD}) = k \log (P_{CD}) = k \log (P_C \cdot P_D)$$

$$= k \log (P_C) + k \log (P_D)$$

$$= H (P_C) + H (P_D) .$$

<div align="right">Q. E. D.</div>

Since, on the other hand, we intuitively feel that a universe that can assume more states than another is also associated with a higher degree of uncertainty, we are forced to give our as yet undetermined constant k a negative sign:

$$H (p) = -k \cdot \log (p) = k \cdot \log \left(\frac{1}{p} \right) = k \cdot \log (n),$$

In the simple situation of universes whose states are equiprobable, we have come to the conclusion that an adequate function that can be used as a measure of uncertainty is of logarithmic form:

$$H (p) = k \cdot \log (n) = - k \cdot \log (p). \qquad (4)$$

However, situations in which all states are equiprobable are relatively rare, and we have to consider the general case in which each

state S_i is associated with a probability p_i which may or may not be equal to the probability of other states. Since, in the equiprobable situation the uncertainty measure turned out to be proportional to the log of the reciprocal of the probability of a single state, it is suggestive to assume that, in the general case, the uncertainty measure is now proportional to the <u>mean value</u> of the log of the reciprocal of each individual state:

$$H = -k \cdot \overline{\log p_i},\tag{5}$$

the bar indicating the mean value of log (p_i) for all states n.

The calculation of a mean value is simple. Take N sticks consisting of n groups, each of which contains N_i sticks of length l_i. What is the mean value of their length? Clearly, the total length of all sticks, divided by the number of sticks:

$$\bar{l} = \frac{\sum\limits_{1}^{n} N_i \, l_i}{N}$$

Call p_i the probability of the occurrence of a stick with length l_i :

$$p_i = N_i \,/\, N,$$

and the expression for the mean length becomes

$$\bar{l} = \sum\limits_{1}^{n} \frac{N_i}{N} \cdot l_i = \sum\limits_{1}^{n} p_i \cdot l_i \,.$$

In other words, the mean value of a set of values is simply obtained by the sum of the products of the various values with the probability of their occurrence. Consequently, the mean value of the various values of log p_i is simply:

$$\overline{\log p_i} = \sum\limits_{1}^{n} p_i \, \log p_i \,,$$

and the uncertainty measure becomes

$$H = -k \, \overline{\log p_i} = -k \sum\limits_{1}^{n} p_i \, \log p_i$$

This measure function fulfills all that we required it to fulfill. It reduces to the simple equation for equiprobable states, because with

$$p_i = \frac{1}{n} \, ,$$

$$H = -k \sum_1^n \frac{1}{n} \log \frac{1}{n} = -k \, \frac{n}{n} \log \frac{1}{n} = k \ln n;$$

consequently, H increases in a monotonic fashion with increasing number of equiprobable states. Furthermore, H vanishes for certainty, which we shall express by assuming only one state to appear with certainty, say $p_1 = 1$, while all others have probability 0. Consequently,

$$H = -k \left[1 \cdot \log 1 + \sum_2^n 0 \cdot \log 0 \right] = 0 \, ,$$

because

$$1 \log 1 = 0, \text{ and also}$$

$$0 \log 0 = 0.$$

Since the latter expression is not obvious, because $\log 0 = -\infty$, we quickly show with l'Hospital's rule that

$$\lim_{x \to 0} (x \cdot \log x) = 0 :$$

$$\lim_{x \to 0} (x \cdot \log x) = \lim \frac{\log x}{\left(\frac{1}{x}\right)} = \lim \frac{\frac{d \log x}{dx}}{\frac{d}{dx} \left(\frac{1}{x}\right)}$$

$$= \lim \frac{1/x}{-1/x^2} = \lim_{x \to 0} (-x) = 0. \qquad \text{Q. E. D.}$$

Finally, we show that the addition theorem for independent universes is preserved, also. Let p_i and p_j be the probabilities of states of two universes, U_1 and U_2, respectively. The probabilities of the

combined universe are

$$p_{ij} = p_i \cdot p_j .$$

Let n and m be the number of states corresponding to U_1 and U_2. For U_1 and U_2 the number of states is $n \cdot m$. The uncertainty of the combined universe is

$$H (U_1 \& U_2) = -k \sum_{ij} p_{ij} \log (p_{ij})$$

$$= -k \sum_{ij} p_i \cdot p_j \log (p_i \cdot p_j)$$

$$= -k \sum_{ij} p_i \cdot p_j \log (p_i) -k \sum_{ij} p_i \cdot p_j \log (p_j)$$

$$= -k \sum_j p_j \sum_i p_i \log (p_i) -k \sum_i p_i \sum_j p_j \log (p_j) .$$

Since $\displaystyle\sum_1^n p_i = \sum_1^m p_j = 1$, the above expression becomes

$$H (U_1 \& U_2) = -k \sum_1^n p_i \log (p_i) - k \sum_1^m p_j \log (p_j)$$

$$= H (U_1) + H (U_2) .$$

Q. E. D.

The only point that remains to be settled in our uncertainty measure H, is an appropriate choice of the as yet undetermined constant k. Again, we are free to let our imagination reign in adjusting this constant. The proposition now generally accepted is to adjust this constant so that a "single unit of uncertainty," usually called "one bit," is associated with a universe that consists of an honest coin. A pedestrian interpretation of this choice may be put forth by suggesting that a universe with just two equiprobable states is, indeed, a good standard with some elementary properties of uncertainty. A more sophisticated argument in favor of this choice is connected with problems of optimal coding (11). However, in this framework, I have no

justification for elaborating on this issue. Let us, therefore, accept the previous suggestion and give H the value of unity for the measure of uncertainty that is associated with a universe consisting of an honest coin:

$$H_{coin} = 1 = -k \left[\frac{1}{2} \log \frac{1}{2} + \frac{1}{2} \log \frac{1}{2} \right]$$

$$= k \log 2 .$$

Hence,

$$k = \frac{1}{\log 2} ,$$

and

$$H = - \frac{1}{\log 2} \sum_{1}^{n} p_i \log p_1 ,$$

or, if we take 2 as the basis of our logarithmic scale:

$$H = - \sum_{1}^{n} p_i \log_2 p_i . \tag{6}$$

With this expression, we have arrived at the desired measure of the uncertainty of a universe that is defined by n states S_i, which occur with probability p_i. It may be worthwhile making a few comments to illustrate some properties of this measure function. First, I would like you to appreciate that for a universe with a fixed number of states, n, the uncertainty measure H is maximum, if all states occur with equal probability. A shift away from this uniform probability distribution immediately reduces the amount of H; in other words, reduces the uncertainty of the universe. Let me illustrate this with a die that is born "honest" but "corrupts" as a consequence of its interaction with bad society. My victim is a die made of a hard cubical shell filled with a highly viscous glue, in the center of which is placed a heavy steel ball. Since there is perfect symmetry in this arrangement, when tossed, the die will go with equal probabilities into its six possible states. However, I am going to teach this fellow to show a preference for the side with one eye. To this end, I place under the table an electromagnet and, whenever one eye comes up, I give the die a short magnetic shock. This moves the steel ball slightly toward the bottom, and gravitation will enhance the chance of its falling at the same side the next time. Table 16 lists

TABLE 16

TIME	p_1	p_2	p_3	p_4	p_5	p_6	H bits
t_0	1/6	1/6	1/6	1/6	1/6	1/6	$2.582 = \log_2 6$
t_1	1/4	1/6	1/6	1/6	1/6	1/12	2.519
t_2	1/3	1/6	1/6	1/6	1/6	0	2.249
t_3	2/3	1/12	1/12	1/12	1/12	0	1.582
t_4	1	0	0	0	0	0	0.000

the probability distribution for the various states as it may look in succeeding intervals, t_0, t_1, t_2, t_3, t_4, between shock treatments. The right hand margin gives the values of H, the uncertainty of this universe, corresponding to the probability distribution at successive states, H is given in "bits," and is calculated according to Equation (6).

Another feature of the uncertainty measure H is that changing from a universe with n equiprobable states to another universe with twice as many equiprobable states, 2n, the uncertainty increases exactly one bit:

$$H_1 = \log_2 n ,$$

$$H_2 = \log_2 2n = \log_2 2 + \log_2 n = 1 + H_1 .$$

Thus, a universe with 1 million states has an uncertainty of about 20 bits. Add 1 million states, and this new universe has about 21 bits uncertainty.

Up to this point, I have referred to our measure function H always as a measure of uncertainty. However, a variety of terms are in use which all refer to the same quantity H as defined in Equation (6). These terms are "entropy," "choice," and "amount of information."

To call H the entropy of a universe, or of a system, is justified by the fact that this thermodynamical variable, when expressed in

terms of the probability distribution of the molecules comprising a thermodynamical system, is defined by an equation almost identical to our Equation (6) for H. The importance usually given to the concept of entropy stems from one of the consequences of the second law of thermodynamics, which postulates that in a closed thermodynamical system the entropy must either remain constant (for thermal equilibrium), or go up, but it can never decrease. This is an expression of the fact that, in "natural systems," the distribution of the probabilities of states tends to uniformity, as exemplified by a bucket of hot water in a cold room. After a while, the thermal energy of the bucket will distribute itself more or less uniformly over the room (equilibrium; all p_i alike; H is maximum). Consequently, a thermodynamicist, unaware of my magnetic contraption, who watches my die violating the second law of thermodynamics as it slowly, but surely, moves from high values of entropy to smaller and smaller ones, will come to the conclusion that a Maxwellian demon is at work who alters selectively the internal organization of the system. And he is so right! I, of course, am the demon who selectively switches on the magnet whenever the die shows one eye.

Sometimes, H, as defined in Equation (6), is referred to as the amount of choice one has in a universe consisting of n items, in which one is permitted to pick items with a probability p_i associated with item S_i. All the considerations of intuitive nature which helped us to define a measure function for uncertainty—in particular, the addition theorem—can as well be applied to a measure function of choice. Consequently, the resulting function, expressing a measure of choice, is identical with the function expressing a measure of uncertainty—even with regard to the units, if a unit of choice is associated with a well-balanced temptation between either one of two choices, as is illustrated so beautifully by Burdidan's Ass.

Finally, H is also associated with an "amount of information" in a situation in which the actual state of a universe, whose uncertainty is H , is transmitted by an observer to a recipient. Before the recipient is in possession of the knowledge of the actual state of the universe, his uncertainty regarding this universe is H. The question arises as to how much he values the information about the actual state of the universe when transmitted to him by the observer. All the considerations of intuitive nature that helped us to define a measure of uncertainty are applicable here, too, and the resulting function expressing a measure of information is identical with H as defined in Equation (6). Since, in a communication situation, the "states" in question are usually symbols, H is usually referred to in bits per symbol. If the observer transmits symbols at a constant rate, H may also be expressed in bits per second.

If, in our vocabulary—presumably consisting of about 8,000

words—we were to use each word with equal probability, our linguistic universe would have an uncertainty H of 13 bits (H = $\log_2$ 8192 = 13). However, due to our using various words with different fre=quencies, the uncertainty of our linguistic universe is somewhat smaller and has been measured to be about 11 bits (12). Consequently, whenever I utter a word, I transmit to you, on the average, 11 bits of information. Since I utter approximately three words/sec, my rate of generating information is about 33 bits/sec. (I hope nobody will tell me that what I generate is not information but noise.)

I hope that in this short outline I have been able to show that one and the same expression, namely, - $\sum p_i \log_2 p_i$, represents a measure for various concepts in various situations. This is reflected in the various names for this expression as, for instance, uncertainty, entropy, choice, and information. This state of affairs may well be compared to mechanics, where the product of a force and a length represents "work" in one context, but "torque" in another context.

The next step in my development is to use these quantitative concepts to construct still another measure function, this time a measure of "order." Again, I shall be guided by intuitive reasoning when selecting from the vast amount of possible measure functions the one that fulfills some desired criteria. First, I would like to suggest that whenever we speak of "order" we mean it in a relative sense, that is, we refer to the state of order of a particular universe. We say that a room is in various states of order or disorder, or that a desk is a mess, and so on. Hence, if we wish to state that a given "universe" is in complete disorder, our function representing a measure of order should vanish. Conversely, perfect order may be represented by a value of unity. Consequently, various states of order of a given universe may be represented by any number between zero and one. Further hints as to the general form of the function that expresses an order measure may be taken from the truism that our uncertainty H about a completely disordered universe is maximum (H = H_{max}), while, in a deterministic universe, (H = 0) order is perfect. This suggests that a measure of order an observer may associate with a particular universe is just the difference between his actual and maximum uncertainty of this universe in reference to maximum uncertainty. Accordingly, we define tentatively, Ω , a measure function of order by:

$$\Omega = \frac{H_{max} - H}{H_{max}} = 1 - \frac{H}{H_{max}} \qquad (7)$$

Clearly, the two conditions as discussed above are fulfilled by this function, because for H = H_{max} the order measure will vanish (Ω = 0), while for perfect order the uncertainty vanishes and the order measure approaches unity (Ω = 1).

I propose to further test this expression, now under addition. Assume two universes, U_1 and U_2, with which we may associate measures of actual and maximum uncertainty, H_1, H_{m1} and H_{m2}. We further assume both universes to be in the same state of order:

$$\Omega_1 = 1 - \frac{H_1}{H_{m1}} = \Omega_2 = 1 - \frac{H_2}{H_{m2}} ,$$

or

$$\frac{H_1}{H_{m1}} = \frac{H_2}{H_{m2}} .$$

This condition is fulfilled if

$$H_2 = k \, H_1 ,$$

and

$$H_{m2} = k \, H_{m1} .$$

I now propose to drop the distinction between the two universes and treat both as parts of a large universe. What is the measure of order for this large universe? With Equation (7) defining Ω, Equation (3) the addition theorem for H, and the above identities, we have

$$\Omega = 1 - \frac{H_1 + H_2}{H_{m1} + H_{m2}} = 1 - \frac{H_1 (1 + k)}{H_{m1} (1 + k)}$$

$$= 1 - \frac{H_1}{H_{m1}} = \Omega_1 = \Omega_2 .$$

In other words, by combining two equally ordered universes, the measure of order remains unchanged, as it intuitively should be.

This measure of order will be helpful in making quantitative estimates of the changes in the internal organization of our inductive inference computer during its interaction with the environment. Presently, let me give you a few examples of systems whose order increases at the cost of external or internal energy consumption. I have already given one example: the magnetic die. Table 17 lists the values of Ω for the various time intervals and uncertainties given in Table 16.

TABLE 17

TIME	t_0	t_1	t_2	t_3	t_4
H	2.582	2.519	2.249	1.582	0.000
Ω	0.000	0.025	0.129	0.387	1.000

Another example may be the growth of crystals in a supersaturated solution. Again, the organization of the system increases as the diffused molecules attach themselves to the crystal lattice, reducing in this process, however, their potential energy.

PRIBRAM: This poses a problem, because the biologist is less interested in this sort of order than he is in the one that involves sequential dependencies. This is why a hierarchical measure of some sort would be more appropriate, unless you can show how you can derive it from your equation.

VON FOERSTER: My measure of order is so general that, I believe, there shouldn't be any difficulty in dealing with the type of order that arises from sequential dependence. Permit me to suggest how I think this may be done. Sequential dependencies express themselves in the form of transition probabilities p_{ij}, that is, the n^2 probabilities for a system that is in state S_i to go into state S_j. If the states of a system are independent of previous states, as is the case with the coin and the die, all p_{ij}'s are, of course, just the p_j's. However, a system that learns will develop strong sequential dependencies as you suggested, and, hence, will shift the p_{ij}'s away from the p_j's. Again, a measure of uncertainty H can be defined for this state of affairs simply by working with the mean value of the various uncertainties H_i, as they can be computed for all states that immediately follow state S_i. Since

$$H_i = - \sum_{j=1}^{n} p_{ij} \log_2 p_{ij} \quad ,$$

we have our rule of developing a mean-value:

$$H = \overline{H}_i = \sum_{1}^{n} p_i H_i \quad .$$

If n, the number of states of the system, remains the same, $H_{max} = \log_2 n$ is unchanged and we may watch how Ω increases steadily as the sequential constraints are going up—that is, as H is going down—while the animal is being trained.

Since changing the internal organization of a system, whether in a spatial or temporal sense, to higher and higher levels of organization is a crucial point in my description of so-called "self-organizing systems" that map environmental order into their own organization, let me establish the criteria which have to be satisfied if we wish our system to be such a self-organizing system. Clearly, for such systems, Ω should increase as time goes on, or:

$$\frac{d\Omega}{dt} > 0.$$

Since our measure of order is a function of H and H_{max}, both of which may or may not be subjected to changes, we obtain the desired criterion by differentiating Equation (7) with respect to time:*

$$\frac{d\Omega}{dt} = -\frac{H_m \frac{dH}{dt} - H \frac{dH_m}{dt}}{H_m^2} > 0.$$

This expression can be transformed into something more tangible. First, we note that for all systems of interest, $H_m > 0$, because only for systems capable of precisely one state $H_m = \log_2 1 = 0$. Second, we divide both sides of the unequality with the product $H \cdot H_m$ and obtain the important relation:

$$\frac{1}{H_m} \frac{dH_m}{dt} > \frac{1}{H} \frac{dH}{dt} \tag{8}$$

This says that if, and only if, the relative increase of maximum uncertainty is larger than the relative increase of the actual uncertainty, then our system is in the process of acquiring higher states of internal organization.

BOWER: Are empirical considerations operating here?

VON FOERSTER: No, not a single empirical consideration. This is a straightforward derivation starting from one definition and using one criterion.

---

*For typographical reasons H_{max} will be written H_m.

BOWER: I can think of several counter-examples in which, although something is being learned, the behavior is becoming increasingly more random or disordered, in your sense.   For example, I am sure I could train a rat to vary his behavior from trial to trial in a way that is completely unpredictable to me; one simply reinforces variance differentially according to some criterion—for example, by reinforcing a rat's lever press only if its latency differs by at least 2 sec from the preceding lever press.   Second, in extinction of learned behavior, the reference response declines in strength and occurs less certainly.   You would describe it as increasing disorder, yet it is a lawful and uniform process, particularly so when competing responses are recorded.

VON FOERSTER: I would say that in such a situation training enlarges the behavioral capacities of the rats by creating new states in the rat's behavioral universe.   Hence, you operate on H_{max} such that dH_m/dt is larger than zero.   If these animals do not deteriorate otherwise, by letting H go up too fast, you have indeed taught them something.

JOHN: Would you define H_{max}, please?

VON FOERSTER: H_{max} can be defined simply as the uncertainty measure of a system with equiprobable states that are also independent of their precursor states.   Under these conditions, as we have seen earlier, H_{max} is just $\log_2 n$, where n is the number of states.

JOHN: Precisely.   Therefore, it seems to me that the derivative of H_m with respect to time must always be zero.

VON FOERSTER: That is an excellent suggestion.   You pointed out one of the fascinating features of this equation, namely, that it accounts for growth.   You have already anticipated the next chapter of my story.

To see immediately that dH_m/dt must not necessarily always be zero, let me replace H_m by $\log_2 n$, or, for simplicity, by a ln n, where a = 1/ln 2 is a scale factor converting base 2 log's into natural log's.   Then:

$$\frac{dH_m}{dt} = a \frac{d \ln n}{dt} = \frac{a}{n} \frac{dn}{dt} \quad . \tag{9}$$

Consider for the moment an organism that grows by cell division.   In the early stages of development the number of cells usually grows exponentially:

$$n = n_o e^{\lambda t} \quad .$$

and, with

$$H_m = a \ln n = a \ln n_0 + a \lambda t , \qquad (10)$$

the rate of change of maximum entropy becomes

$$\frac{dH_m}{dt} = a \lambda ,$$

which is a positive constant! This means that the actual entropy H of the system must not necessarily decrease in order for the system to acquire higher states of organization According to Equation (8), it is sufficient that the relative rate of change of H remains just below that of H_{max} . An observer who pays attention only to the increase of the actual uncertainty H may get the impression that his system goes to pot. If we look at our cities today we may easily get this impression. However, if we consider their rapid growth, they represent centers of increasing organization. But, to go back to the population of dividing and differentiating cells: Organized growth of tissue represents considerable constraints on the possible arrangement of cells, hence, the probability distribution of their position is far from uniform. Consequently, H changes during the growth phase of the organism very slowly indeed, if at all. Nevertheless, a conservative model for a tentative expression for the growth of H after Equation (10) is:

$$H = a \ln n_0 + \mu a \lambda t , \qquad (11)$$

where $\mu < 1$. The measure of order for the growing organism becomes, with Equations (10) and (11),

$$\Omega = 1 - \frac{\ln n_0 + \mu \lambda t}{\ln n_0 + \lambda t} .$$

At early stages of its development (t = 0), we have

$$\Omega (0) = 0 ,$$

while at its mature state ($t \to \infty$) :

$$\Omega (\infty) = 1 - \mu > 0 ;$$

that is, the organism is indeed an " organism "

The possibility of accounting for the acquisition of higher states of organization by incorporation of new states into the system—provided that this incorporation takes place in an orderly fashion—has been illuminated for me by the beautiful work reported here by Dr. Hydén. Let us take the neuron nucleus as the universe under consideration. The organizational increase can be quite formidable, as can be seen from Equation (9), which sets the absolute rate of maximum uncertainty proportional to the percentage rate of increase in the number of ordered elements. Since there are only 80,000 molecules in the nucleus, only a couple of thousand molecules, modified according to Dr. Hyden's ingenious mechanisms, may add substantially to the Ω of the system.

These have been somewhat general remarks about the use of numbers in describing systems in various states of order, expressing an amount of uncertainty, complexity, or "perplexity," as one of my students suggested, associated with these systems and, finally, the amount of information that is required to specify them. I will return now to the topic of our Conference where these numbers may become useful.

At issue is an important property of the functioning of our nervous system. We call it "memory." In looking for mechanisms that can be made responsible for this property, I strongly suggested that we not look upon this system as if it were a recording device. Instead, I have proposed looking at this system as if it were a computer whose internal organization changes as a result of its interaction with an environment that possesses some order. The changes of the internal organization of this computer take place in such a way that some constraints in the environment which are responsible for its orderliness are mapped into the computer's structure. This homomorphism "environment-system" reveals itself as "memory" and permits the system to function as an inductive inference computer. States of the environment which are, so to say, "incompatible with the laws of nature," are also incompatible with output-states of the computer.

I am going to apply now the numerology of information theory to some of the known features of the nervous system, and we shall see what kind of conclusion can be derived from these numbers. The first number I am going to derive is an estimate of the amount of information necessary to specify a "brain." As I pointed out earlier, in order to make any progress in making such estimates, one has first to specify the "universe"—in this case it is the brain—in terms of a finite number of states and the probability of their occurrence. If this is done, H (Brain) can be calculated from Equation (7). To this end, I suggest interpreting "brain" as a set of finite number of elements, the neurons, which are interconnected to each other in some fashion, forming a huge network. I propose to make these connections "directional" by putting

imaginary arrows on the connection lines in order to suggest unidirectionality of the propagation of impluses along the axons away from the cell body. The universe under consideration is, then, all possible networks that can be formed by connecting the elements, and a particular state of this universe is a particular network. I have now to estimate the number of states of this universe, in other words, the number of different networks that can be composed by directionally connecting n elements. The question as to what differentiates one network from another can be approached from two different points. One approach is a purely structural one, where the operational modalities of the nodal elements are ignored; the other approach takes these modalities into consideration. I suggest looking first into the purely structural features, and only later into the possible operations that are carried out by each neuron in a structurally defined network.

The problem of counting the number of nets which can be formed by directionally connecting n elements is solved easily with the aid of a connection matrix. This is a square matrix of n rows and n columns labeled according to the label of each element (Fig. 95). If element E_i is connected to element E_j, a "one" is inserted into the i-th row at the intersection of the j-th column. Otherwise, a "zero" is inserted. Thus, the particular way in which "ones" and "zeros" appear in the matrix uniquely determines the corresponding network. Hence $\mathcal{N}$, the number of ways in which "ones" and "zeros" can be distributed over the n^2 entries of the matrix, is also the number of different nets that can be constructed by directionally connecting n elements. With two choices at each entry, this number is

$$\mathcal{N} = 2^{n^2}$$

Since my ignorance is complete in regard to the question of whether one net is more probable than another, my universe is populated by equiprobable states, and, hence, I associate with it an uncertainty:

$$H = \log_2 2^{n^2} = n^2 \text{ bits/net.}$$

In other words, n^2 bits of information are necessary to specify a particular network, as could have been seen directly from the connection matrix where n^2 binary choices had to be made when putting "ones" or "zeros" into the n^2 entries in order to specify a particular network.

Estimates of the number of neurons in a human brain center around 10 billion. Consequently,

NET
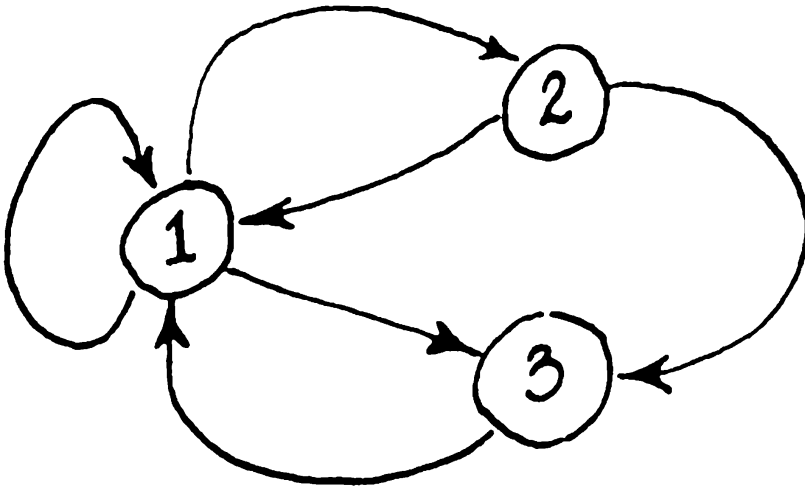
MATRIX

Figure 95. Representations of networks: (a) Graph; (b) Matrix. These representations are equivalent.

$$H = (10^{10})^2 = 10^{20} \text{ bits/brain.}$$

Let us see whether or not the information that is needed to specify just the connection structure of the nervous system—not to speak of the specifications of the operational modalities of its elements—can be genetically determined: Fortunately, there exist good estimates for the information content of the genetic program. The most careful one,

I believe, is still the one made by Dancoff and Quastler (3). From various considerations, they arrived at an upper and lower limit for the amount of uncertainty H_G in a single zygote:

$$10^{5.} < H_G < 10^{12}$$

In other words, the program that is supposed to define the structure is, by a factor of, say, 10^{10} off the required magnitude! This clearly indicates that the genetic code, which determines far more than just the nervous system, is incapable of programming nets of the unrestricted generality, as considered before. One way out of this dilemma is to assume that, de facto, only an extraordinary small amount in the structure of the nervous system is genetically specified, while the overwhelming portion is left to chance. Although the idea of leaving some space to chance-connections is not to be rejected entirely, it does not seem right to assume that only one hundredth of 1%, or less, of all neurons have specified connectivities. This assumption, for instance, would make neuroanatomy impossible, because the differences in brains would far outweigh their likenesses.

Another way out of this dilemma is to assume that the genetic code is indeed capable of programming a large variety of networks, each of which, however, involves only a small number of neurons, and each of which is repeated in parallel over and over again. Repetition of a particular structure in parallel requires very little information indeed, the only command being: "Repeat this operation until stop." The various kinds of networks may then be stacked in the form of a cascade (Fig. 96). In a very crude way, the appeal of this picture is that there is some resemblance to the various laminate structures that are observed in the distribution of neurons in the outer folds of the brain. Let us see now what numbers we obtain if we assume that the whole system of parallel networks in cascade is specified by the genetic program.

I propose to consider a small elementary net that involves only $2n$ neurons, half of which are located in one layer, say L_1, and the other half in an adjacent layer L_2 (Fig. 96). The axons of neurons in L_1 contact those in L_2, but there are no return pathways assumed in this simple model. Since the total number of neurons located in each layer is supposed to be large, say N, the complete connection scheme for the two layers is established by shifting the elementary network parallel to itself in both directions along the surface of the layers. The number of parallel networks is, thus,

$$P = \frac{N}{n} .$$

Again, a connection matrix for the elementary network can be
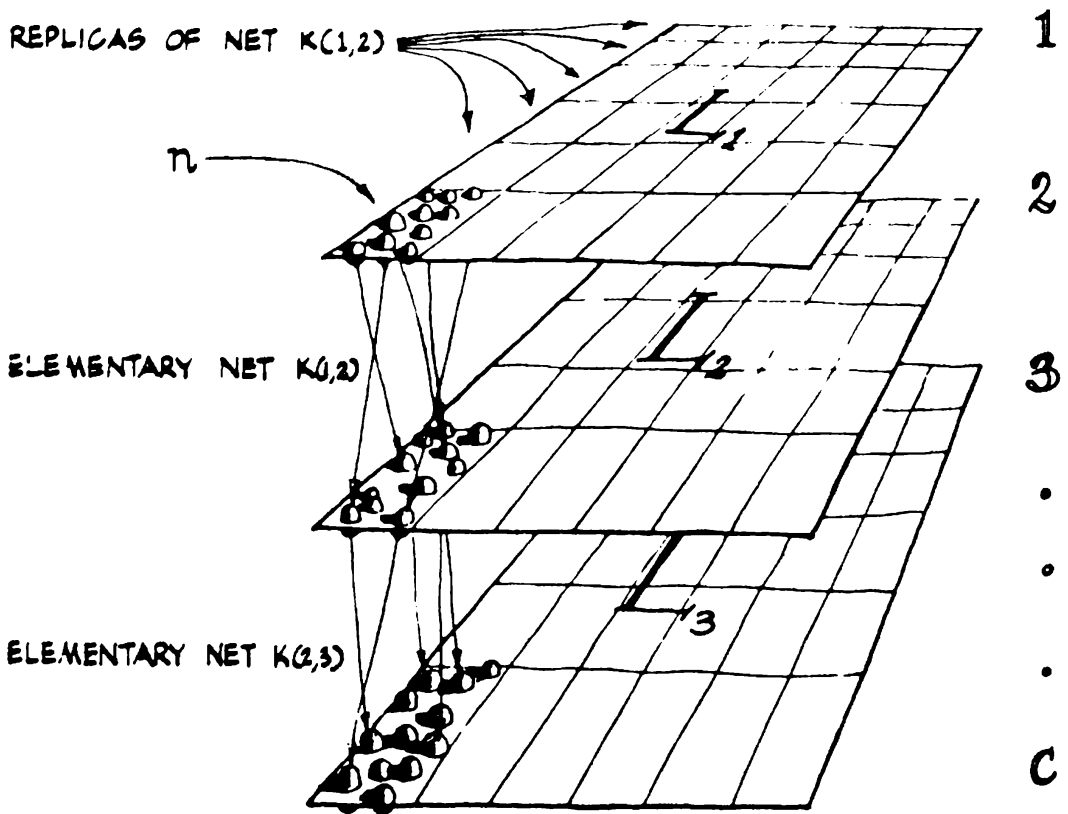
Figure 96. Cascade of C unlike networks composed of P like nets in parallel.

drawn with n rows and n columns, corresponding to neurons in layers L_1 and L_2, respectively, where, at the intersection of a row with a column, the absence or presence of a connection between the appropriate neuron in L_1 with a neuron in L_2 is indicated by a "zero" or a "one." Consequently, the number of nets is again 2^{n^2}, hence, the uncertainty of this elementary network is

$$H_n = n^2.$$

Although P such nets are working in parallel between layers L_1 and L_2, the uncertainty for the whole network connecting these two layers is still only n^2, because there is no freedom for even a single connection in any one of the P networks to change without the corresponding changes in all other nets, since their connectivity is determined by the connection matrix which functions as a genetic mold from which all P nets are cast.

I propose to assume that a different connection matrix controls the connections between the next pair of layers (L_2, L_3), and so on, in the cascade of C layers. The uncertainty of the system as a whole is, therefore,

$$H_S = n^2 C$$

and is assumed to be specified by the genetic code. Hence:

$$H_S = H_G$$

On the other hand, we have to accommodate in the whole system a totality of $\overline{N}$ neurons which are distributed among C layers of nP neurons each:

$$\overline{N} = nPC.$$

Eliminating n, the number of cells in an elementary network, from the two equations above, we obtain a relation between the number of cascades and the number of parallel channels in each cascade

$$H_G = \frac{\overline{N}^2}{CP^2} \, ,$$

or

$$P = \frac{\overline{N}}{H_G} \cdot \frac{1}{C} \cdot$$

TABLE 18

H_G bits								
10^6			10^8			10^{10}		
C	P	n	C	P	n	C	P	n
10^2	1.10^6	100	10^2	1.10^5	1000	10^2	1.10^4	10^4
10^3	3.10^5	30	10^3	3.10^4	300	10^3	3.10^3	3000
10^4	1.10^5	10	10^4	1.10^4	100	10^4	1.10^3	1000
10^5	3.10^4	3	10^5	3.10^3	30	10^5	3.10^2	300
10^6	1.10^4	1	10^6	1.10^3	10	10^6	1.10^2	100

Table 18 gives for three reasonable values of the genetic information H_G, a set of five values for triplets C, P, n, which satisfy the above equation.

Among the various choices that are given in Table 18, it seems to me that, for an assumed genetic information of 10^8 bits/zygote, a system that, on the average, consists of 1,000 layers (C = 10^3), each layer incorporating 30,000 parallel elementary networks (P = 3.10^4) which involve 300 neurons each, is, in the crudest sense, a structural sketch of cortical organization which may, perhaps, not be dismissed immediately for being completely out of the question, quantitatively.

Although this picture is extremely crude, the merit of it—if there is any merit at all—lies purely in its way of suggesting, among the vast amount of possible features, certain ones that may deserve closer inspection.

I would now like to point out a few implications as they seem to me of relevance to our topic. First, the possibility of parallel channels permits us to deal with relatively small nets for which an adequate theory may eventually be devised. I shall report briefly in a moment on the present state of affairs in the theory of small computing nets. Second, a network that consists of periodic repetitions of one and the same elementary network computes on its stimuli the same functions, irrespective of a linear translation of the stimulus distribution. Hence, parallelism implies translational invariance.

I have just referred to the elementary network as "computing nets," and I owe you an explanation for why this may be an appropriate term. At the same time, we shall see what these networks compute, what their computational possibilities are as constituents of large parallel nets, and, finally, how they may modify their operational modalities as a consequence of the results of earlier computations.

Since 1958, at the University of Illinois, we have been looking at the computational possibilities of periodic networks. We have been encouraged in our activity by the various findings of Lettvin (8), Maturana (9), Mountcastle (10), Hubel (5, 6), and others concerning the computation of abstracts in small nets arranged in parallel. The principle idea (1, 7) is to associate geometrical concepts with the connection scheme that prevails between two layers which are the loci of evenly distributed computational elements (Fig. 97).

Assume that from a small area, dA, located at $\vec{r}_1$ in layer L_1, fibers descend in all directions to synapse with elements located in layer L_2. Consider the bundle that synapses with elements located at r_2 in L_2. Along this bundle a certain fraction, λ, of the activity, $\sigma(\vec{r}_1)$ dA, that prevails in the vicinity of r_1, is passed along and elicits an infinitesimal response, $d\rho(\vec{r}_2)$, in the elements located in the vicinity of $\vec{r}_2$ in L_2. Let this response be proportional to the stimulus activity in $\vec{r}_1$:
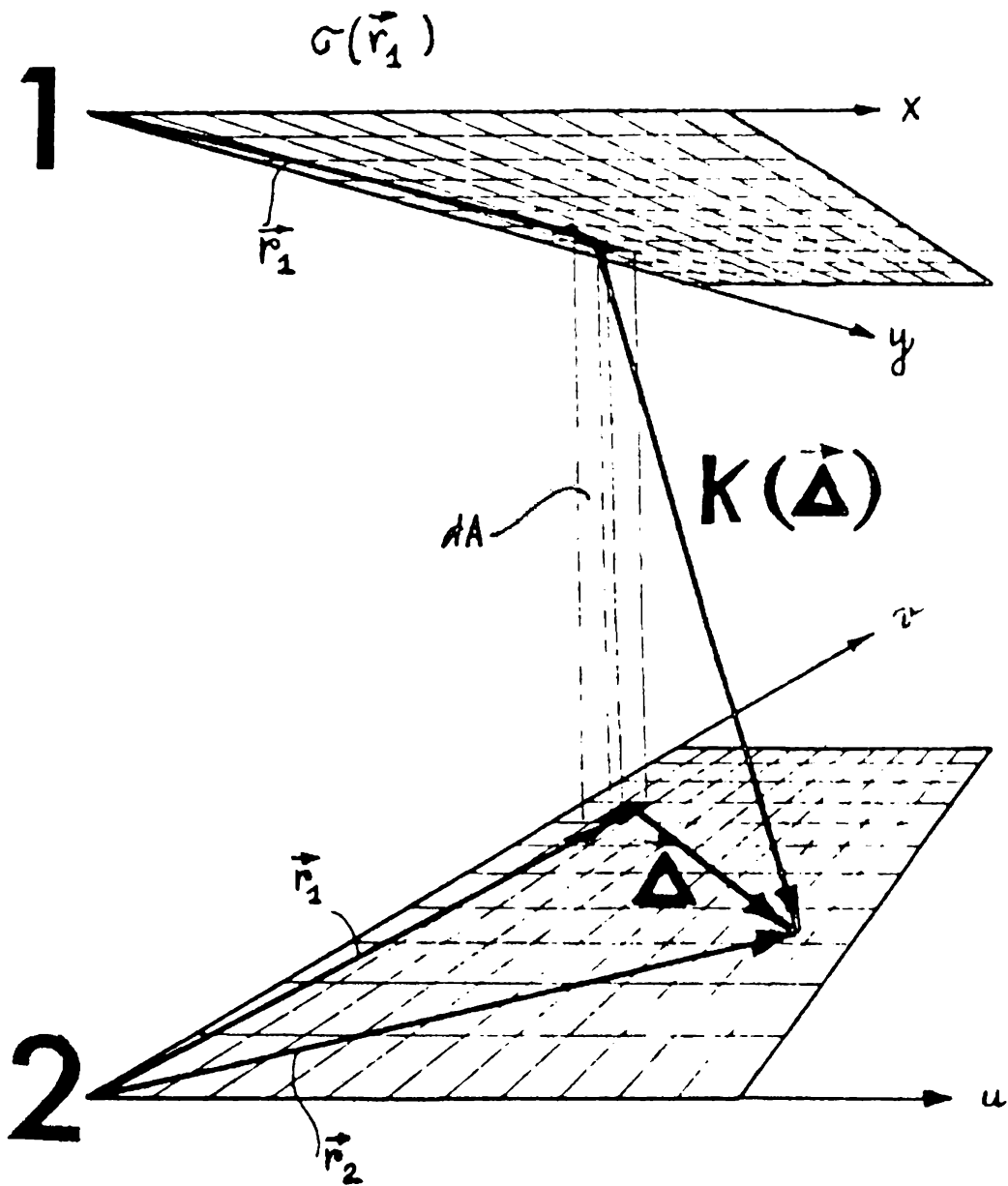
Figure 97. Geometrical relationships in an action network.

$$d\rho(\vec{r}_2) = \lambda K (\vec{r}_1, \vec{r}_2) \sigma (\vec{r}_1) dA,$$

with $K (\vec{r}_1, \vec{r}_2)$ the proportionality "constant," which may have different values from point to point in the stimulus layer L_1 as well as in the response layer L_2. This is suggested by letting K be a function

of the loci $\vec{r}_1$ and $\vec{r}_2$. Since K is uniquely associated with the bundle of fibers that connect elements at $\vec{r}_1$ with those at $\vec{r}_2$, we have in K the parameter that represents the action of the elementary network which connects the two layers L_1 and L_2. From here on, I shall refer to K as the "action function" of the elementary network. This network is supposed to repeat itself with periodicity, say p, in both directions, hence

$$K \ (x + ip, \ y + jp; \ u + ip, \ v + jp) \ = \ K \ (x, \ y; \ u, \ v)$$

$$i, j \ = \ 0; \ \pm 1; \ \pm 2; \ \pm 3; \ \ldots \ldots$$

and, consequently, K is a function of only the distance

$$\vec{\Delta} = \vec{r}_2 - \vec{r}_1$$

between the two points under consideration:

$$K \ (\vec{r}_1, \ \vec{r}_2) \ = \ (\vec{r}_2 - \vec{r}_1) \ = \ K \ (\vec{\Delta}) \ .$$

The response at $\vec{r}_2$ that is elicited by the contribution from $\vec{r}_1$ is, of course only a fraction of the response elicited at $\vec{r}_2$. In order to obtain the total response at $\bar{r}_2$ we have to add the elementary contributions from all regions in the stimulus layer:

$$\rho(\vec{r}_2) \ = \ \int_{L_1} K \ (\vec{\Delta}) \quad \sigma(\vec{r}_1) \ dA. \tag{13}$$

If the action function K ($\vec{\Delta}$) is specified—and my suggestion was that it is specified by the genetic program—then, for a given stimulus distribution, the response distribution is determined by the above expression. The physiological significance of the action function may best be illuminated by breaking this function up into a product of two parts:

$$K \ (\vec{\Delta}) \ = \ D \ (\vec{\Delta}) \quad . \quad T \ (\vec{\Delta}) \ , \tag{14}$$

where D ($\vec{\Delta}$) represents a change in the density of fibers that arise in $\vec{r}_1$ and converge in $\vec{r}_2$. Hence, D ($\vec{\Delta}$) is a structural parameter. T ($\vec{\Delta}$), on the other hand, describes the local transfer function for fibers that arise in $\vec{r}_1$ and synapse with neurons in the vicinity $\vec{r}_2$. Hence, T ($\vec{\Delta}$) is a functional parameter.

I hope that I have not unduly delayed an account of what is com-

puted by these nets. Unfortunately, I cannot give a detailed account of the various computational results that can be obtained by considering various action functions, K. Let me go only so far as to say that the results of these computations are invariants, or abstracts, of the stimulus distribution. I have already discussed invariance against translation as the prime bonus of using networks in parallel. Furthermore, it may not be too difficult to imagine the kind of abstracts that are computed if the action function, K, possesses certain symmetry properties. Consider, for instance, the three fundamental types of symmetry to hold for three types of action functions: the symmetric, the antisymmetric and the circular symmetric action function defined as follows:

$$K_s (\vec{\Delta}) = K_s (- \vec{\Delta}),$$

$$K_a (\vec{\Delta}) = -K_a (- \vec{\Delta}),$$

$$K_c (\vec{\Delta}) = K_c (|\Delta|).$$

Clearly, K_s gives invariance to reversals of stimuli symmetric to axes $y = 0$ and $x = 0$, that is, a figure 3 into ε, or M into W; while K_a gives invariance to reversals of stimuli symmetric to lines $y = x$, that is, $\sim$ into S, and $>$ into V. Finally, K_c gives invariance to rotations, that is, almost all previous reversals plus N into Z.

Maybe you have recognized in some of the properties of these action functions a resemblance with properties Hubel and others have observed in the response pattern of what they called the "receptor field." There is indeed a very close relation between these two concepts, because knowing the "domain" of the action function, that is, the cells in the target layer that are "seen" by a single cell in the source layer, enables one at once to establish the domain of the "receptor function." That is, the cells in the source layer that are seen by a single cell in the target layer. Let $G(\vec{\Delta})$ be the receptor function, then we have

$$G_s = K_s; \quad G_a = -K_a; \quad G_c = K_c.$$

At this stage of my presentation it may be argued that all this has to do with now well-established filter operations in the nervous system, but what has it to do with memory? It is true that in the foregoing I have indeed attempted to suggest a rigorous framework in which we can discuss these filter operations. But what if they arise from interaction with the environment? How would we interpret the computation

of an invariant if it results from past experience, or if it is the result of some learning process? I venture to say that we would interpret such adaptive computations of abstracts precisely as the functioning of a "memory" which responds with a brief categorization when interrogated with a flood of information. My remembering Karl Pribram cannot consist of interrogating a past record of all attitudes, gestures, etc., that I have stored of him. The probability that I find among these the ones with which I am presented at this moment is almost nil. Most probably, they are not even there. What I have built in, I believe, is a net of computers that compute on the vast input information which is pumped through my visual system all that is "Pribramish" — that is all categories that define him and only him — and come up with a terse name for all these categories which is "Karl Pribram."

In order to support this thesis, my task is now to show that it is indeed possible to teach the kind of computing networks which I have discussed to shift their computational habits from the computation of one type of abstracts to the computation of other abstracts. In the language of the theory of these computing networks, my task is now to show that the action function K, which uniquely determines the computed invariant, is not necessarily an unalterable entity, but can vary under the influence of various agents. In other words, the temporal variation of K does not necessarily vanish:

$$\delta K \neq O \ .$$

Before I go into the details of my demonstration, I have to confess that, to my knowledge, we are today still far from a satisfying theory of adaptive abstracting networks. The kind of mathematical problems which are soon encountered are of fundamental nature and there is as yet little help to be found in the literature. Hence, I will not be able to present spectacular results today. On the other hand, I hope that the following two simple examples will be sufficiently explicit as to indicate the main line of approach.

In my brief outline of some features of parallel networks, I suggested that under certain conditions the genetic program may suffice to specify all networks in the system. In my terminology, a network connecting two layers is specified if the action function, $K (\vec{\Delta})$, for the elementary network is defined. The whole network is then generated by simply shifting the elementary network along all points of the confining layers. In my following discussion of variable action functions, I still propose to think of a genetically programmed action function, say, $K_0 (\vec{\Delta})$; but now I consider this action function to be subjected to some perturbation. Due to my earlier suggestion [Equation (14)] that we think about this action function as being composed of two parts, a structural part $D (\vec{\Delta})$ and a functional part $T (\vec{\Delta})$,
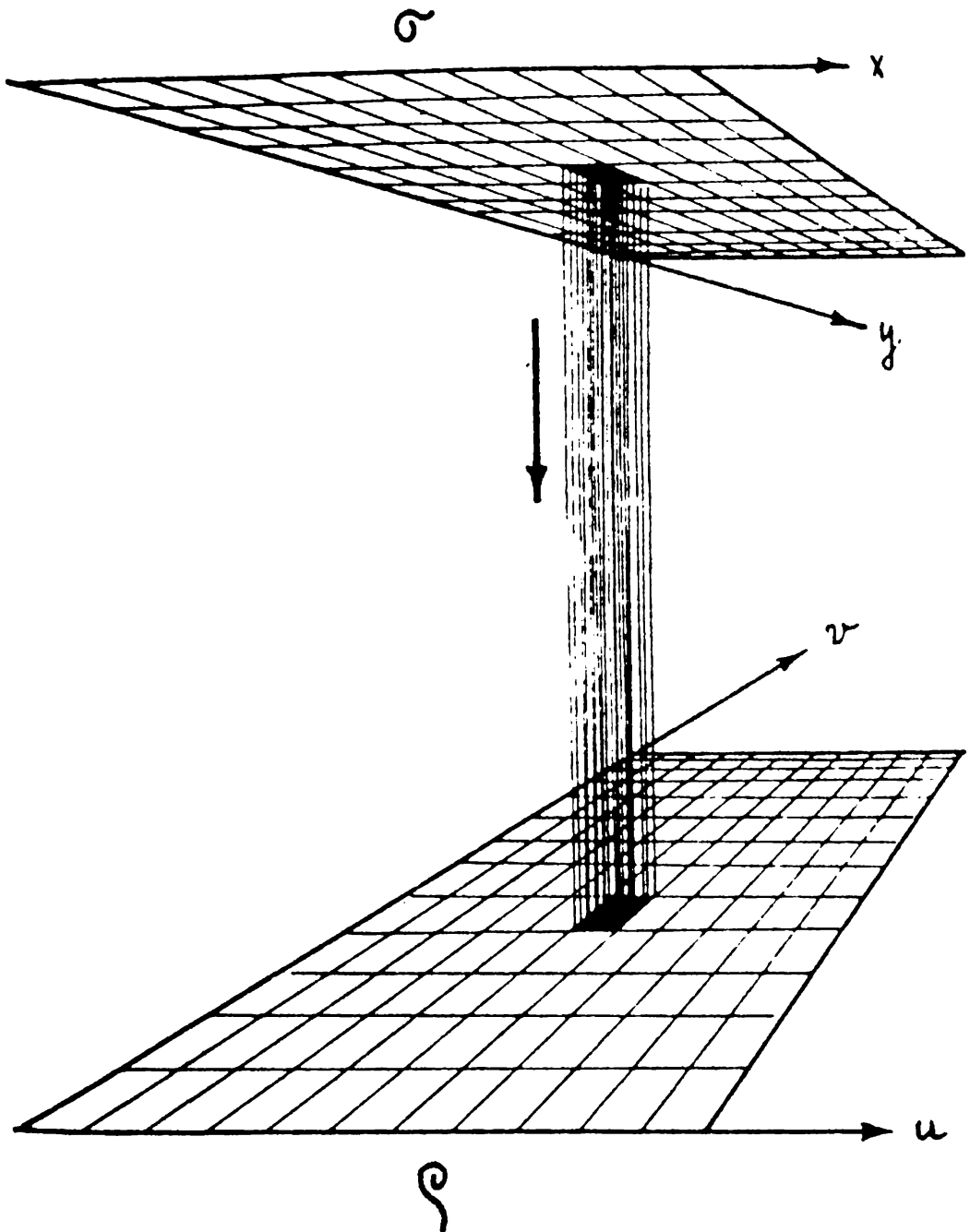
Figure 98. Geometrical relationships in an action function that is defined by an ideal one-to-one mapping.

we may now consider the perturbation to act on either T or D , or on both. Although I realize that the assumption of structural variation in neural nets is not too popular, I will give as my first example the sults of just such a structural perturbation which you may, perhaps, accept as a possibility which prevails during the early phases of the formation of these networks.

Assume that the simplest of all possible elementary networks is being programmed genetically (Fig. 96). It is a net consisting of precisely two elements, one in layer L_1, the other in layer L_2. Since $2n = 2$, we have $n = 1$, and the uncertainty for this net is

$$H = 1 \text{ bit/net.}$$

The genetic program says: "Connect these two elements!" In mathematical terms, the mapping function, $D(\vec{\Delta})$, that represents the structure of the net can be expressed by the Dirac delta function

$$D(\vec{\Delta}) = \delta(|\Delta|),$$

where

$$\delta(x - x_0) = \begin{cases} 0 \text{ for } x \neq x_0 \\ \infty \text{ for } x = x_0 \end{cases},$$

and

$$\int_{-\infty}^{+\infty} \delta(x - x_0)\, dx = 1.$$

For simplicity, let us assume the transfer function $T(\vec{\Delta})$ to be just a constant:

$$T(\vec{\Delta}) = a.$$

With these, the action function is simply

$$K(\vec{\Delta}) = a\,\delta(|\Delta|),$$

and the response in layer L_2 for a given stimulus in L_1 is after Equation (13):

$$\rho(\vec{r}_2) = a\lambda \int_{L_1} \delta(|\Delta|)\, \sigma(\vec{r}_1)\, dA = a\lambda\sigma(\vec{r}_2).$$

In other words, the response is an identical replica of the stimulus with modified amplitude due to the factor $a\lambda$, as was to be expected by this simple connection scheme.

Assume now that during the process of the realization of this network, it is impossible for the fibers descending from L_1 to make appropriate contacts in L_2, and suffer random deviations due to the presence of the intermediate glia cells. A fiber bundle leaving at $\vec{r}_1$, and destined to arrive at $\vec{r}_2$, will be scattered according to a Gaussian
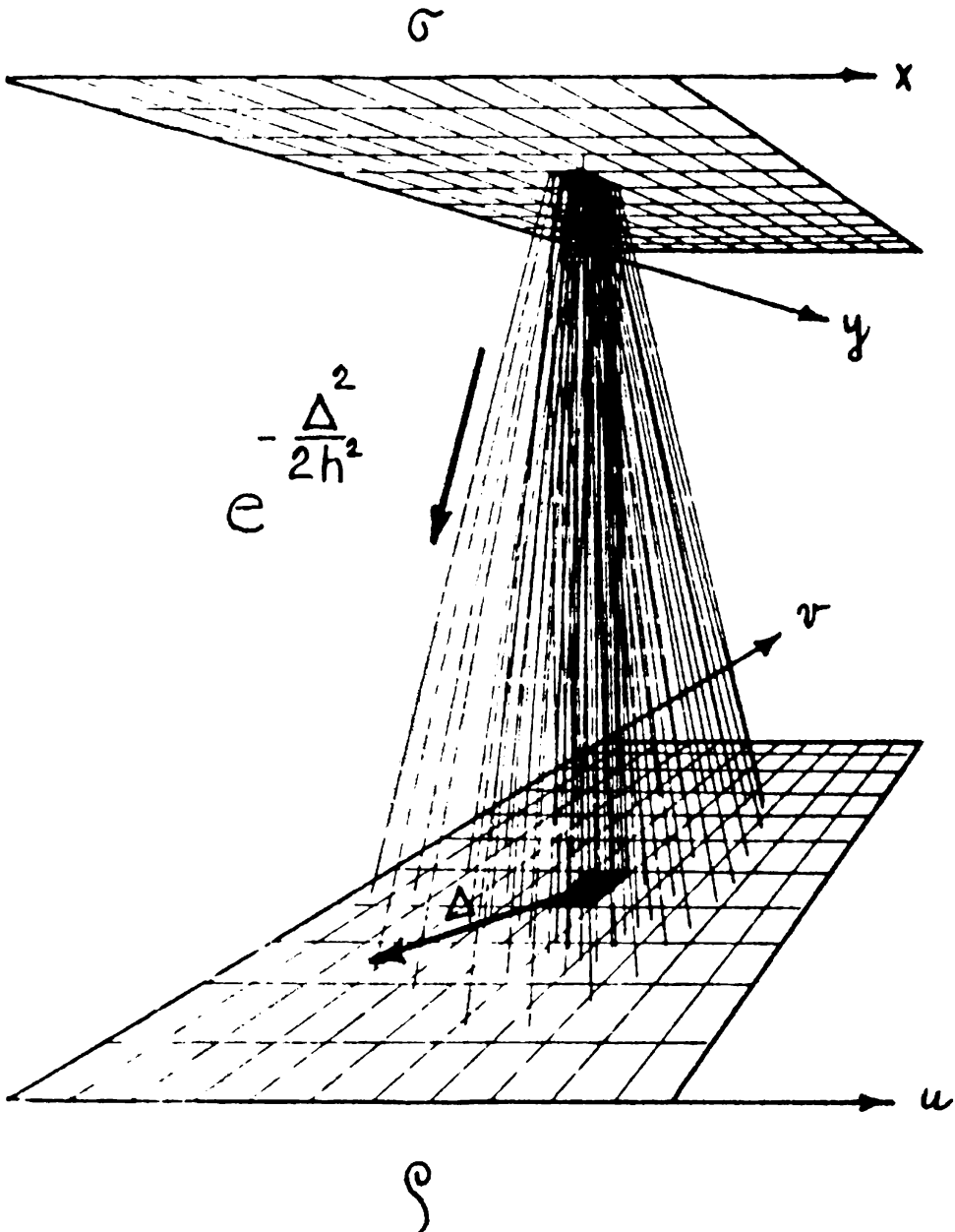
Figure 99. Geometrical relationships in an action function that arises from a random perturbation of an ideal one-to-one mapping.

distribution. Hence, the mapping function becomes:

$$D (\vec{\Delta}) \propto \exp [-\Delta^2/2h^2] ,$$

with h representing the variance of this distribution (Fig. 99).

Assume, furthermore, that the transfer functions for inhibitory and excitation connections are constants, but different for the two kinds,
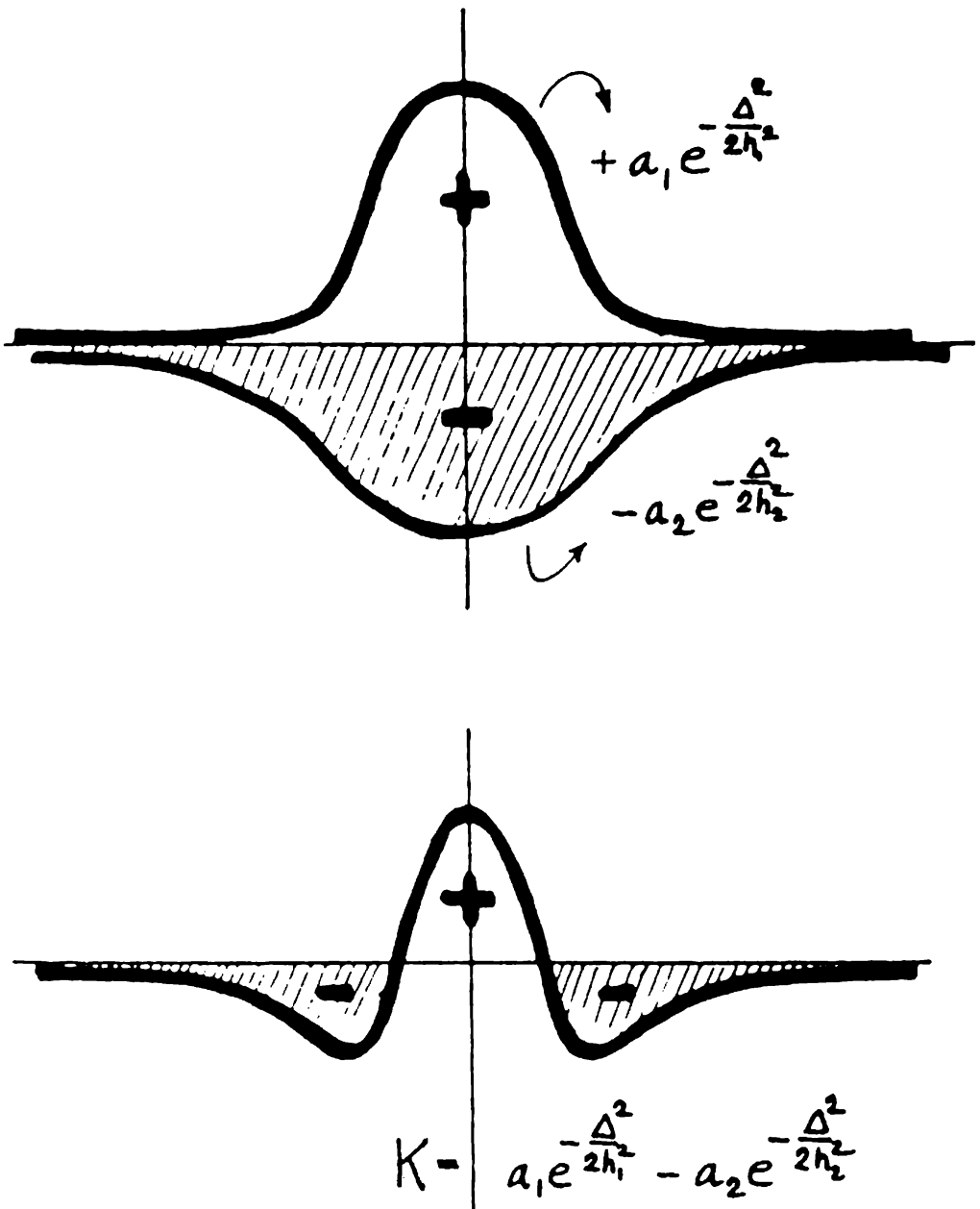
Figure 100. Graphical representation of a Gaussian (Normal) distributed action function. (a) Excitatory and inhibitory distribution separated. (b) Composite of excitatory and inhibitory distribution function.

$(^+a_1)$ and $(-a_2)$. Introducing corresponding variances, h_1 and h_2, the action function of the elementary network becomes:

$$K(\vec{\Delta}) = a_1 \exp\left[-\Delta^2/2h_1^2\right]$$

$$- a_2 \exp\left[-\Delta^2/2h_2^2\right],$$
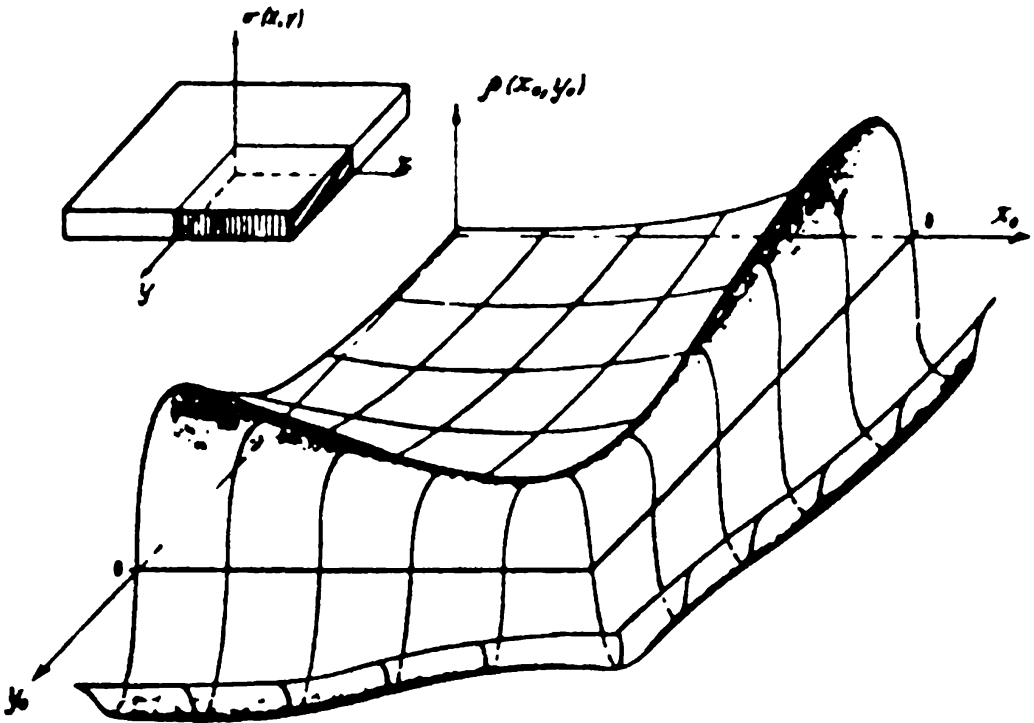
Figure 101. Response distribution elicited by a uniform stimulus confined to a square. Contour detection is the consequence of an action function obtained by superposition of an excitatory and inhibitory Gaussian distribution.

and the stimulus-response relation can be established with Equation (13). For uniform stimulus, $\sigma = \sigma_0$, the resulting response is also uniform:

$$\rho(\vec{r}_2) = \sigma_0 \lambda \quad (a_1 h_1^2 - a_2 h_2^2) = \text{const.}, \qquad (15)$$

and vanishes for

$$a_1 h_1^2 = a_2 h_2^2 . \qquad (16)$$

This condition is assumed in the graphical representation of the Gaussian action function in Fig. 100, and also in Fig. 101, which shows the response activity in layer L_2, if the stimulus pattern is a uniformly illuminated square. The interesting feature in the composite picture of Fig. 100 as a pronounced lateral inhibition of the fiber bundles acting on neurons located in L_2. The consequence of this action function is a computation of a contour by the elements in L_2. This contour is most conspicuously present if uniform stimulation elicits no response, or, in other words, if the condition expressed in Equation (16) is fulfilled.

Since there was nothing in our assumed program that could have specified such an exacting condtion, adaptation is the only mechanism that can be made responsible to accomplish this. In other words, the action function that has evolved so far has to be submitted to further changes, if the ideal connection conditions expressed in Equation (16) have not perchance established themselves spontaneously. This chance, however, is extremely small and, after formation, we may expect the network to be in a state where the equality sign in Equation (16) should be replaced by an inequality sign.

The question now arises as to what is going to change, how is the change to be accomplished, and what is the cause of this change? If I understood correctly some of the remarks that have been made during this meeting, then I am not permitted to change the structure of the network as it has evolved so far, because after the synaptic connections are established they are rigidly maintained. The only escape which is now left to me is to change $T(\overset{\rightarrow}{\triangle})$, the transfer function, or, as I referred to it earlier, the operational modality of the nodal element in this network.

ROBERTS: Why are you not permitted to change structure?

VON FOERSTER: I am just following the suggestions that were given earlier, I believe, which do not permit me to move synapses around. They are fixed.

SPERRY: But you can't do that because, earlier in the Conference, at least from the point of view of chemistry and microbiology, I didn't hear any loud objections when we were discussing the idea that there could be structural changes at synapses.

VON FOERSTER: Yes, but in my terminology, these would be functional changes of the transfer function, caused by some sub-microscopic changes in the synaptic structure. I don't know whether they can be seen in the microscope. Let me just add that I would be very happy indeed if I were allowed to make changes in the network structure. But what I am trying to say is that, even if no structural changes are permitted in my networks, I am still able to devise a computer that changes its internal organization, because I can now change the transfer function of my elementary components. I believe this is what Sir John was showing when he pointed out the variation in the efficacy of synaptic transmission.

Going back to my action function, I propose now not to change the structural part, $D(\overset{\cdot}{\triangle})$, which established itself as a random distribution of connecting fibers, and to permit the local transfer function, $T(\triangle)$, to be submitted to some perturbation. The transfer function, as I introduced it earlier, was of the most simple kind. It consisted of just two constants, $(+a_1)$ and $(-a_2)$, for excitation and inhibition, respectively. Assume that, after the contacts have been established, these two constants have some initial values, say, $(+a_{10})$

and $(-a_{20})$. However, the crucial condition for optimum contour extraction (Equation 16) is not initially fulfilled:

$$ a_{10} h_1^2 \neq a_{20} h_2^2 . $$

Since in this equation the variances of inhibitory and excitory connections are structual parameters (and according to my rules they are tabu), I have to let something operate on either a_{10} or a_{20} in order to obtain the desired equality. Among the many possibilities that offer themselves from a purely hypothetical point of view, let me submit one that has, for me, at least, the ring of plausibility. Let, for example, the left-hand side in the inequality exceed the right-hand side. I shall now assume that at least one of these transfer functions, in this case, a_2, the inhibitory transfer function, will increase its efficacy due to the overall activity of the responding network. In other words, I don't make this feedback loop a local affair, but rather an affair in which the activity of whole cell complexes, plus their surrounding tissue, may be involved. Formulated precisely, I have

$$ \frac{\partial a_2}{\partial t} = cR, $$

where R stands for the total response of the layer L_2:

$$ R = \int_{L_2} \rho(\vec{r}_2) \, dA . $$

In order to get a rough picture of what happens under these conditions, let the universe of our system be a dull one, with occasional appearances of objects that are "seen" by the elements in L_1, but otherwise is uniformly illuminated. In this case, R can be evaluated at once with the aid of Equation (15):

$$ R = (a_{10} h_1^2 - a_2 h_2^2) \cdot S(t) , $$

with $S(t)$ the total stimulation applied to layer L_1 as a function of time. Inserting this into Equation (17), one obtains a differential equation in a_2:

$$ \frac{\partial a_2}{\partial t} = c\lambda (a_{10} h_1^2 - a_2 h_2^2) \cdot S(t) , $$

with the initial condition

$$ t = 0 \ldots a_2 = a_{20} $$

and its solution:

$$a_2 h_2^2 = a_{10} h_1^2 - (a_{10} h_1^2 - a_{20} h_2^2) \exp(-h_2^2 c \; \lambda \int_0^t S(t) \, dt \,)$$

The first term represents the desired condition as expressed in Equation (16), while the second term, due to the decaying exponential eventually must vanish, whatever the history of the stimulus $\int_0^t S dt$ might have been.

Although this is admittedly an almost trivial example of an adaptive network, it enabled me, without too much acrobatics, to get my point across that a computer originally conceived as an ideal repeater, when exposed to the roughness of a real world, transforms itself to a perfect contour detector. This is the more surprising if one considers for a moment that this machine of considerable sophistication grew from a genetic specification that required, as you may remember, only one single bit of information!

With this example, I could retire now if another argument would not loom in the back of my mind: That, again, I could be accused of having merely presented the story of an adaptive filter mechanism, without having even touched the profound problem of memory.

Fortunately, in this dilemma a charming, but not immediately obvious, property of the Gaussian action function comes to my rescue. It turns out that if the condition in Equation (16) is fulfilled, in the vicinity of points $\vec{r}_2$ along the contour in the response layer L_2, this connection scheme produces a mean response density, $\vec{\rho}(\vec{r}_2)$, that is proportional to the curvature of the contour. Take $\underline{R}(\vec{r}_2)$ to be the radius of curvature of the contour at point $\vec{r}_2$, and k to be a constant, then

$$\vec{\rho}(\vec{r}_2) = k / \underline{R}(\vec{r}_2) \; .$$

In the first approximation, the total response, R, is the activity integrated over the whole contour

$$R = \int \vec{\rho}(\vec{r}_2) \, ds = k \int \frac{ds}{\underline{R}(\vec{r}_2)} \; ,$$

where ds is a line-element along the contour. But

$$ds = \underline{R}(\vec{r}_2) \, d\psi \quad ,$$

with ψ representing the polar angle from the center of curvature.

Hence,

$$R = k \int_{\psi_1}^{\psi_2} \frac{\underline{R}\ (\vec{r}_2)\ d\psi}{\underline{R}\ (\underline{r}_2)} = k \int_{\psi_1}^{\psi_2} d\psi = \text{size invariant.}$$

In other words, the only function that carries the information about the size of the object, namely $\underline{R}$, cancels out, and the resulting response, R , is invariant with respect to the size of different objects and varies only with their shape.

Imagine now that our system is in contact with a universe which has the peculiar property of being populated by a fixed number of objects of various sizes, but of similar shapes, that move about freely, so that the limited "visual field" of our system perceives at any instant only a certain fraction of the number of objects. Since each object seen will elicit the same response, R_0, Q objects in the visual field will elicit a total response of Q . R_0. In other words, our system is able to count! Since nothing of this sort was a property of the system when it was born — we are still dealing with only one bit of genetic information — it must have acquired its mathematical abilities as a result of interactions with its environment. Moreover, one may contrive a large variety of educational methods, that accelerate the process by which the system acquires its knowledge about numbers. When it finally masters this task, a tutor, ignorant about our system's internal workings may speculate about the location of its memory. We, of course, know that it is all over the place, realized in the structure of the connection scheme and in the operational modalities of all nodal elements of this network.

I shall now conclude my remarks by again giving a few numbers. First, let me take our tutor, who does not know the simple evolutionary history of our system, and who wants to study its anatomy, he will find an extraordinary complicated network consisting of n elementary nets, each different from any other. Applying good statistics, he can summarize his data by showing that the connections of all elementary nets have a Gaussian distribution, each network involving on the average, say, m contacted elements. The uncertainty H_e for a single elementary net is approximately*

$$H_e = \log_2 \sqrt{m} .$$

---

*Due to the constraint present in a normal probability distribution, the uncertainty is not $\log_2 m$, but approximately only $\frac{1}{2} \log_2 m$.

and, for the network consisting of N elements,

$$H = N \log_2 \sqrt{m}$$

If I use the figures from Table 18, as suggested earlier, the uncertainty for this system, consisting of only one net (C = 1), is:

$$H = 3 \cdot 10^4 \log_2 \sqrt{300} = 1.2 \cdot 10^5 \text{ bits.}$$

The anatomist may attribute this to the genetic program and will estimate its information content to be about 100,000 bits. We, however, know that the original specifications were written only in terms of one single bit. Where does all this information come from? The answer is, of course, from the "noise" that was introduced into the network when it made its feeble attempt to carry out the genetic command. In this case, as in so many others, it is the noise that enriches a structure that was extremely poor to begin with.

Let us turn now to the performance of our system. With N independent binary elements in its "sensory" layer, L_1, which may be ready for a new sensation in a time interval of t_0 seconds, the system's rate of input information is

$$H_{in} = N/t_0 \text{ bits/second.}$$

Its output rate depends, of course, upon the demands of our system's environment. Assume that the number of objects seen by our animal varies according to a normal distribution, with a variance of, say, M objects. It reports in intervals of t_0 seconds about the state of its visual universe. Hence, its output information rate:

$$H_{out} = \frac{1}{t_0} \log_2 \sqrt{M} \quad ,$$

and the ratio between output and input:

$$H_{out} / H_{in} = (\log_2 \sqrt{M}) / N.$$

However, M is always smaller than N , because one cannot see more objects than there are cones and rods. In the most optimistic

case, $M = N$, and the optimum output-input ratio is

$$\left(\frac{H_{out}}{H_{in}} \right)_{opt} = \frac{\log_2 N}{.\,2N}\;.$$

Again, for $N = 30,000$ this ratio becomes:

$$\left(\frac{H_{out}}{H_{in}} \right)_{opt} = 0.00075\;.$$

In other words, the system reduces considerably the uncertainty of its environment by its power of classification and abstraction which, in this case, consists of identifying individual objects.

Up to now, I have only stressed the miracles that are obtained when disorder is introduced into an ordered genetic program. One may ask now: Where is the internal organization that arises in my system from interactions with its environment, as I proclaimed in my earlier statements? It is clear that the consequences of such interactions can only be accounted for after the system is capable of interacting at all; but this is only after the network has established itself. The only quantity that changes after that is the inhibitory transfer function, a_2, which, for simplicity, I shall denote by "a" without index. Let a_{min} be the smallest interval detectable in observing this transfer function, then, n, the number of states of this "universe," is determined by the maximum value a_{max} that this function can assume:

$$n = a_{max} / a_{min}\;.$$

Consequently, H_{max}, the maximum uncertainty, is

$$H_{max} = \log_2 n.$$

With

$$n_o = a_o / a_{max} \ll n\;,$$

its initial uncertainty is

$$H_o = \log_2 (n - n_o),$$

that is, the log of the number of states still available, hence, its initial measure of organization is:

$$\Omega_0 = 1 - \frac{H_0}{H_m} = \frac{n_0/n}{\log_2 n} \, .$$

If, after exposure, the system drifts in its mature state over, say, n_1 distinguishable states, its final state of organization is:

$$\Omega_\infty = 1 - \frac{\log_2 n_1}{\log_2 n} = \frac{\log_2 (n/n_1)}{\log_2 n}$$

Comparison between initial and final state of organization shows that the measure of the final state of organization is greater than that of the initial state of organization, if

$$\log_2 (n/n_1) > (n_0/n) \, ,$$

Or if, approximately,

$$n > n_0 + n_1 \, .$$

But this is always the case. In fact, if the drift around the final equilibrium state is very small indeed, say $\bar{n}_1 = 1 + \epsilon$:

$$\Omega_\infty = 1 - \left[\epsilon/\log_2 n \right] = 1 \, ,$$

and the system is in almost perfect order.

I hope that these remarks suggest the possibility to recognize learning and memorizing systems as computers, changing their operational modalities as a consequence of interactions with their environment. Their operations change so as to remove more and more uncertainty of the environment, until the output of these systems keeps them in equilibrium with their universe.

REFERENCES
(Starred references have not been verified.)

*1. Babcock, M. L., A. Inselberg, L. Lofgren, H. Von Foerster, P. Weston, and G. W. Zopf, Jr. 1960. Some principles of pre-organization in self-organizing systems. Tech. Rep. #2 Contract

Nonr 1834(21). Electrical Engineering Research Lab., Engineering Experiment Station, Univ. of Illinois, Urbana.

2. Brillouin, L. 1962. Science and Information Theory. Academic Press, New York.

3. Dancoff, S. M. and H. Quastler. 1953. The information content and error rate of living things. In H. Quastler (Ed.), Information Theory in Biology. University of Illinois Press, Urbana. Pp. 263-273.

4. Defares, J. G. and I. N. Sneddon. 1961. The mathematics of medicine and biology. Year Book Medical Publ. Inc., Chicago.

5. Hubel, D. H. and T. N. Wiesel. 1959. Receptive fields of single neurons in the cat's striate cortex. J. Physiol. 148: 547-595.

6. Hubel, D. H. and T. N. Wiesel, 1962. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. J. Physiol. 160: 106-154.

*7. Inselberg, A. and H. Von Foerster. 1962. Property extraction in linear networks. Tech. Rep. #2 N. S. F. Grant 17414; Electrical Engineering Research Laboratory, Engineering Experiment Station, Univ. of Illinois, Urbana.

*8. Lettvin, J. Y., H. R. Maturana, W. S. McCulloch and W. Pitts. 1959. What the frog's eye tells the frog's brain. Proc. I. R. E. 47: 1940-1951.

9. Maturana, H. R. and S. Frenk. 1963. Directional movement and horizontal edge detectors in the pigeon retina. Science. 142: 977-978.

*10. Mountcastle, V. B. 1963. Abstract in Symposium on information processing in the nervous system. 22nd international congress of physiological science. Excerpta Medica Foundation, Amsterdam, 1, pt. 2, p. 930.

11. Quastler, H. 1958. A primer on information theory. In H. P. Yockey, R. L. Platzman and H. Quastler (Eds.), Symposium on Information Theory in Biology. Pergamon Press, New York. Pp. 3-49.

*12. Shannon, C. E. 1951. Prediction and entropy in printed English. The Bell Syst. Tech. J. 30: 50-64.

13. Shannon, C. E. and W. Weaver. 1949. The Mathematical Theory of Communication. University of Illinois Press, Urbana.

*14. Von Foerster, H. 1960. On self-organizing systems and their environments. In M. C. Yovits and S. Cameron (Eds.), Self Organizing Systems. Pergamon Press, New York. Pp. 31-50.

TIME AND MEMORY [*]

HEINZ VON FOERSTER (*University of Illinois, Urbana, Ill.*): I am sorry that Dr. Whitrow is absent this morning, because I hoped he would elaborate upon some of his delightful remarks of yesterday evening on the relationship between time and memory. Since I believe that this relationship may be of interest to some participants in this conference, I wish to add just a few words.

Dr. Whitrow said yesterday that we know little about "memory." I whole-heartedly agree but I would like to add that we know even less about "time." The cause for this deficiency I see in the superior survival value for all perceptive and cognitive living organisms if they succeed to eliminate quickly all temporal aspects in a sequence of events or, in other words, if "time" is abandoned as early as possible in the chain of cognitive processes. I believe that I can give at least two plausible arguments to support this proposition.

The first argument is purely numerical and attempts to show the infinitely superior economy of a "time-less" memory compared to a record which is isomorphic to the temporal flow of events. Consider a finite "universe" which may assume at any particular instant precisely one of n possible states, S_1; S_2; S_3; . . . S_n. Let m be the length of a sequence of states:

$$1 \quad 2 \quad 3 \quad 4 \quad . \: . \: . \: . \: . \quad m$$

e.g.: $\quad S_{15} \: S_{23} \: S_4 \: S_{105} \: . \: . \: . \: . \: S_{22}.$

The number, N, of distinguishable sequences of length m is equivalent to the number of combinations of n distinct objects taken m at a time, multiplied with the number of permutation of m distinct objects. This is because the sequence $S_a \: S_b$ is, of course, different from the sequence $S_b \: S_a$. Hence

$$N = \binom{n}{m} \cdot m! = \frac{n!}{(n-m)!}.$$

If both n and m are large numbers, but n is much larger than m, one may approximate

$$N \simeq n^m.$$

The number of binary relays to hold this number—or the "amount of information" to be stored—is approximately

---

$$H = \log_2 N = m \log_2 n = m\, H_o \text{ bits,}$$

where H_o is the amount of information necessary to specify one state of the universe. As an example, let us consider the retinal mosaic of various excited states of its rods and cones as the states of our "visual universe." Conservative estimates suggest 32 distinguishable states for each receptor of which there are about two hundred millions in both eyes. Hence, the visual universe has

$$n = (32)^{2 \cdot 10^8}$$

distinguishable states and

$$H_o = \log_2 n = 10^9 \text{ bits.}$$

If only ten states per second are processed by the retina—a regular movie projector presents 24 pictures per second—then during one second a sequence of length $m = 10$ is to be stored, i.e.,

$$H \text{ (1 second)} = 10^{10} \text{ bits.}$$

However, the entire brain has "only" 10^{10} neurons at its disposal. Let us be optimistic and assume one thousand bits stored within each neuron, then in one thousand seconds—or slightly over a quarter of an hour—the whole brain is flooded with information, most of which may be completely worthless.

On the other hand, if it were possible to integrate the sequence of m events into a single "operator" that permits reconstruction of the sequence, say, "a galloping elephant," "a flight over the Atlantic," etc., the number of distinguishable "macro-states" becomes

$$N^* = \frac{n}{m}$$

and

$$H^* = H_o - \log_2 m.$$

The compression ratio between these two methods of "storage" is simply

$$\frac{H^*}{H} = \frac{1}{m}\left(1 - \frac{\log_2 m}{\log_2 n}\right) \simeq \frac{1}{m}.$$

If the length of the sequence is extended, this reduction in uncertainty may take on values of considerable magnitude.

I hasten to demystify these "operators" which I have just mentioned. Indeed, they are perpetually computed in our perceptive system and their abstracting powers become apparent in their linguistic representation, usually in form of *names* for spatial abstracts and of *verbs* for temporal abstracts. In fact, without these abstracting operators we could not conceive of motion or of change, and—as an ultimate abstraction—of the flow of time. Contemplate for a moment the somewhat formalized representation of a unicellular animal moving from one spot to another by extending a tubular pseudopod
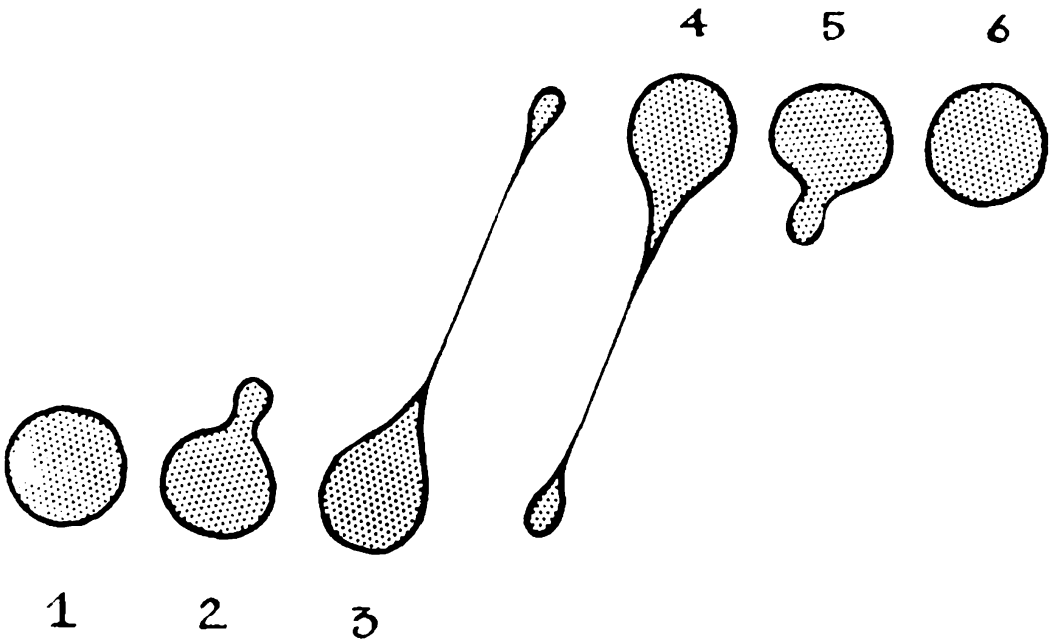
FIGURE 1. Schematic representation of a unicellular animal moving from one spot to another by extending a tubular pseudopod and pulling itself up through this extended capillary.

and pulling itself up through this extended capillary (See FIGURE 1, Stages 1 → 6). However, what we see in fact is a sequence of six apparitions of quite distinct shapes of which it is difficult—I believe even impossible—to assert that they represent the "same object" unless, of course, the set of transformations is specified under which the properties of this "object" remain invariant. These transformations may accommodate spatial as well as temporal aspects of the object under consideration, as can be seen by the linguistic representation of the totality of events depicted in FIGURE 1: "a unicellular animal moving from one spot to another by extending a tubular pseudopod and pulling itself up through this extended capillary." The invariance of the sequence 1, 2, 3, 4, 5, 6 is suggested by a feeling of "inappropriateness" if any permutation of this sequence, say, 214356 is proposed as an alternate possibility.

It is clear that it is due to memory that temporal abstracts can be computed and stored. Although memories may have some charming aspects, their crucial test lies in their efficacy to anticipate sequences of events, in other words, to permit inductive inferences. The conceptual construct of "time" is, so far as I see it, just a by-product of our memory, which in some instances may use "time" as a convenient parameter—a *tertium comparatum*, so to say—to indicate synchronism of events belonging to two or more spatially separated sequences. Of course, there is *no need* to refer to time in such

comparison, for it is always sufficient to take one sequence as "standard" and to associate with standard events the events of another sequence as, for instance, the anticipation of the sequence of events regarding Peter:[1] ". . . Verily I say unto thee, that this day, even in this night, before the cock crows twice, thou shalt deny me thrice."

In parenthesis it may be interesting to note that this prediction would immediately lose its punch if it would use temporal reference, for instance: ". . . even in this night, before 6:30 a.m., thou shalt deny me thrice." The intellectual slump one suffers in this version stems, I believe, from the fact that idealized absolute, or Newtonian, time carries no information: $H \rightarrow 0$. Unfortunately, in spite of considerable efforts by clock designers to build a perfect clock, this ideal goal has not yet been attained. There is still a small residue of uncertainty $H = \epsilon > 0$, which is due to small inaccuracies of even the best clocks and, ultimately, there will be quantum noise which will set an absolute limit to this enterprise.

Since these assertions which represent my second argument may, at first glance, sound surprising, let me first demonstrate that Newtonian time is a useful but unnecessary parameter in a complete description of the universe; second, let me briefly state what I mean by an accurate clock.

In practice an approximation to Newtonian absolute time, called "ephemeris time," is obtained in two steps as Dr. McVittie pointed out earlier this morning. First the equations of motion in Newtonian mechanics are solved for various celestial bodies, in particular for the different planets P_i. These solutions usually express the positions $\vec{r_i}$ of these bodies in terms of a linear parameter, t, called time:

$$\vec{r_i} = \vec{r_i}(t).$$

Second, the scale of this parameter is adjusted so as to give the best fit between observation and the theoretical solutions. In fact, this established scale fits the observations so well that there is a residual error of only one unit in about 10^{11} units.[2] With this estimate one may calculate the uncertainty H of reading this astronomical clock. With the probability p of making an error

$$p = 10^{-11},$$

and with the definition of H for a binary choice:

$$H = - p \log_2 p - (1 - p) \log_2 (1 - p),$$

one finds the uncertainty to be

$$H = 3.8 \cdot 10^{-10} \text{ bits/unit}.$$

It may be interesting to determine whether this small residue of uncertainty is due to "noise" in the observed system, that is the planets refuse to obey Newton by occasionally performing small extravagancies in their other-

wise predictable behavior, or whether this noise is introduced by the transmission channel, that is by inaccuracies of observation due to random fluctuations in the atmosphere, in the optical equipment or in the evaluation of data.

If the latter should be the case, then this uncertainty can be made arbitrarily small on the basis of Shannon's[3] far reaching theorem which states that if the *same* sequence of signals is repeated again and again the effect of errors (within certain limits) that are introduced during transmission can be made arbitrarily small. This principle of "Reliability through Redundancy" is most effectively applied in all periodic or repetitive phenomena. All "clocks," for instance, are based on periodic event sequences, and environmental periodicities can be recognized, or absorbed, by the genetic code of reproducing and mutating living organism by programming these periodicities into the organism in the form of circadian rhythms: "Even if you read me poorly, if you read me often enough, you will get my message."

If I should make a guess as to the causes of the small residue of uncertainty left in establishing ephemeris time, I would venture to say that they come indeed from a certain capriciousness of the planets, for the channel noise has most probably been eliminated by the prolonged observation of these periodic events.

After having assured ourselves that this parameter "time" is universal* and does not change scale from planet to planet, we can now eliminate it from the set of equations which represent the positions of the planets as a function of this parameter by selecting the positions of one planet $\vec{r_o}$ as reference for the positions of all others:

$$\vec{r_i} = R_i(\vec{r_o}).$$

In other words, one would, for instance, tell the position of Mars in reference to the position of Venus, etc. Adhering to this scheme, appointments would be made in this form: ". . . we shall meet after the sun has risen twice." Of course, this is precisely the method outlined earlier in which simultaneity of events belonging to different sequences are used to establish the "when" of an event of one sequence by reference to a particular event of a standard sequence.

For practical purposes, however, it is convenient to have a highly redundant signal generator—a reliable clock—which facilitates the estimates of the simultaneity of events in a large number of sequences. Clearly, such a device should not inject into the universe of observation unwanted uncertainties, i.e., each subsequent state should be well determined by its predecessor. This is most easily accomplished if this device goes at a constant rate, which gives the additional advantage that such a device may read ephemeris time which is a useful parameter in Newton's equations of motion.

This brings me to my second point, namely, what do I mean when I speak of a "reliable," of an "accurate," clock that goes at a "uniform rate." As Dr McVittie has already pointed out, comparison of one clock with another may never establish which one goes ahead and which one falls back. Since cross-correlation between two clocks leads us nowhere, I propose to consider for a moment auto-correlation of one clock with itself.

FIGURE 2 shows such an arrangement where an illuminated diagonal slit in a circular rotating disk, representing the clock, is located at the center of curvature R of a convex mirror. In this arrangement, the clock is optically mapped onto itself.†

If the disc rotates with angular velocity ω, the angular position ϕ_1 of the slit and ϕ_2 of its image are given in parametric form (time t is parameter):

$$\phi_1 = \omega t$$

$$\phi_2 = \omega(t - 2R/c)$$

where c is the velocity of light. Eliminating t from this pair of equations expresses the position of the slit's image as a function of the position of the slit:

$$\phi_2 = \phi_1 - 2R\omega/c.$$
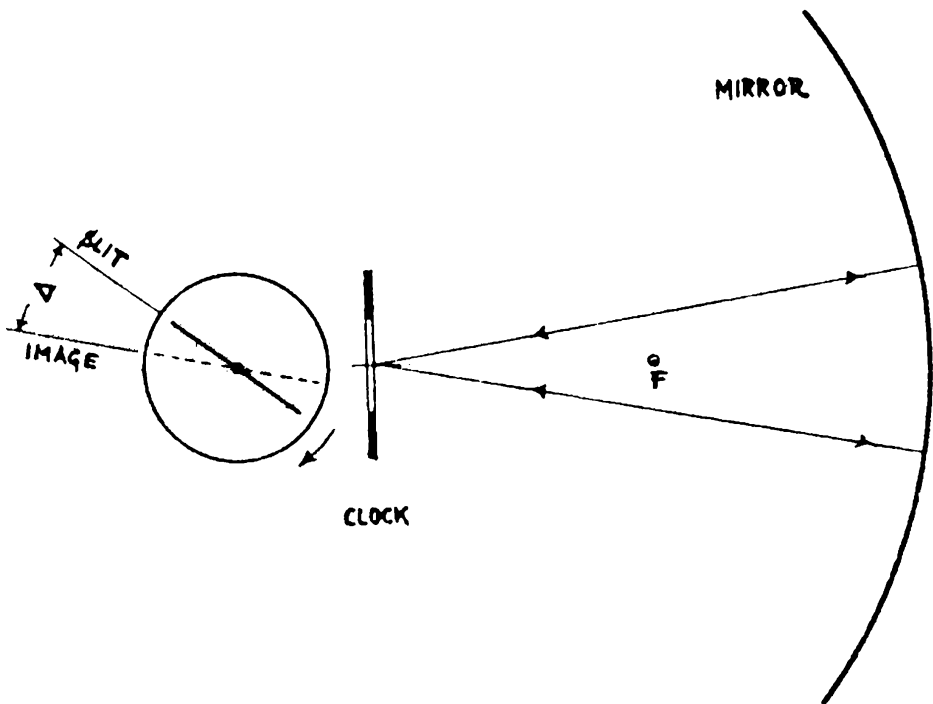
FIGURE 2. Determination of the accuracy of a clock through auto-correlation.

*That is within the framework of Newton's equations of motion.
†The inversion is compensated by making the hand a diagonal slit rather than a radial one.

This means that the slit will be trailed by its image at an angle of

$$\phi_2 - \phi_1 \equiv \Delta = - 2R\omega/c.$$

If the curvature of the mirror, its distance from the slit and the velocity of light are for the moment assumed to be constant, then it appears that the angular displacement Δ between slit and image will reflect all variations in the rate by which this clock proceeds. If ω increases, so will Δ and conversely, a reduced ω will result in a smaller displacement. One is tempted to say that a constant Δ indeed ascertains an accurate clock that proceeds at a constant rate.

Nevertheless, this naive interpretation has been subjected to a severe criticism by E. A. Milne,[††] who rightly points out that a constant displacement Δ means only that its variation vanishes. With

$$\Delta = - 2R\omega/c$$

this means that

$$\delta\Delta = 0$$

or

$$\frac{\delta\Delta}{\delta R}\, dR + \frac{\delta\Delta}{\delta\omega}\, d\omega + \frac{\delta\Delta}{\delta c}\, dc = 0.$$

Calculating the partial derivatives suggested above, this expression may be rewritten to read

$$\frac{d\omega}{\omega} + \frac{dR}{R} = \frac{dc}{c}.$$

If in spite of constancy of the displacement a dependency of ω with the position ϕ_1 of the slot is suspected, one may write:

$$\frac{1}{\omega}\frac{d\omega}{d\phi_1} + \frac{1}{R}\frac{dR}{d\phi_1} = \frac{1}{c}\frac{dc}{d\phi_1}.$$

This expression suggests that indeed a variation of rate at which the clock proceeds

$$\frac{d\omega}{d\phi_1} \neq 0,$$

may go unnoticed by relying on an invariable displacement Δ, if the mirror flaps or wiggles, or if the velocity of light jerks back and forth precisely in such a manner as to compensate for the variation of ω. One may imagine a "law of nature" that couples the three quantities R, c and ϕ_1 so that the relation

––––––––––

[††] Milne's criticism is addressed against a "Gedanken experiment" which incorporates entirely different physical devices. However, the basic features in both experiments are equivalent.

obtains

$$\frac{1}{c}\frac{dc}{d\phi_i} \qquad \frac{1}{R}\frac{dR}{d\phi_i} \qquad \Omega(\phi_i)$$

where the function on the right hand side represents the expression

$$\Omega(\phi_i) \qquad \frac{1}{\omega(\phi_i)}\frac{d\omega(\phi_i)}{d\phi_i} \; .$$

There are, of course, an infinite number of functions $c(\phi_i)$ or $R(\phi_i)$ or couplings between R and c, which will satisfy the above differential equation.

However, the amusing upshot of this side issue is that whatever these laws may be, they must transmit with great accuracy the fluctuations in the clock $\omega(\phi_i)$ to the mirror and force it to wiggle or to flap, or to the velocity of light to jiggle in precise compensation of the fluctuation of the clock so as to make the deviation Δ to appear unchanged or, at least, to change so little that ephemeris time can still be determined with the great precision mentioned earlier. Hence, these processes—imaginary or real—are of such redundancy that they do not interfere with our use of the parameter "time" as a *tertium comparatum* which facilitates highly accurate determinations of the simultaneity of events belonging to different cognitive sequences.

References

1. ST. MARK: 14-30.
2. G. M. MCVITTIE. 1961. Fact and Theory in Cosmology : 25. Eyre & Spottiswoode London. (1961).
3. C. E. SHANNON & W. WEAVER. 1949. The Mathematical Theory of Communication : 39. University of Illinois Press. Urbana, Ill.
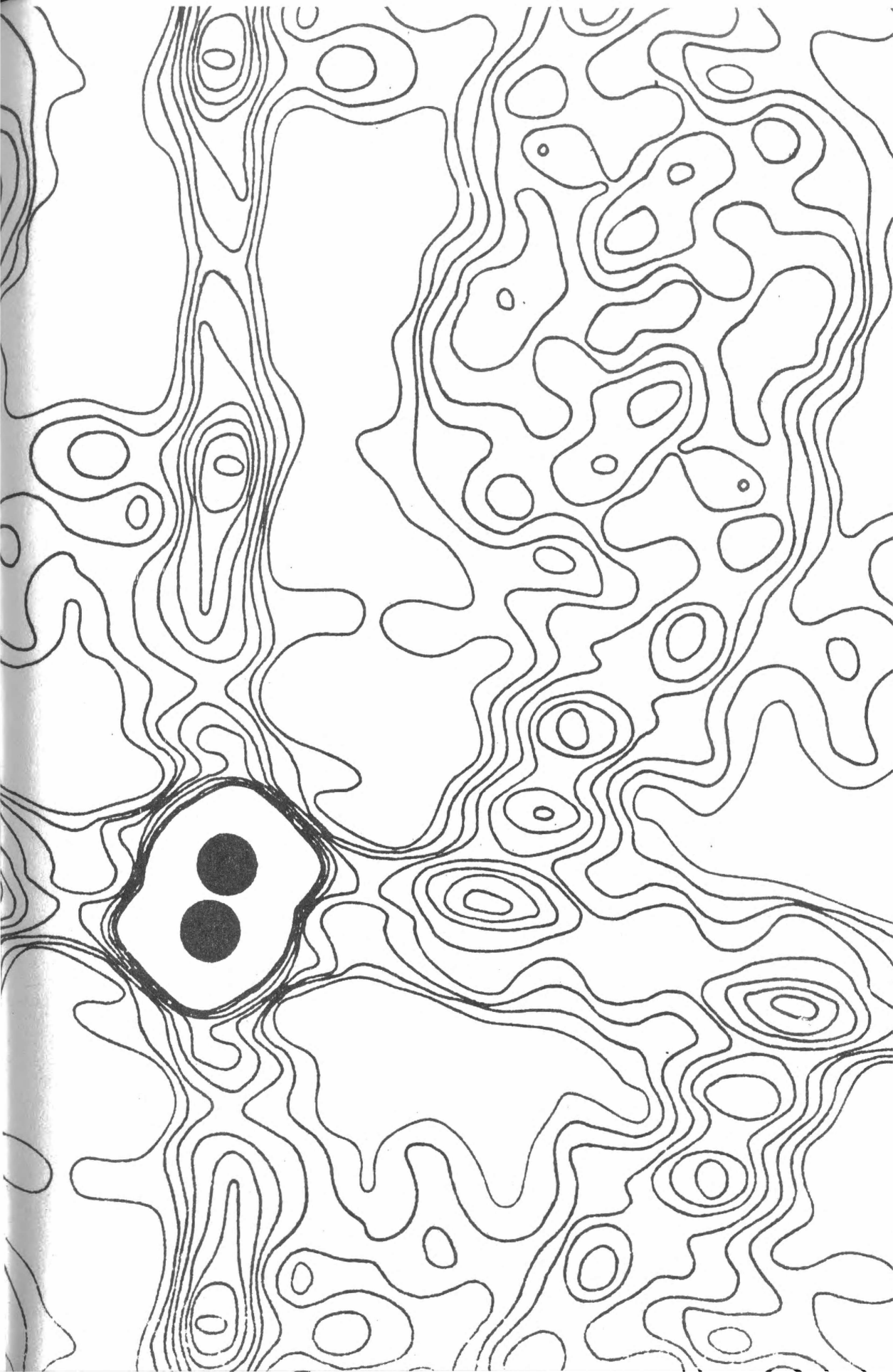
MOLECULAR ETHOLOGY

An Immodest Proposal for Semantic Clarification

Heinz Von Foerster

Departments of Biophysics and Electrical Engineering
University of Illinois
Urbana, Illinois

I. INTRODUCTION

Molecular genetics is one example of a successful bridge that links a phenomenology of macroscopic things experienced directly (a taxonomy of species; intraspecies variations; etc.) with the structure and function of a few microscopic elementary units (in this case a specific set of organic macromolecules) whose properties are derived from other, independent observations. An important step in building this bridge is the recognition that these elementary units are not necessarily the sole constituents of the macroscopic properties observable in things, but are determiners for the synthesis of units that constitute the macroscopic entities. Equally helpful is the metaphor which considers these units as a "program," and the synthesized constituents in their macroscopic manifestation as the result of a "computation," controlled and initiated by the appropriate program. The genes for determining blue eyes are not blue eyes, but in blue eyes one will find replicas of genes that determine the development of blue eyes.

Stimulated by the success of molecular genetics, one is tempted to search again for a bridge that links another set of macroscopic phenomena, namely the behavior of living things, with the structure and function of a few microscopic elementary units, most likely the same ones that are responsible for shape and organization of the living organism. However, "molecular ethology" has so far not yet been blessed by success, and it may be worthwhile to investigate the causes.

One of these appears to be man's superior cognitive powers in dis-

criminating and identifying forms and shapes as compared to those powers which allow him to discriminate and identify change and movement. Indeed, there is a distinction between these two cognitive processes, and this distinction is reflected by a difference in semantic structure of the linguistic elements which represent the two kinds of apparitions, namely different nouns for things distinct in form and shape, and verbs for change and motion.

The structural distinction between nouns (cl_i^k) and verbs (v_i) becomes apparent when lexical definitions of these are established. Essentially, a noun signifies a class (cl^1) of objects. When defined, it is shown to be a member of a more inclusive class (cl^2), denoted also by a noun which, in turn, when defined is shown to be a member of a more inclusive class (cl^3), etc., [pheasant $\rightarrow$ bird $\rightarrow$ animal $\rightarrow$ organism $\rightarrow$ thing]. We have the following scheme for representing the definition paradigm for nouns:

$$cl^n = \{cl_{i_{n-1}}^{n-1}\{cl_{i_{n-2}}^{n-2}\{\cdots\{cl_{i_m}^m\}\}\}\} \tag{1}$$

where the notation $\{\epsilon_i\}$ stands for a class of elements ϵ_i $(i = 1, 2, \ldots, p)$, and subscripted subscripts are used to associate these subscripts with the appropriate superscripts. The highest order n in this hierarchy of classes is always represented by a single undefined term "thing," "entity," "act," etc., which appeals to basic notions of being able to perceive at all. A graphic representation of the hierarchical order of nouns is given in Fig. 1 and a more detailed discussion of the properties of these (inverted) "noun-chain-trees" can be found elsewhere (Weston, 1964; Von Foerster, 1967a).

Essentially, a verb (v_i) signifies an action, and when defined is given by a set of synonyms $\{v_j\}$, by the union or by the intersection of the meaning of verbs denoting similar actions. [hit $\rightarrow$ {strike, blow, knock} $\rightarrow$
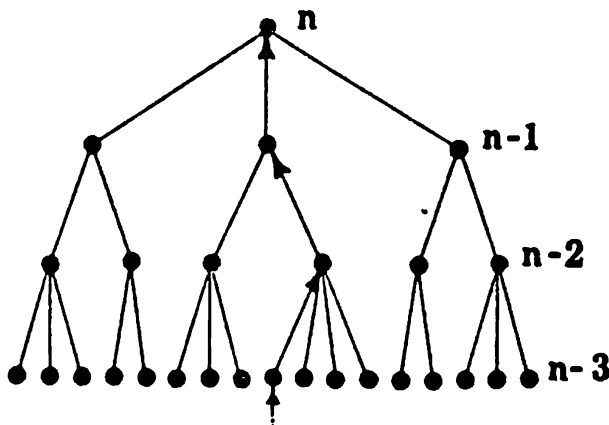
Fig. 1. Ascending hierarchical definition structure for nouns. (Nouns are at nodes; arrow heads: definiens; arrow tails: definiendum.)
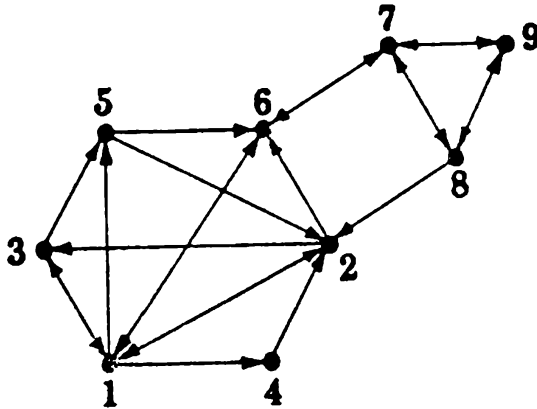
Fig. 2. Closed heterarchical definition structure
for verbs. (Verbs are at nodes; arrow heads:
definiens; arrow tails: definiendum.)

{(hit, blow, ...) (stir, move air, sound, soothe, lay eggs, ..., boast)
(strike, blow, bump, collide ...)} → etc.]

$$v_i = \{v_j\} \vee \sum v_k \vee \prod v_e \qquad (2)$$

A graphic representation of this basically closed heterarchical structure
is given in Fig. 2, and its corresponding representation in form of finite
matrices is discussed elsewhere (Von Foerster, 1966).

The essential difference in the cognitive processes that allow for
identification of forms and those of change of forms is not only reflected
in the entirely different formalisms needed for representing the different
definition structures of nouns [Eq. (1)] and of verbs [Eq. (2)], but also
by the fact that the set of invariants that identify shape under various
transformations can be computed by a single *deductive* algorithm (Pitts
and McCulloch, 1947), while identification of even elementary notions of
behavior requires *inductive* algorithms that can only be computed by
perpetual comparison of present states with earlier states of the system
under consideration (Von Foerster *et al.*, 1968).

These cognitive handicaps put the ethologist at a considerable dis-
advantage in developing a phenomenology for his subject matter when
compared to his colleague the geneticist. Not only are the tools of ex-
pressing his phenomena devoid of the beautiful isomorphism which prevails
between the hierarchical structures of all taxonomies and the definition
of nouns that describe them, but, he may fall victim to a semantic trap
which tempts him to associate with a conceptually isolable function a
corresponding isolable mechanism that generates this function. This temp-
tation seems to be particularly strong when our vocabulary suggests a
variety of conceptually separable higher mental faculties as, for instance
"to learn," "to remember," "to perceive," "to recall," "to predict," etc.,

and the attempt is made to identify and localize within the various parts of our brain the mechanisms that learn, remember, perceive, recall, predict, etc. The hopelessness of a search for mechanisms that represent these functions in isolation does not have a physiological basis as, for instance, "the great complexity of the brain," "the difficulty of measurement," etc. This hopelessness has a purely semantic basis. Memory, for instance, contemplated in isolation is reduced to "recording," learning to "change," perception to "input," and so on. In other words, in separating these functions from the totality of cognitive processes, one has abandoned the original problem and is now searching for mechanisms that implement entirely different functions which may or may not have any semblance to some processes that are subservient to the maintenance of the integrity of the organism as a functioning unit (Maturana, 1969).

Consider the two conceivable definitions for memory:

(a) An organism's potential awareness of past experiences.
(b) An observed change of an organism's response to like sequences of events.

While definition A postulates a faculty (memory$_A$) in an organism whose inner experience cannot be shared by an outside observer, definition B postulates the same faculty (memory$_A$) to be operative in the observer only—otherwise he could not have developed the concept of "change"—but ignores this faculty in the organism under observation, for an observer cannot "in principle" share the organism's inner experience. From this follows definition B.

It is definition B which is generally believed to be the one which obeys the ground rules of "the scientific method," as if it were impossible to cope scientifically with self-reference, self-description, and self-explanation, i.e., closed logical systems that include the referee in the reference, the descriptor in the description, and the axioms in the explanation.

This belief is unfounded. Not only are such logical systems extensively studied (e.g., Gunther, 1967; Löfgren, 1968), but also neurophysiologists (Maturana *et al.*, 1968), experimental psychologists (Konorski, 1962), and others (Pask, 1968; Von Foerster, 1969) have penetrated to such notions.

These preliminaries suggest that the explorer of mechanisms of mentation has to resolve two kinds of problems, only one of which belongs to physiology or, as it were, to physics; the other one is that of semantics. Consequently, it is proposed to reexamine some present notions of learning and memory as to the category to which they belong, and to sketch a conceptual framework in which these notions may find their proper place.

The next section, "Theory," reviews and defines concepts associated with learning and memory in the framework of a unifying mathematical

formalism. In the Section III various models of interaction of molecules with functional units of higher organization are discussed.

II. THEORY

A. General Remarks

Since we have as yet no comprehensive theory of behavior, we have no theory of learning and, consequently, no theory of memory. Nevertheless, there exists today a whole spectrum of conceptual frameworks ranging from the most naive interpretations of learning to the most sophisticated approaches to this phenomenon. On the naive side, "learning" is interpreted as a change of ratios of the occurrence of an organism's actions which are predetermined by an experimenter's ability to discriminate such actions and his value system, which classifies these actions into "hits" and "misses." Changes are induced by manipulating the organism through electric shocks, presentations of food, etc., or more drastically by mutilating, or even removing, some of the organism's organs. "Teaching" in this frame of mind is the administration of such "reinforcements" which induce the changes observed on other occasions.

On the sophisticated side, learning is seen as a process of evolving algorithms for solving categories of problems of ever-increasing complexity (Pask, 1968), or of evolving domains of relations between the organism and the outside world, of relations between these domains, etc. (Maturana, 1969). Teaching in this frame of mind is the facilitation of these evolutionary processes.

Almost directly related to the level of conceptual sophistication of these approaches is their mathematical naiveté, with the conceptually primitive theories obscuring their simplicity by a smoke screen of mathematical proficiency, and the sophisticated ones failing to communicate their depth by the lack of a rigorous formalism. Among the many causes for this unhappy state of affairs one seems to be most prominent, namely, the extraordinary difficulties that are quickly encountered when attempts are made to develop mathematical models that are commensurate with our epistemological insight. It may require the universal mind of a John von Neumann to give us the appropriate tools. In their absence, however, we may just browse around in the mathematical tool shop, and see what is available and what fits best for a particular purpose.

In this paper the theory of "finite state machines" has been chosen as a vehicle for demonstrating potentialities and limitations of some concepts in theories of memory, learning, and behavior mainly for two reasons. One is that it provides the most direct approach to linking a system's

external variables as, e.g., stimulus, response, input, output, cause, effect, etc., to states and operations that are internal to the system. Since the central issue of a book on "molecular mechanisms in memory and learning" must be the development of a link which connects these internal mechanisms with their manifestations in overt behavior, the "finite state machine" appears to be a useful model for this task.

The other reason for this choice is that the interpretations of its formalism are left completely open, and may as well be applied to the animal as a whole; to cell assemblies within the animal; to single cells and their operational modalities, for instance, to the single neuron; to subcellular constituents; and, finally, to the molecular building blocks of these constituents.

With due apologies to the reader who is used to a more extensive and rigorous treatment of this topic, the essential features of this theory will be briefly sketched to save those who may be unfamiliar with this formalism from having to consult other sources (Ashby, 1956; Ashby, 1962; Gill, 1962).

B. Finite State Machines

1. Deterministic Machines

Essentially, the theory of finite state machines is that of computation. It postulates two finite sets of external states called "input states" and "output states," one finite set of "internal states," and two explicitly defined operations (computations) which determine the instantaneous and temporal relations between these states.*

Let $\mathcal{X}_i$ ($i = 1, 2, \ldots, n_x$) be the n_x receptacles for inputs x_i each of which can assume a finite number, $v_i > 0$, of different values. The number

---

* Although the interpretation of states and operations with regard to observables is left completely open, some caution is advisable at this point if these are to serve as mathematical models, say, for the behavior of a living organism. A specific physical spatiotemporal configuration which is identifiable by the experimenter who wishes that this configuration be appreciated by the organism as a "stimulus" cannot *sui modo* be taken as "input state" for the machine. Such a stimulus may be a stimulant for the experimenter, but be ignored by the organism. An input state, on the other hand, cannot be ignored by the machine, except when explicitly instructed to do so. More appropriately, the distribution of the activity of the afferent fibers has to be taken as an input, and similarly, the distribution of activity of efferent fibers may be taken as the output of the system.

of distinguishable input states is then

$$X = \prod_{i=1}^{n_x} v_i \tag{3}$$

A particular input state $x(t)$ at time t (or x for short) is then the identification cf the values x_i on all n_x input receptacles $\mathfrak{X}_i$ at that "moment":

$$x(t) \equiv x = \{x_i\}, \tag{4}$$

Similarly, let $\mathfrak{Y}_j$ ($j = 1, 2, \ldots, n_y$) be the n_y outlets for outputs y_j, each of which can assume a finite number, $v_j > 0$, of different values. The number of distinguishable output states is then

$$Y = \prod_{j=1}^{n_y} v_j \tag{5}$$

A particular output state $y(t)$ at time t (or y for short) is then the identification of the values y_i on all n_y outlets $\mathfrak{Y}_j$ at that "moment":

$$y(t) \equiv y = \{y_i\} \tag{6}$$

Finally, let Z be the number of internal states z which, for this discussion (unless specified otherwise), may be considered as being not further analyzable. Consequently, the values of z may just be taken to be the natural numbers from 1 to Z, and a particular output state $z(t)$ at time t (or z for short) is the identification of z's value at that "moment":

$$z(t) \equiv z \tag{7}$$

Each of these "moments" is to last a finite interval of time, Δ, during which the values of all variables x, y, z are identifiable. After this period, i.e., at time $t + \Delta$, they assume values $x(t + \Delta)$, $y(t + \Delta)$, $z(t + \Delta)$ (or x', y', z' for short), while during the previous period $t - \Delta$ they had values $x(t - \Delta)$, $y(t - \Delta)$, $z(t - \Delta)$ (or x^*, y^*, z^* for short).

After having defined the variables that will be operative in the machine we are now prepared to define the operations on these variables. These are two kinds and may be specified in a variety of ways. The most popular procedure is first to define a "driving function" which determines at each instant the output state, given the input state and the internal state at that instant:

$$y = f_y(x, z) \tag{8}$$

Although the driving function f_y may be known and the time course of input states x may be controlled by the experimenter, the output states y as time goes on are unpredictable as long as the values of z, the internal

157

states of the machine, are not yet specified. A large variety of choices are open to specify the time course of z as depending on x, on y, or on other newly to be defined internal or external variables. The most profitable specification for the purposes at hand is to define z recursively as being dependent on previous states of affairs. Consequently, we define the "state function" f_z of the machine to be:

$$z = f_z(x^*, z^*) \tag{9a}$$

or alternately and equivalently

$$z' = f_z(x, z) \tag{9b}$$

that is, the present internal state of the machine is a function of its previous internal state and its previous input state; or alternately and equivalently, the next internal machine state is a function of both its present internal and input states.

With the three sets of states $\{x\}$, $\{y\}$, $\{z\}$ and the two functions f_y and f_z, the behavior of the machine, i.e., its output sequence, is completely determined if the input sequence is given.

Such a machine is called a sequential, state-determined, "nontrivial" machine and in Fig. 3a the relations of its various parts are schematically indicated.

Such a nontrivial machine reduces to a "trivial" machine if it is insensitive to changes of internal states, or if the internal states do not change (Fig. 3b):

$$z' = z = z_0 = \text{constant} \tag{10a}$$

$$y = f_y(x, \text{constant}) = f(x) \tag{10b}$$
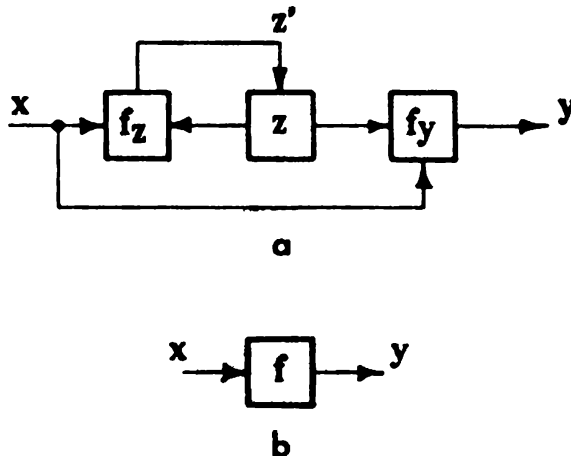
Fig. 3. Signal flow in a finite state machine (a); input–output relation in a trivial machine (b).

In other words, a trivial machine is one which couples deterministically a particular input state with a specific output state or, in the language of naive reflexologists, a particular stimulus with a specific response.

Since the concept of "internal states" is crucial in appreciating the difference between a trivial and a nontrivial machine, we shall now give various formal interpretations of these states to lift them from the limbo of "being not further analyzable."

First, it may appear that by an artifice one can get rid of these mysterious states by defining the driving function f_ν in a recursive form. However, as we shall see shortly, these states reappear in just another form.

Consider the driving function [Eq. (8)] at time t and one step later $(t + \Delta)$:

$$y = f_\nu(x, z)$$
$$y' = f_\nu(x', z') \tag{8'}$$

and assume there exists an "inverse function" to f_ν:

$$z = \phi_\nu(x, y) \tag{11}$$

We now enter the state function [Eq. (9b)] for z' into Eq. (8') and replace z by Eq. (11):

$$y' = f_\nu(x', f_z(x, \phi_\nu(x, y))) = F_\nu^{(1)}(x', x, y) \tag{12}$$

or alternately and equivalently

$$y = F_\nu^{(1)}(x, x^*, y^*) \tag{13}$$

However, y^* is given recursively through Eq. (13)

$$y^* = F_\nu^{(1)}(x^*, x^{**}, y^{**}) \tag{13*}$$

and inserting this into Eq. (13) we have

$$y = F_\nu^{(2)}(x, x^*, x^{**}, y^{**})$$

and for n recursive steps

$$y = F_\nu^{(n)}(x, x^*, x^{**}, x^{***} \cdots x^{(n)*}, y^{(n)*}) \tag{14}$$

This expression suggests that in a nontrivial machine the output is not merely a function of its present input, but may be dependent on the particular sequence of inputs reaching into the remote past, and an output state at this remote past. While this is only to a certain extent true—the "remoteness" is carried only over Z recursive steps and, moreover, Eq. (14) does not uniquely determine the properties of the machine—this dependence of the machine's behavior on its past history should not tempt one to project into this system a capacity for memory, for at best it may

look upon its present internal state which may well serve as *token* for the past, but without the powers to recapture for the system all that which has gone by.

This may be most easily seen when Eq. (13) is rewritten in its full recursive form for a linear machine (with x and y now real numbers)

$$y(t + \Delta) - ay(t) = bx(t) \qquad (15a)$$

or in its differential analog expanding $y(t + \Delta) = y(t) + \Delta dy/dt$:

$$\frac{dy}{dt} - \alpha y = x(t) \qquad (15b)$$

with the corresponding solutions

$$y(n\Delta) = a^n\Big[y(0) + b \sum_{i=0}^{n} a^{-i}x(i\Delta)\Big] \qquad (16a)$$

and

$$y(t) = e^{\alpha t}\left[y(0) + \int_0^t e^{-\alpha\tau}x(\tau)d\tau\right] \qquad (16b)$$

From these expressions it is clear that the course of events represented by $x(i\Delta)$ (or $x(\tau)$) is "integrated out," and is manifest only in an additive term which, nevertheless, changes as time goes on.

However, the failure of this simple machine to account for memory should not discourage one from contemplating it as a possible useful element in a system that remembers.

While in these examples the internal states z provided the machine with an appreciation—however small—of its past history, we shall now give an interpretation of the internal states z as being a selector for a specific function in a set of multivalued logical functions. This is most easily seen when writing the driving function f_v in form of a table.

Let a, b, c ... X be the input values x; α, β, γ ... Y be the output values y; and $1, 2, 3 \ldots Z$ be the values of the internal states. A particular driving function f_v is defined if to all pairs $\{xz\}$ an appropriate value of y is associated. This is suggested in Table I.

Clearly, under $z = 1$ a particular logical function, $y = F_1(x)$, relating y with x is defined; under $z = 2$ another logical function, $y = F_2(x)$, is defined; and, in general, under each z a certain logical function $y = F_z(x)$ is defined.

Hence, the driving function f_v can be rewritten to read

$$y = F_z(x), \qquad (17)$$

TABLE I

Computing Z Logical Function $F_s(x)$ on Inputs x

z	1	1	1	... 1	2	2	2	... 2	...	Z	Z	Z	... Z
x	a	b	c	... X	a	b	c	... X	...	a	b	c	... X
y	γ	α	β	... δ	α	γ	β	... ϵ	...	β	ϵ	γ	... δ

which means that this machine computes another logical function $F_{s'}$ on its inputs x, whenever its internal state z changes according to the state function $z' = f_z(x, z)$.

Or, in other words, whenever z changes, the machine becomes a *different* trivial machine.

While this observation may be significant in grasping the fundamental difference between nontrivial and trivial machines, and in appreciating the significance of this difference in a theory of behavior, it permits us to calculate the number of internal states that can be effective in changing the *modus operandi* of this machine.

Following the paradigm of calculating the number $\mathfrak{N}$ of logical functions as the number of states of the dependent variable raised to the power of the number of states of the independent variables

$$\mathfrak{N} = \text{(no. of states of dep. variables)}^{\text{(no. of states of indep. variables)}} \qquad (18)$$

we have for the number of possible trivial machines which connect y with x

$$\mathfrak{N}_T = Y^X \qquad (19)$$

This, however, is the largest number of internal states which can effectively produce a change in the function $F_s(x)$, for any additional state has to be paired up with a function to which a state has been already assigned, hence such additional internal states are redundant or at least indistinguishable. Consequently

$$Z \leq Y^X$$

Since the total number of driving functions $f_y(x, z)$ is

$$\mathfrak{N}_D = Y^{XZ}, \qquad (20)$$

its largest value is:

$$\bar{\mathfrak{N}}_D = Y^{X Y^X} \qquad (21)$$

Similarly, for the number of state functions $f_z(z, x)$ we have

$$\mathfrak{N}_S = Z^{X \cdot Z} \tag{22}$$

whose largest effective value is

$$\bar{\mathfrak{N}}_S = Y^{X \cdot X Y^X} = [\bar{\mathfrak{N}}_D]^X \tag{23}$$

These numbers grow very quickly into meta-astronomical magnitudes even for machines with most modest aspirations.

Let a machine have only one two-valued output ($n_y = 1$; $v_y = 2$; $y = \{0; 1\}$; $Y = 2$) and n two-valued inputs ($n_x = n$; $v_x = 2$; $x = \{0; 1\}$; $X = 2^n$). Table II gives the number of effective internal states, the number of possible driving functions, and the number of effective state functions for machines with from one to four "afferents" according to the equations

$$Z = 2^{2^n}$$

$$\mathfrak{N}_D = 2^{2^{2n+n}}$$

$$\mathfrak{N}_S = 2^{2^{2n+2n}}$$

These fast-rising numbers suggest that already on the molecular level without much ado a computational variety can be met which defies imagination. Apparently, the large variety of results of genetic computation, as manifest in the variety of living forms even within a single species, suggests such possibilities. However, the discussion of these possibilities will be reserved for the next section.

TABLE II

The Number of Effective Internal States Z, the Number of Possible Driving Functions $\mathfrak{N}_D$, and the Number of Effective State Functions $\mathfrak{N}_S$ for Machines with One Two-Valued Output and with from One to Four Two-Valued Inputs

n	Z	$\mathfrak{N}_D$	$\mathfrak{N}_S$
1	4	256	65536
2	16	$2 \cdot 10^{19}$	$6 \cdot 10^{76}$
3	256	10^{600}	$300 \cdot 10^{4 \cdot 10^3}$
4	65536	$300 \cdot 10^{4 \cdot 10^3}$	$1600 \cdot 10^{7 \cdot 10^4}$

2. *Interacting Machines*

We shall now discuss the more general case in which two or more such machines interact with each other. If some aspects of the behavior of an organism can be modeled by a finite state machine, then the interaction of the organism with its environment may be such a case in question, if the environment is likewise representable by a finite state machine. In fact, such two-machine interactions constitute a popular paradigm for interpreting the behavior of animals in experimental learning situations, with the usual relaxation of the general complexity of the situation, by chosing for the experimental environment a trivial machine. "Criterion" in these learning experiments is then said to have been reached by the animal when the experimenter succeeded in transforming the animal from a nontrivial machine into a trivial machine, the result of these experiments being the interaction of just two trivial machines.

We shall denote quantities pertaining to the environment (E) by Roman letters, and those to the organism (Ω) by the corresponding Greek letters. As long as E and Ω are independent, six equations determine their destiny. The four "machine equations," two for each system

$$E: \quad y = f_y(x, z) \tag{24a}$$

$$z' = f_z(x, z) \tag{24b}$$

$$\Omega: \quad \eta = f_\eta(\xi, \zeta) \tag{25a}$$

$$\zeta' = f_\zeta(\xi, \zeta) \tag{25b}$$

and the two equations that describe the course of events at the "receptacles" of the two systems

$$x = x(t); \quad \xi = \xi(t) \tag{26a, b}$$

We now let these two systems interact with each other by connecting the (one step delayed) output of each machine with the input of the other. The delay is to represent a "reaction time" (time of computation) of each system to a given input (stimulus, cause) (see Fig. 4). With these connections the following relations between the external variables of the two systems are now established:

$$x' = \eta = u'; \quad \xi' = y = v' \tag{27a, b}$$

where the new variables u, v represent the "messages" transmitted from $\Omega \to E$ and $E \to \Omega$ respectively. Replacing x, y, η, ξ, in Eqs. (24)(25) by u, v according to Eq. (27) we have

$$v' = f_y(u, z); \quad u' = f_\eta(v, \zeta)$$
$$z' = f_z(u, z); \quad \zeta' = f_\zeta(v, \zeta) \tag{28}$$
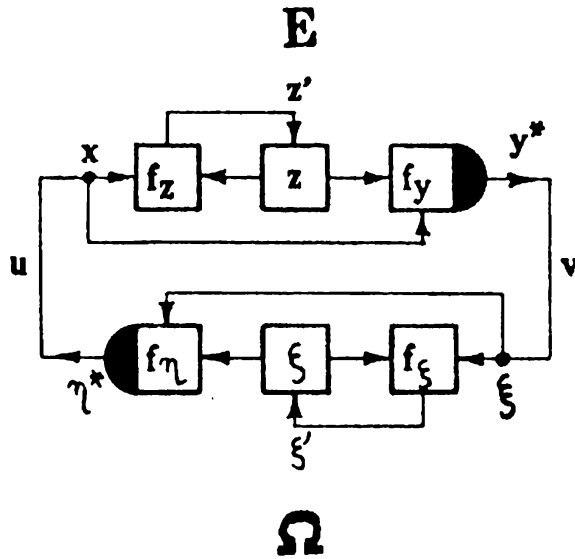
Fig. 4. Two finite state machines (E) (Ω) connected via delays (black semicircles).

These are four recursive equations for the four variables u, v, z, ζ, and if the four functions f_v, f_z, f_η, f_ζ are given, the problem of "solving" for $u(t)$, $v(t)$, $z(t)$, $\zeta(t)$, i.e., expressing these variables explicitly as functions of time, is purely mathematical. In other words, the "meta-system" $(E\Omega)$ composed of the subsystems E and Ω, is physically as well as mathematically "closed," and its behavior is completely determined for all times. Moreover, if at a particular time, say $t = 0$ (initial condition), the values of all variables $u(0)$, $v(0)$, $z(0)$, $\zeta(0)$ are known, it is also completely predictable. Since this meta-system is without input, it churns away according to its own rules, coming ultimately to a static or dynamic equilibrium, depending on the rules and the initial conditions.

In the general case the behavior of such systems has been extensively studied by computer simulation (Walker, 1965; Ashby and Walker, 1966; Fitzhugh, 1963), while in the linear case the solutions for Eqs. (28) can be obtained in straight-forward manner, particularly if the recursions can be assumed to extend over infinitesimally small steps:

$$w' = w(t + \Delta) = w(t) + \Delta \frac{dw}{dt} \qquad (29)$$

Under these conditions the four Eqs. (28) become

$$\dot{w}_i = \sum_{j=1}^{4} a_{ij} w_j \qquad (30)$$

where the w_i ($i = 1, 2, 3, 4$) are now the real numbers and replace the four variables in question, $\dot{w}$ represents the first derivative with respect to time, and the 16 coefficients a_{ij} ($i, j = 1, 2, 3, 4$) define the four linear functions under consideration. This system of simultaneous, first-order, linear differential equations is solved by

$$w_i(t) = \sum_{j=1}^{4} A_{ij} e^{\lambda_j t} \tag{31}$$

in which λ_j are the roots of the determinant

$$| a_{ij} - \delta_{ij}\lambda | = 0 \tag{32}$$

$$\delta_{ij} = \begin{cases} 1 \ldots i = j \\ \\ 0 \ldots i \neq j \end{cases}$$

and the A_{ij} depend on the initial conditions. Depending on whether the λ_j turn out to be complex, real negative or real positive, the system will ultimately oscillate, die out, or explode.*

While a discussion of the various modes of behavior of such systems goes beyond this summary, it should be noted that a common behavioral feature in all cases is an initial transitory phase that may move over a very large number of states until one is reached that initiates a stable cyclic trajectory, the dynamic equilibrium. Form and length of both the transitory and final equilibrial phases are dependent on the initial conditions, a fact which led Ashby (1956) to call such systems "multistable." Since usually a large set of initial conditions maps into a single equilibrium, this equilibrium may be taken as a *dynamic representation* of a set of events, and in a multistable system each cycle as an "abstract" for these events.

With these notions let us see what can be inferred from a typical learning experiment (e.g., John *et al.*, 1969) in which an experimental animal in a Y-maze is given a choice ($\xi_0 \equiv C$, for "choice") between two actions ($\eta_1 \equiv L$, for "left turn"; $\eta_2 \equiv R$, for "right turn"). To these the environment E, a trivial machine, responds with new inputs to the animal ($\eta_1 = x_1' \rightarrow y_1' = \xi_1'' \equiv S$, for "shock"; or $\eta_2 = x_2' \rightarrow y_2' = \xi_2'' \equiv F$, for "food"), which, in turn, elicit in the animal a pain ($\eta_3 \equiv$ "−") or pleasure ($\eta_4 \equiv$ "+") response. These responses cause E to return the animal to the original choice situation ($\xi_0 \equiv C$).

Consider the simple survival strategy built into the animal by which

---

* This result is, of course, impossible in a finite state machine. It is obtained here only because of the replacement of the discrete and finite variables u, v, z, ζ, by w_i which are continuous and unlimited quantities.

under neutral and pleasant conditions it maintains its internal state $[\zeta' = \zeta$, for $(C\zeta)$ and $(F\zeta)]$, while under painful conditions it changes it $[\zeta' \neq \zeta$, for $(S\zeta)]$. We shall assume eight internal states $(\zeta = i;\ i = 1, 2, 3, \ldots, 8)$.

With these rules the whole system (ΩE) is specified and its behavior completely determined. For convenience, the three functions, $f_y = f$ for the trivial machine E, f_η and f_ζ for Ω are tabulated in Tables IIIa, b, c.

With the aid of these tables the eight behavioral trajectories for the (ΩE) system, corresponding to the eight initial conditions, can be written. This has been done below, indicating only the values of the pairs $\xi\zeta$ as

TABLE IIIa

$$y = f(x)$$

$x\ (= \eta^*)$	$y\ (= \xi')$
L	S
R	F
$-$	C
$+$	C

TABLE IIIb

$$\eta = f_\eta(\xi, \zeta)$$

$\eta\ (= x')$		ζ							
		1	2	3	4	5	6	7	8
	C	L	L	L	L	R	R	R	R
$\xi\ (= y^*)$	S	$-$	$-$	$-$	$-$	$-$	$-$	$-$	$-$
	F	$+$	$+$	$+$	$+$	$+$	$+$	$+$	$+$

TABLE IIIc

$$\zeta' = f_\zeta(\xi, \zeta)$$

ζ'		ζ							
		1	2	3	4	5	6	7	8
	C	1	2	3	4	5	6	7	8
$\xi\ (= y^*)$	S	2	3	4	5	6	7	8	1
	F	1	2	3	4	5	6	7	8

they follow each other as consequences of the organism's responses and the environment's reactions.

C1 → S1 → C2 → S2 → C3 → S3 → C4 → S4 → C5 → F5 → C5 ⌐

C2 ──────────┘

C3 ─────────────────────┘

C4 ───────────────────────────────────┘

C5 ───┘

C6 → F6 → C6 ──┘

C7 → F7 → C7 ──┘

C8 → F8 → C8 ──┘
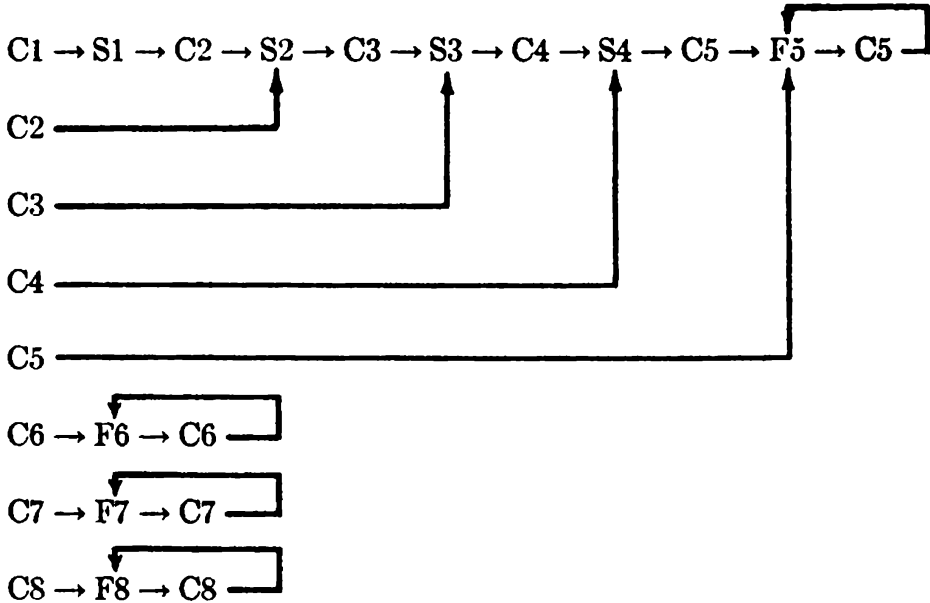
These trajectories show indeed the behavior as suggested before, initial transients depending in length on the initial conditions, and ultimately a dynamic equilibrium flipping back and forth between two external states without internal state change. The whole system, and its parts, have become trivial machines. Since, even with maximum semantic tolerance, one cannot say a trivial machine has memory, one wonders what is intended to be measured when at this stage it is opened and the internal workings are examined. Does one wish to inspect its present workings? Or, to see how much it has changed since earlier examinations? At best, these are tests of the experimenter's memory, but whether the machine can appreciate any changes cannot, in principle, be inferred from experiments whose conceptual design eliminates the quality which they intend to measure.

3. Probabilistic Machines

This dilemma can be seen in still another light if we adopt for the moment the position of statistical learning theory (Skinner, 1959; Estes, 1959; Logan, 1959). Here either the concept of internal states is rejected or the existence of internal states is ignored. But whenever the laws which connect causes with effects are ignored, either through ignorance or else by choice, the theory becomes that of probabilities.

If we are ignorant of the initial state in the previous example, the chances are 50/50 that the animal will turn left or right on its first trial. After one run the chances are 5/8 for turning right, and so on, until the animal has turned from a "probabilistic (nontrivial) machine" to a "deterministic (trivial) machine," and henceforth always turns right. While a statistical learning theoretician will elevate the changing probabilities in each of the subsequent trials to a "first principle," for the finite state machinist this is an obvious consequence of the effect of certain inputs on the internal states of his machine: they become inaccessible when paired with "painful inputs." Indeed, the whole mathematical machinery of statistical learning theory can be reduced to the paradigm of drawing balls of different color from an urn while observing certain non-replacement rules.

Let there be an urn with balls of m different colors labeled 0, 1, 2, ..., $(m - 1)$. As yet unspecified rules permit or prohibit the return of a certain colored ball when drawn. Consider the outcomes of a sequence of n drawings, an "n-sequence," as being an n digit m-ary number (e.g., $m = 10$; $n = 12$):

$$\nu = 1\ 5\ 7\ 3\ 0\ 2\ 1\ 8\ 6\ 2\ 1\ 4$$

Last
drawn

First
drawn

From this it is clear that there are

$$\mathfrak{N}(n, m) = m^n$$

different n-sequences. A *particular* n-sequence will be called a ν-number, i.e.:

$$0 \leq \nu(m, n) = \sum_{i=1}^{n} j(i) m^{(i-1)} \leq m^n_{-1} \tag{33}$$

where $0 \leq j(i) \leq (m - 1)$ represents the outcome labeled j at the ith trial.

The probability of a *particular* n-sequence (represented by a ν-number) is then

$$p_n(\nu) = \prod_{i=1}^{n} p_i[\, j(i)\,] \tag{34}$$

where $p_i[\, j(i)\,]$ gives the probability of the color labeled j to occur at the ith trial in accordance with the specific ν-number as defined in Eq. (33).

Since after each trial with a "don't return" outcome all probabilities are changed, the probability of an event at the nth trial is said to depend

on the "path," i.e., on the past history of events, that led to this event. Since there are m^{n-1} possible paths that may precede the drawing of j at the nth trial, we have for the probability of this event:

$$p_n(j) = \sum_{\nu=0}^{m^{n-1}-1} p_n(j \cdot m^{n-1} + \nu(n-1, m))$$

where $j \cdot m^{n-1} + \nu(n-1, m)$ represent a $\nu(n, m)$-number which begins with j.

From this a useful recursion can be derived. Let j^* be the colors of balls which when drawn are *not* replaced, and j the others. Let n_{j^*} and n_j be the number of preceding trials on which j^* and j came up respectively $(\sum n_{j^*} + \sum n_j = n - 1)$, then the probability for drawing j (or j^*) at the nth trial with a path of $\sum n_{j^*}$ withdrawals is

$$p_n(j) = \frac{N_j}{N - \sum n_{j^*}} \cdot p_{n-1}(\sum n_{j^*}) \tag{35a}$$

and

$$p_n(j^*) = \frac{N_{j^*} - n_{j^*}}{N - \sum n_{j^*}} \cdot p_{n-1}(\sum n_{j^*}) \tag{35b}$$

where $N = \sum N_j + \sum N_{j^*}$ is the initial number of balls, and N_j and N_{j^*} the initial number of balls with colors j and j^* respectively.

Let there be N balls to begin with in an urn, N_w of which are white, and $(N - N_w)$ are black. When a white ball is drawn, it is returned; a black ball, however, is removed. With "white" $\equiv 0$, and "black" $\equiv 1$, a particular n-sequence $(n = 3)$ may be

$$\nu(3, 2) = 1\ 0\ 1$$

and its probability is:

$$p_3(1\ 0\ 1) = \frac{N - N_w - 1}{N - 1} \cdot \frac{N_w - 1}{N - 1} \cdot \frac{N - N_w}{N}$$

The probability of drawing a black ball at the third trial is then:

$$p_3(1) = p_3(1\ 0\ 0) + p_3(1\ 0\ 1) + p_3(1\ 1\ 0) + p_3(1\ 1\ 1)$$

We wish to know the probability of drawing a white ball at the nth trial. We shall denote this probability now by $p(n)$, and that of drawing a black ball $q(n) = 1 - p(n)$.

By iteratively approximating [through Eq. (35)] trial tails of length m as being path independent $[p_i(j) = p_1(j)]$ one obtains a first-order

approximation for a recursion in $p(n)$:

$$p(n) = p(n - m) + \frac{m}{N} q(n - m) \tag{36}$$

or for $m = n - 1$ (good for $p(1) \approx 1$, and $n/N \ll 1$):

$$p(n) = p(1) + \frac{n - 1}{N} q(1) \tag{37}$$

and for $m = 1$ (good for $p(1) \approx 1$):

$$p(n) = p(n - 1) + \frac{1}{N} q(n - 1) \tag{38}$$

A second approximation changes the above expression to

$$p(n) = p(n - 1) + \theta q(n - 1) \tag{39}$$

where $\theta = \theta(N, N_w)$ is a constant for all trials. With this we have

$$p(n) - p(n - 1) = \Delta p = \theta(1 - p) \tag{40}$$

which, in the limit for

$$\lim_{\Delta n \to 0} \frac{\Delta p}{\Delta n} = \frac{dp}{dn}$$

gives

$$\frac{dp}{dn} = \theta(1 - p(n))$$

with the solution

$$p(n) = 1 - (1 - p_0)e^{-\theta n} \tag{41}$$

This, in turn, is an approximation for $p \approx 1$ of

$$p(n) = \frac{p_0}{p_0 + (1 - p_0)e^{-\theta n}} \tag{42}$$

which is the solution of

$$\frac{dp}{dn} = \theta p(1 - p) \tag{43}$$

or, recursively expressed, of

$$p(n) = p(n - 1) + \theta p(n - 1) \cdot q(n - 1) \tag{44}$$
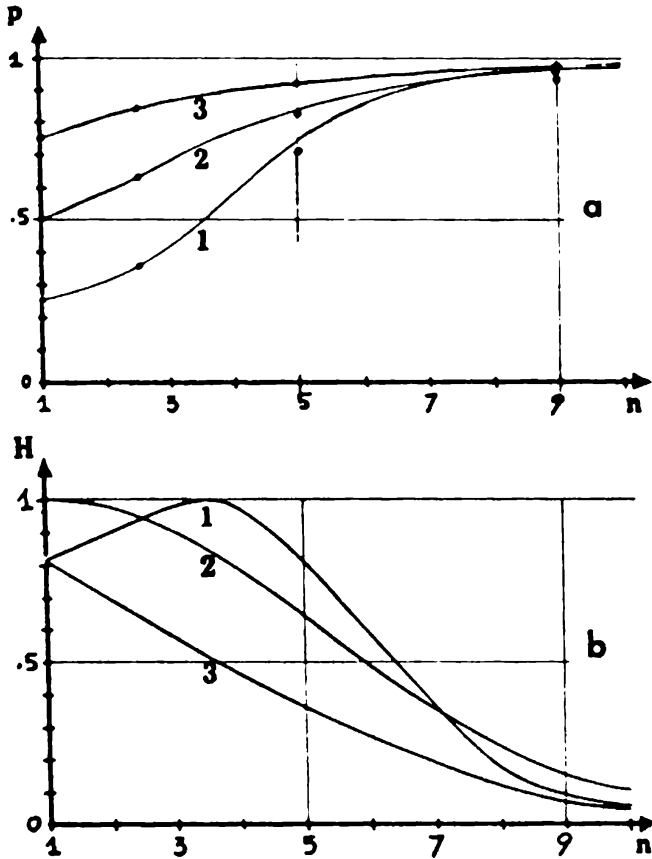
Fig. 5. Probability for drawing a white ball at the nth trial from an urn having initially four balls of which 1, 2, or 3 are white, the others black. White balls are replaced, black are not (a). Entropy at the nth trial (b).

Figure 5a compares the probabilities $p(n)$ for drawing a white ball at the nth trial, as calculated through approximation [Eq. (42)] (solid curves), with the exact values computed by an IBM 360/50 system with a program kindly supplied by Mr. Atwood for an urn with initially four balls ($N = 4$) and for the three cases in which one, two, or three of these are white ($N_w = 1$; $N_w = 2$; $N_w = 3$). The entropy* $H(n)$ in bits per trial corresponding to these cases is shown in Fig. 5b, and one may note that while for some cases [$p(1) \leq 0.5$] it reaches a maximum in the course of this game, it vanishes in all cases when certainty of the outcome is approached [$p(n) \to 1$].

Although the sketch on probabilities dealt exclusively with urns, balls, and draws, students of statistical learning theory will have recognized in Eqs. (39), (41), and (42) the basic axioms of this theory [Estes, 1959;

---

* Or the "amount of uncertainty"; or the "amount of information" received by the outcome of each trial, defined by $-H(n) = p(n) \log_2 p(n) + q(n) \log_2 q(n)$.

Eqs. (5), (6), and (9)], and there is today no doubt that under the given experimental conditions animals will indeed trace out the learning curves derived for these conditions.

Since the formalism that applies to the behavior of these experimental animals applies as well to our urn, the question now arises: can we say an urn learns? If the answer is "yes," then apparently there is no need for *memory* in learning, for there is no trace of black balls left in our urn when it finally "responds" correctly with white balls when "stimulated" by each draw; if the answer is "no," then by analogy we must conclude it is not *learning* that is observed in these animal experiments.

To escape this dilemma it is only necessary to recall that an urn is just an urn, and it is animals that learn. Indeed, in these experiments learning takes place on two levels. First, the experimental animals learned to behave "urnlike," or better, to behave in a way which allows the experimenter to apply urnlike criteria. Second, the experimenter learned something about the animals by turning them from nontrivial (probabilistic) machines into trivial (deterministic) machines. Hence, it is from studying the experimenter whence we get the clues for memory and learning.

C. Finite Function Machines

1. *Deterministic Machines*

With this observation the question of where to look for memory and learning is turned into the opposite direction. Instead of searching for mechanisms in the environment that turn organisms into trivial machines, we have to find the mechanisms within the organisms that enable them to turn their environment into a trivial machine.

In this formulation of the problem it seems to be clear that in order to manipulate its environment an organism has to construct—somehow—an internal representation of whatever environmental regularities it can get hold of. Neurophysiologists have long since been aware of these abstracting computations performed by neural nets from right at the receptor level up to higher nuclei (Lettvin *et al.*, 1959; Maturana *et al.*, 1968; Eccles *et al.*, 1967). In other words, the question here is how to compute functions rather than states, or how to build a machine that computes programs rather than numerical results. This means that we have to look for a formalism that handles "finite *function* machines." Such a formalism is, of course, one level higher up than the one discussed before, but by maintaining some pertinent analogies its essential features may become apparent.

Our variables are now functions, and since relations between functions are usually referred to as "functionals," the essential features of a calculus of recursive functionals will be briefly sketched.

Consider a system like the one suggested in Fig. 3a, with the only difference that it operates on a finite set of functions of two kinds, $\{f_{yi}\}$ and $\{f_{zj}\}$. These functions, in turn, operate on their appropriate set of states $\{y_i\}$ and $\{z_j\}$. The rules of operation for such a finite function machine are modeled exactly according to the rules of finite state machines. Hence:

$$f_y = F_y[x, f_z] \tag{45a}$$

$$f_z' = F_z[x, f_z] \tag{45b}$$

where F_y and F_z are the functionals which generate the driving functions f_y and the subsequent internal function f_z' from the present internal function f_z and an input x. One should note, however, that the input here is still a state. This indicates an important feature of this formalism, namely, the provision of a link between the domain of states with the entirely different domain of functions. In other words, this formalism takes notice of the distinction between entities and their representations and establishes a relation between these two domains.

Following a procedure similar to that carried out in Eqs. (10) through (14), the functions of type f_z can be eliminated by expressing the present driving function as result of earlier states of affairs. However, due to some properties that distinguish functionals from functions, these earlier states of affairs include both input states as well as output functions. We have for n recursive steps:

$$f_y = \Phi_y^{(n)}[x, x^*, x^{**}, x^{***}, \ldots, x^{(n)*}; f_y^*, f_y^{**}, \ldots, f_y^{(n)*}] \tag{46}$$

Comparing this expression with its analog for finite state machines [Eq. (14)], it is clear that here the reference to past events is not only to those events that were the system's history of inputs $\{x^{(i)*}\}$, but also to its history of potential actions $\{f_y^{(i)*}\}$. Moreover, when this recursive functional is solved explicitly for time ($t = k\Delta; k = 0, 1, 2, 3, \ldots$;) [compare with Eq. (16)], it is again the history of inputs that is "integrated out"; however, the history of potential actions remains intact, because of a set of n "eigenfunctions" which satisfy Eq. (46). We have explicitly for $(k\Delta)$, and for the ith eigenfunction:

$$f_y^i(k\Delta) = K_i(k\Delta) \cdot [\pi_i(f_y^{(i)*}) + G_i(x, x^*, x^{**}, \ldots, x^{(n)*})] \tag{47}$$

$$i = 1, 2, 3, \ldots, n$$

with K_i and G_i being functions of $(k\Delta)$, the latter one giving a value that depends on a tail of values in $x^{(j)*}$ which is n steps long. π_i is again a

functional, representing the output function f_ν of i steps in the past in terms of another function.

Although this formalism does not specify any mechanism capable of performing the required computations, it provides us, at least, with an adequate description of the functional organization of memory. Access to "past experience" is given here by the availability of the system's own *modus operandi* at earlier occasions, and it is comfortable to see from expression (47) that the subtle distinction between an experience in the past ($f_\nu^{(i)*}$), and the present experience of an experience in the past $[\pi_i(f_\nu^{(i)*})]$ —i.e., the distinction between "experience" and "memory"—is indeed properly taken care of in this formalism. Moreover, by the system's access to its earlier states of functioning, rather than to a recorded collection of accidental pairs $\{x_i, y_i\}$ that manifest this functioning, it can compute a stream of "data" which are consistent with the system's past experience. These data, however, may or may not contain the output values $\{y_i\}$ of those accidental pairs. This is the price one has to pay for switching domains, from states to functions and back again to states. But this is a small price indeed for the gain of an infinitely more powerful "storage system" which computes the answer to a question, rather than stores all answers together with all possible questions in order to respond with the answer when it can find the question (Von Foerster, 1965).

These examples may suffice to interpret without difficulty another property of the finite function machine that is in strict analogy to the finite state machine. As with the finite state machine, a finite function machine will, when interacting with another system, go through initial
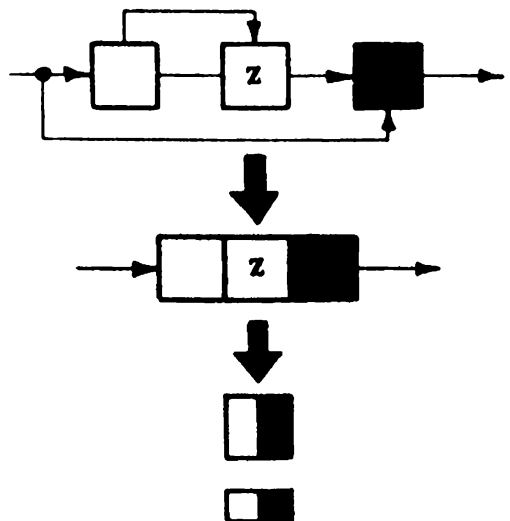
Fig. 6. Symbolization of a finite state machine by a computational tile. Input region white; output region black.

transients depending on initial conditions and settle in a dynamic equilibrium. Again, if there is no internal function change ($f_z' = f_z = f_0$) we have a "trivial finite function machine" with its "goal function" f_0. It is easy to see that a trivial finite function machine is equivalent to a nontrivial finite state machine.[*]

Instead of citing further properties of the functional organization of finite function machines, it may be profitable to have a glance at various possibilities of their structural organization. Clearly, here we have to deal with aggregates of large numbers of finite state machines, and a more efficient system of notation is required to keep track of the operations that are performed by such aggregates.

2. Tessellations

Although a finite state machine consists of three distinct parts, the two computers, f_y and f_z, and the store for z, (see Fig. 3a), we shall represent the entire machine by a single square (or rectangle); its input region denoted white, the output region black (Fig. 6). We shall now treat this unit as an elementary computer—a "computational tile," T_i—which, when combined with other tiles, T_j, may form a mosaic of tiles—a "computational tesselation," 3. The operations performed by the ith tile shall be those of a finite state machine, but different letters, rather than subscripts, will be used to distinguish the two characteristic functions. Subscripts shall refer to tiles.

$$y_i = f_i(x_i, z_i)$$

$$z_i = g_i(x_i, z_i) \tag{48}$$

Figure 7/I sketches the eight possible ways (four each for the parallel and the antiparallel case) in which two tiles can be connected. This results in three classes of elementary tesselations whose structures are suggested in Fig. 7/II. Cases I/1 and I/3, and I/2 and I/4 are equivalent in the parallel case, and are represented in II/1 ("chain") and II/2 ("stack") respectively. In the antiparallel case the two configurations I/1 and I/3 are ineffective, for outputs cannot act on outputs, nor inputs on inputs; cases I/2 and I/4 produce two autonomous elementary tesselations $A = [a^+, a^-]$, distinct only by the sense of rotation in which the signals are processed.

Iterations of the same concatenations result in tesselations with the

---

[*] In the case of several equilibria $\{f_{0i}\}$, we have, of course, a set of nontrivial finite state machines that are the outcomes of various initial conditions.
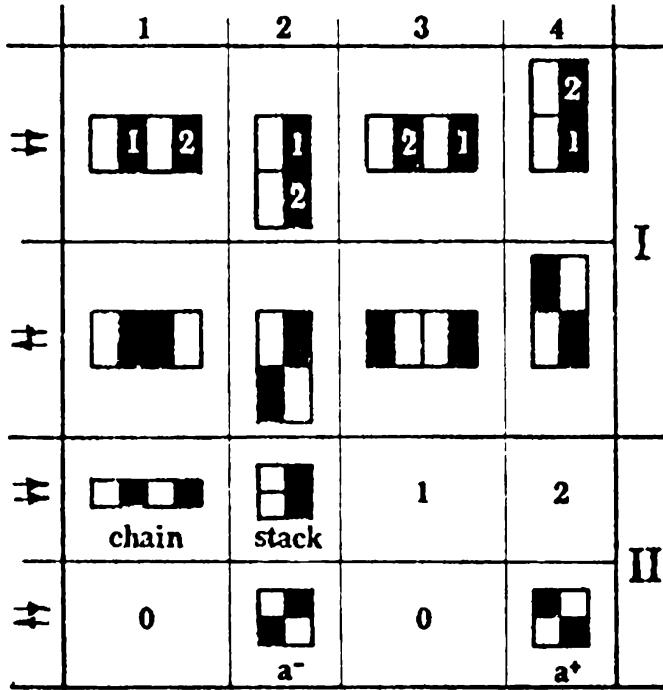
Fig. 7. Elementary tessellations.

following functional properties (for n iterations):

1. *Stack*

$$nT: \qquad y = \sum_1^n f_i(x_i, z_i) \qquad (49)$$

2. *Chain*

$$T^n: \qquad y = f_n(f_{n-1}(f_{n-2}\ldots(x^{(n)*}, z^{(n)*})\ldots z^{**}_{n-2})z^*_{n-1})z_n \qquad (50)$$

3. $A = \{a^+, a^-\}$

$$\left.\begin{array}{l} a^+a^- \\ a^-a^+ \end{array}\right\} = 0 \qquad \left.\begin{array}{l} a^+a^+ \\ a^-a^- \end{array}\right\} \neq 0$$

(i) *Stack* $\qquad\qquad nA^n \qquad (51)$

(ii) *Chain* $\qquad\qquad A^n \qquad (52)$

Introducing a fourth elementary tessellation by connecting horizontally
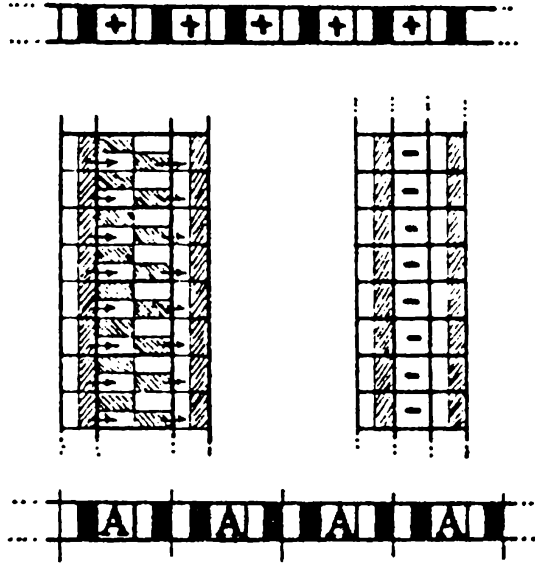
Fig. 8. Some examples of simple tesselations.

$T \rightarrow A \rightarrow T$, or TAT, we have

4. TAT

(i) *Stack* $\qquad n(TA^nT)$ $\qquad$ (53)

(ii) *Chain* $\qquad (TAT)^n$ $\qquad$ (54)

Figure 8 suggests further compositions of elementary tesselations. All of these contain autonomous elements, for it is the presence of at least two such elements as, e.g., in $(TAT)^2$, which constitute a finite function machine. If none of these elements happens to be "dead"—i.e., are locked into a single state static equilibrium—they will by their interaction force each other from one dynamic equilibrium into another one. In other words, under certain circumstances they will turn each other from one trivial finite function machine into another one, but this is exactly the criterion for being a nontrivial finite function machine.

It should be pointed out that this concept of formal mathematical entities interacting with each other is not new. John von Neumann (1966) developed this concept for self-reproducing "automata" which have many properties in common with our tiles. Lars Löfgren (1962) expanded this concept to include self-repair of certain computational elements which are either stationary or freely moving in their tesselations, and Gordon Pask (1962) developed similar ideas for discussing the social self-organization of aggregates of such automata.

It may be noted that in all these studies ensembles of elements are contemplated in order to achieve logical closure in discussing the proprietory concept and autonomous property regarding the elements in question as, e.g., *self*-replication, *self*-repair, *self*-organization, *self*-explanation, etc. This is no accident, as Löfgren (1968) observed, for the prefix "self-" can be replaced by the term to which it is a prefix to generate a second-order concept, a concept of a concept. Self-explanation is the explanation of an explanation; self-organization is the organization of an organization (Selfridge, 1962), etc. Since cognition is essentially a self-referring process (Von Foerster, 1969), it is to be expected that in discussing its underlying mechanisms we have to contemplate function of functions and structure of structures.

Since with the build-up of these structures their functional complexity grows rapidly, a detailed discussion of their properties would go beyond the scope of this article. However, one feature of these computational tesselations can be easily recognized, and this is that their operational modalities are closely linked to their structural organization. Here function and structure go hand in hand, and one should not overlook that perhaps the lion's share of computing has been already achieved when the system's topology is established (Werner, 1969). In organisms this is, of course, done mainly by genetic computations.

This observation leads us directly to the physiology and physics of organic tesselations.

III. BIOPHYSICS

A. General Remarks

The question now arises whether or not one can identify structural or functional units in living organisms which can be interpreted in terms of the purely mathematical objects mentioned previously, the "tiles," the "automata," the "finite function machines," etc. This method of approach, first making an interpretation and then looking for confirming entities, seems to run counter to "the scientific method" in which the "facts" are supposed to precede their interpretation. However, what is reported as "fact" has gone through the observer's cognitive system which provides him, so to say, with a priori interpretations. Since our business here is to identify the mechanisms that observe observers (i.e., becoming "self-observers"), we are justified in postulating first the necessary functional structure of these mechanisms. Moreover, this is indeed a popular approach, as seen by the frequent use of terms like "trace," "engram," "store," "read-in," "read-out," etc., when mechanisms of memory are discussed.

Clearly, here too the metaphor precedes the observations. But metaphors have in common with interpretations the quality of being neither true nor false; they are only useful, useless, or misleading.

When a functional unit is conceptually isolated—an *animal*, a *brain*, the *cerebellum*, *neural nuclei*, a *single neuron*, a *synapse*, a *cell*, the *organelles*, the *genomes*, and other molecular building blocks—in its abstract sense the concept of "machine" applied to these units is useful, if it were only to discipline the user of this concept to identify properly the structural and functional components of his "machine." Indeed, the notions of the finite state machine, or all its methodological relatives, have contributed—explicitly or implicitly—much to the understanding of a large variety of such functional units. For instance, the utility of the concepts "transcript," "en-coding," "de-coding," "computation," etc., in molecular genetics cannot be denied.

Let the n-sequence of the four bases ($b = 4$) of a particular DNA molecule be represented by a ν-number $\nu(n, b)$ [see Eq. (33)]; let $Tr(\nu) = \bar{\nu}$ be an operation which transforms the symbols $(0, 1, 2, 3) \rightarrow (3, 2, 1, \emptyset)$, in that order, with $0 \equiv$ thymine, $1 \equiv$ cytosine, $2 \equiv$ guanine, $3 \equiv$ adenine, and $\emptyset \equiv$ uracil, and I be the identity operation $I(\nu) = \nu$; finally, let $\Phi[\bar{\nu}(n, b)] = \nu(n/3, a) = \mu(m, a)$, with $a = 20$, and $j = 0, 1, \ldots, 19$, representing the 20 amino acids of the polypeptide chain. Then

(i) DNA replication: $\nu = I(\nu)$ (55a)

(ii) DNA/RNA transcript: $\bar{\nu} = Tr(\nu)$ (55b)

(iii) Protein synthesis: $\mu = \Phi(\bar{\nu})$ (55c)

While the operations I and Tr require only trivial machines for the process of transcription, Φ is a recursive computation of the form

$$j(i) \equiv y(i) = y(i - 1) + a^i f(x) \tag{56}$$

Using the suggested recursion [compare with Eq. (14)]:

$$y(i) = a^i f(x) + a^{i-1} f(x^*) + a^{i-2} f(x^{**}) \ldots$$

or

$$y(i) = \sum_{k=0}^{i} a^{i-k} f(x^{(k)*}) \tag{57}$$

and

$$y(m) \equiv \mu(m, a)$$

The function f is, of course, computed by the ribosome which reads the codon x, and synthesizes the amino acids which, in turn, are linked together by the recursion to a connected polypeptide chain.

Visualizing the whole process as the operations of a sequential finite state machine was probably more than just a clue in "breaking the genetic code" and identifying as the input state to this machine the triplet (u, v, w) of adjacent symbols in the ν-number representation of the messenger RNA.

A method for computing ν-numbers of molecular sequences directly from properties of the generated structure was suggested by Pattee (1961). He used the concept of a sequential "shift register," i.e., in principle that of an autonomous tile. For computing periodic sequences in growing helical molecules, the computation for the next element to be attached to the helix is solely determined by the present and some earlier building block. No extraneous computing system is required.

If on a higher level of the hierarchical organization the neuron is taken as a functional unit, the examples are numerous in which it is seen as a recursive function computer. Depending on what is taken to be the "signal," a single pulse, an average frequency code, a latency code, a probability code (Bullock, 1968), etc., the neuron becomes an "all or nothing" device for computing logical functions (McCulloch and Pitts, 1943), a linear element (Sherrington, 1906), a logarithmic element, etc., by changing in essence only a single parameter characteristic for that neuron (Von Foerster, 1967b). The same is true for neural nets in which the recursion is achieved by loops or sometimes directly through recurring fibers. The "reverberating" neural net is a typical example of a finite state machine in its dynamic equilibrium.

In the face of perhaps a whole library filled with recorded instances in which the concept of the finite state machine proved useful, it may come as a surprise that on purely physical grounds these systems are absurd. In order to keep going they must be nothing less than perpetual motion machines. While this is easily accomplished by a mathematical object, it is impossible for an object of reality. Of course, from a heuristical point of view it is irrelevant whether or not a model is physically realizable, as long as it is self-consistent and an intellectual stimulus for further investigations.

However, when the flow of energy between various levels of organization is neglected, and the mechanisms of energy conversion and transfer are ignored, difficulties arise in matching descriptive parameters of functional units on one level to those of higher or lower levels. For instance, a relation between the code of a particular nuclear RNA molecule and, say, the pulse frequency code at the same neuron cannot be established, unless mechanisms of energy transfer are considered. As long as the question as to what keeps the organism going and how this is done is not asked, the gap between functional units on different levels of organization remains open. Can it be closed by thermodynamics?

Three different kinds of molecular mechanisms that offer themselves readily for this purpose will be briefly discussed. All of them make use of various forms of energy as radiation (νh), potential energy (V, structure), work ($p\Delta v$), and heat ($k\Delta T$), and its various conversions from one form to another.

RADIATION $\longrightarrow$ STRUCTURE

STRUCTURE

ENERGY

STRUCTURE

WORK

HEAT

HEAT
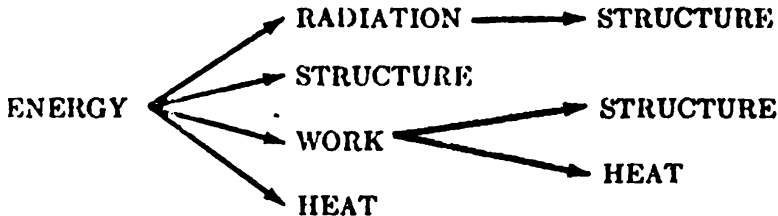
We remain in the terminology of finite state machines and classify the three kinds of mechanism according to their inputs and their outputs, dropping, however, for the moment all distinctions of forms of energy, except that of potential energy (structure) as distinct from all other forms (energy).

(i) Molecular store: Energy in,
Energy out.

(ii) Molecular computer: Energy in,
Structure out.

(iii) Molecular carrier: Energy and structure in,
Energy out.

These three cases will now be briefly reviewed.

B. Molecular Store

Probably the most obvious, and hence perhaps the oldest, approach to link macroscopic behavior, as for example, the forgetting of nonsense syllables (Ebbinghaus, 1885), with the quantum mechanical decay of the available large number of excited metastable states in macromolecules, assumes no further analyzable "elementary impressions" that are associated with a molecule's meta-stable state (Von Foerster, 1948; Von Foerster, 1949). By a nondestructive read-out they can be transferred to another molecule, and a record of these elementary impressions may either decay or else grow, depending on whether the product of the quantum decay time constant with the scanning rate of the read-out is either smaller or else larger than unity. While this model gives good agreement between macroscopic variables such as forgetting rates, temperature dependence of conceived

lapse of time (Hoagland, 1951; Hoagland, 1954), and such microscopic variables as binding energies, electron orbital frequencies, it suffers the malaise of all recording schemes, namely, it is unable to infer anything from the accumulated records. Only if an inductive inference machine which computes the appropriate behavior functions is attached to this record can an organism survive (Von Foerster *et al.*, 1968). Hence, one may abandon speculations about systems that just record specifics, and contemplate those that compute generalizations.

C. Molecular Computer

The good match between macroscopic and microscopic variables of the previous model suggests that this relation should be pursued further. Indeed, it can be shown (Von Foerster, 1969) that the energy intervals between excited meta-stable states are so organized that the decay times in the lattice vibration band correspond to neuronal pulse intervals, and their energy levels to a polarization potential of from 60 mV to 150 mV. Consequently, a pulse train of various pulse intervals will "pump" such a molecule up into higher states of excitation, depending on its initial condition. However, if the excitation level reaches about 1.2 eV, the molecule undergoes configurational changes with life spans of 1 day or longer. In this "structurally charged" state it may now participate in various ways in altering the transfer function of a neuron, either transmitting its energy to other molecules or facilitating their reaction. Since in this model undirected electrical potential energy is used to cause specific structural change, it is referred to as "energy in—structure out." This, however, gives rise to a concept of molecular computation, the result of which is deposition of energy on a specific site of utilization. This is the content of the next and last model.

D. Molecular Carrier

One of the most widely used principles of energy dissemination in a living organism is that of separation of sites of synthesis and utilization. The general method employed in this transfer is a cyclic operation that involves one or many molecular carriers which are "charged" at the site where environmental energy can be absorbed, and are "discharged" where this energy must be used. Charging and discharging is usually accomplished by chemical modifications of the basic carrier molecules. One obvious example of the directional flow of energy and the cyclic flow of matter is, of course, the complementarity of the processes of photosynthesis and respiration (Fig. 9). Light energy, νh, breaks the stable bonds of inorganic oxides and transforms them into energetically charged organic molecules.

ORG.

νh → **Ph.S.**   **Res.** → **k ΔT**
                                      pΔv

INORG.
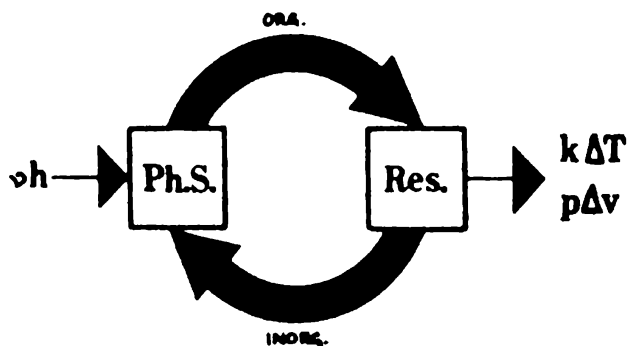
Fig. 9. Directional flow of energy and cyclic flow of matter in photosynthesis coupled with respiration.

These, in turn, are burned up in the respiratory process, releasing the energy in the form of work, $p\Delta v$, or heat, $k\Delta T$, at the site of utilization and return again as inorganic oxides to the site of synthesis.

Another example is the extremely involved way in which in the mitochondria the uphill reaction is accomplished. This reaction not only synthesizes adenosine triphosphate (ATP) by coupling a phosphate group to adenosine diphosphate (ADP), but also charges the ATP molecule with considerable energy which is effectively released during muscular contraction; the contraction process converts ATP back again into ADP by losing the previously attached phosphate group.

Finally, the messenger RNA may be cited as an example of separate sites for synthesis and utilization, although in this case the energetics are as yet not so well established as in the other cases. Here, apparently it is structure which is to be transferred from one place to another, rather than energy.

Common in all these processes is the fact that during synthesis not only a releasable package of energy, ΔE, is put on this molecular carrier but also an address label saying where to deliver the package. This address requires an additional amount of organization, $-\Delta H$, (negentropy), in order to locate its destination. Hence we have the crucial condition

$$\frac{\Delta E}{\Delta H} < 0 \tag{58}$$

which says "for high energy have a low entropy, and for low energy have a high entropy." This is, of course, contrary to the usual course of events in which these two quantities are coupled with each other in a positive relationship.

It can be shown, however, that if a system is composed of constituents which in the ground state are separated, but when "excited" hang together

by "reasonably stable" metastable states, it fulfills the crucial condition above (Von Foerster, 1964).

Let

$$V = \pm\left(Ae^{-x/\kappa} + B \sin\frac{2\pi x}{p}\right) \tag{59}$$

with

$$A/B \gg 1 \quad \text{and} \quad \kappa/p \gg 1$$

be the potential distribution in two one-dimensional linear "periodic crystals," C^+ and C^-, where the $\pm$ refer to corresponding cases. The essential difference between these two linear structures which can be envisioned as linear distributions of electric charges changing their sign (almost) periodically is that energy is required to put "crystal" C^+ together, while for "crystal" C^- about the same energy is required to decompose it into its constituents. These linear lattices have metastable equilibria at

$$C^+ \rightarrow x_1, x_3, x_5 \ldots$$

$$C^- \rightarrow x_0, x_2, x_4 \ldots$$

which are solutions of

$$e^{x/\kappa} \cos\frac{2\pi x}{p} = \frac{1}{2\pi}\frac{Ap}{B\kappa} \approx 1$$

These states are protected by an energy threshold which lets them stay in this state on the average of amount an time

$$\tau = \tau_0 e^{\Delta v/kT} \tag{60}$$

where τ_0^{-1} is an electron orbital frequency, and ΔV is the difference between the energies at the valley and the crest of the potential wall $[\pm\Delta V_n = V(x_n) - V(x_{n+1})]$.

In order to find the entropy of this configuration, we solve the Schrödinger equation (given in normalized form)·

$$\psi'' + \psi[\lambda - V(x)] = 0 \tag{61}$$

for its eigenvalues λ_i and eigenfunctions ψ_i, ψ_i^*, which, in turn, give the probability distribution for the molecule being in the ith eigenstate:

$$\left(\frac{dp}{dx}\right)_i = \psi_i \cdot \psi_i^* \tag{62}$$

with, of course,

$$\int_{-\infty}^{+\infty} \psi_i \cdot \psi_i^* dx = 1 \tag{63}$$

whence we obtain the entropy

$$H_i = -\int_{-\infty}^{+\infty} \psi_i \cdot \psi_i^* \ln \psi_i \cdot \psi_i^* \tag{64}$$

for the ith eigenstate.

It is significant that indeed for the two crystals C^+ and C^- the change in the ratio of energy to entropy for charging ($\Delta E = e(V(x_n) - V(x_{n+2}))$) goes into opposite directions.

$$C^- \rightarrow \left(\frac{\Delta E}{\Delta H}\right)^- > 0$$

$$C^+ \rightarrow \left(\frac{\Delta E}{\Delta H}\right)^+ < 0$$

This shows that the two crystals are quite different animals: one is dead (C^-), the other is alive (C^+).

IV. SUMMARY

In essence this paper is a proposal to restore the original meaning of concepts like memory, learning, behavior, etc. by seeing them as various manifestations of a more inclusive phenomenon, namely, cognition. An attempt is made to justify this proposition and to sketch a conceptual machinery of apparently sufficient richness to describe these phenomena in their proper extension. In its most concise form the proposal was presented as a search for mechanisms within living organisms that enable them to turn their environment into a trivial machine, rather than a search for mechanisms in the environment that turn the organisms into trivial machines.

This posture is justified by realizing that the latter approach—when it succeeds—fails to account for the mechanisms it wishes to discover, for a trivial machine does not exhibit the desired properties; and when it fails does not reveal the properties that made it fail.

Within the conceptual framework of finite state machines, the calculus of recursive functionals was suggested as a descriptive (phenomenological) formalism to account for memory as potential awareness of previous interpretations of experiences, hence for the origin of the *concept of*

"change," and to account for transitions in domains that occur when going from "facts" to "description of facts" and—since these in turn are facts too—to "descriptions of descriptions of facts" and so on.

Elementary finite function machines can be strung together to form linear or two-dimensional tesselations of considerable computational flexibility and complexity. Such tesselations are useful models for aggregates of interacting functional units at various levels in the hierarchical organization of organisms. On the molecular level, for instance, a stringlike tesselation coiled to a helix may compute itself (self-replication) or, in conjunction with other elements, compute other molecular functional units (synthesis).

While in the discussion of descriptive formalisms the concept of recursive functionals provides the bridge for passing through various descriptive domains, it is the concept of energy transfer connected with entropic change that links operationally the functional units on various organizational levels. It is these links, conceptual or operational, which are the prerequisites for interpreting structures and function of a living organism seen as an autonomous self-referring organism. When these links are ignored, the concept of "organism" is void, and its unrelated pieces becomes trivialities or remain mysteries.

ACKNOWLEDGMENT

REFERENCES

Ashby, W. R., 1956, "An Introduction to Cybernetics," Chapman and Hall, London.

Ashby, W. R., 1962, The Set Theory of Mechanisms and Homeostasis, Technical Report 7, NSF Grant 17414, Biological Computer Laboratory, Electrical Engineering Department, University of Illinois, Urbana, 44 pp.

Ashby, W. R., and Walker, C., 1966, On Temporal Characteristics of Behavior in Certain Complex Systems, Kybernetik 3:100.

Bullock, T. H., 1968, Biological Sensors, in "Vistas in Science" (D. L. Arm, ed.) pp. 176-206, University of New Mexico, Albuquerque.

Ebbinghaus, H., 1885, "Über das Gedächtnis: Untersuchungen zur experimentellen Psychologie," Drucker & Humbold, Leipzig.

Eccles, J. C., Ito, M., and Szentagothai, J., 1967, "The Cerebellum as a Neuronal Machine," Springer-Verlag, New York.

Estes, W. K., 1959, The Statistical Approach to Learning Theory, in "Psychology: A Study of a Science, I/2" (S. Koch, ed.) pp. 380-491, McGraw-Hill, New York.

Fitzhugh, H. S. II, 1963, Some considerations of polystable systems, IEEE Transactions 7:1.

Gill, A., 1962, "Introduction to the Theory of Finite State Machines," McGraw-Hill, New York.

Gunther, G., 1967, Time, timeless logic and self-referential systems, in "Interdisciplinary Perspectives of Time" (R. Fischer, ed.) pp. 396-406, New York Academy of Sciences, New York.

Hoagland, H., 1951, Consciousness and the chemistry of time, in "Problems of Consciousness Tr. First Conf." (H.A. Abramson, ed.) pp. 164-198, Josiah Macy Jr. Foundation, New York.

Hoagland, H., 1954, (A remark), in "Problems of Consciousness Tr. Fourth Conf." (H. A. Abramson, ed.) pp. 106-109, Josiah Macy Jr. Foundation, New York.

John, E. R., Shimkochi, M., and Bartlett, F., 1969, Neural readout from memory during generalization, Science 164:1534.

Konorski, J., 1962, The role of central factors in differentiation, in "Information Processing in the Nervous System" (R. W. Gerard and J. W. Duyff, eds.) Vol. 3, pp. 318-329, Excerpta Medica Foundation, Amsterdam.

Lettvin, J. Y., Maturana, H. R., McCulloch, W. S., and Pitts, W., 1959, What the frog's eye tells the frog's brain, Proc. I.R.E. 47:1940.

Löfgren, L., 1962, Kinematic and tesselation models of self-repair, in "Biological Prototypes and Synthetic Systems" (E. E. Bernard and M. R. Kare, eds.) pp. 342-369, Plenum Press, New York.

Löfgren, L., 1968, An axiomatic explanation of complete self-reproduction, Bull. of Math. Biophysics 30(3):415.

Logan, F. A., 1959, The Hull-Spence approach, in "Psychology: A Study of a Science, I/2" (S. Koch, ed.) pp. 293-358, McGraw-Hill, New York.

McCulloch, W. S., and Pitts, W., 1943, A logical calculus of the ideas immanent in nervous activity, Bull. of Math. Biophysics 5:115.

Maturana, H. R., 1969, Neurophysiology of cognition, in "Cognition   A Multiple View" (P. L. Garvin, ed.) in press, Spartan Books, New York.

Maturana, H. R., Uribe, G., and Frenk, S., 1968, A Biological Theory of Relativistic Colour Coding in the Primate Retina, Supplemento No. 1, Arch. Biologia y Med. Exp., University of Chile, Santiago, 30 pp.

Pask, G., 1962, A proposed evolutionary model, in "Principles of Self-Organization" (H. Von Foerster and G. W. Zopf, Jr., eds.) pp. 229-254, Pergamon Press, New York.

Pask, G., 1968, A cybernetic model for some types of learning and mentation, in "Cybernetic Problems in Bionics" (H. L. Oestreicher and D. R. MOore, eds.) pp. 531-586, Gordon & Breach, New York.

Pattee, H. H., 1961, On the origin of macro-molecular sequences, Biophys. J. 1:683.

Pitts, W., and McCulloch, W. S., 1947, How we know universals; the perception of auditory and visual forms, Bull. of Math. Biophysics 9:127.

Selfridge, O. G., 1962, The organization of organization, in "Self-Organizing Systems" (M. C. Yovits, G. T. Jacoby and G. D. Goldstein, eds.) pp. 1-8, Spartan Books, New York.

Sherrington, C. S., 1906, "Integrative Action of the Nervous System," Yale University Press, New Haven.

Skinner, B. F., 1959, A case history in scientific method, in "Psychology: A Study of a Science, I/2" (S. Koch, ed.) pp. 359-379, McGraw-Hill, New York.

Ungar, G., 1969, Chemical transfer of learning, in "The Future of the Brain Sciences" (S. Bogoch, ed.) pp. 373-374, Plenum Press, New York.

Von Foerster, H., 1948, "Das Gedächtnis; Eine quantenmechanische Untersuchung," F. Deuticke, Vienna.

Von Foerster, H., 1949, Quantum mechanical theory of memory, in "Cybernetics, Transactions of the Sixth Conference" (H. Von Foerster, ed.) pp. 112-145, Josiah Macy Jr. Foundation, New York.

Von Foerster, H., 1964, Molecular bionics, in "Information Processing by Living Organisms and Machines" (H. L. Oestreicher, ed.) pp. 161-190, Aerospace Medical Division, Dayton.

Von Foerster, H., 1965, Memory without record, in "The Anatomy of Memory" (D. P. Kimble, ed.) pp. 388-433, Science and Behavior Books, Palo Alto.

Von Foerster, H., 1966, From stimulus to symbol, in "Sign, Image, Symbol" (G. Kepes, ed.) pp. 42-61, George Braziller, New York.

Von Foerster, H., 1967a, Biological principles of information storage and retrieval, in "Electronic Handling of Information: Testing and Evaluation" (A. Kent et al., eds.) pp. 123-147, Academic Press, London.

Von Foerster, H., 1967b, Computation in neural nets, Currents Mod. Biol. 1:47.

Von Foerster, H., 1969, What is memory that it may have hindsight and foresight as well?, in "The Future of the Brain Sciences" (S. Bogoch, ed.) pp. 19-64, Plenum Press, New York.

Von Foerster, H., Inselberg, A., and Weston, P., 1968, Memory and inductive inference, in "Cybernetic Problems in Bionics" (H. L. Oestreicher and D. R. MOore, eds.) pp. 31-68, Gordon & Breach, New York.

von Neumann, J., 1966, "The Theory of Self-Reproducing Automata," (A. Burks, ed.) University of Illinois Press, Urbana.

Walker, C., 1965, A Study of a Family of Complex Systems, An Approach to the Investigation of Organism's Behavior, Technical Report 5, AF-OSR Grant 7-65, Biological Computer Laboratory, Electrical Engineering Department, University of Illinois, Urbana, 251 pp.

Werner, G., 1969, The topology of the body representation in the somatic afferent pathways, in "The Neurosciences, II" Rockefeller University Press, New York.

Weston, P., 1964, Noun chain tress, unpublished manuscript.

PART II
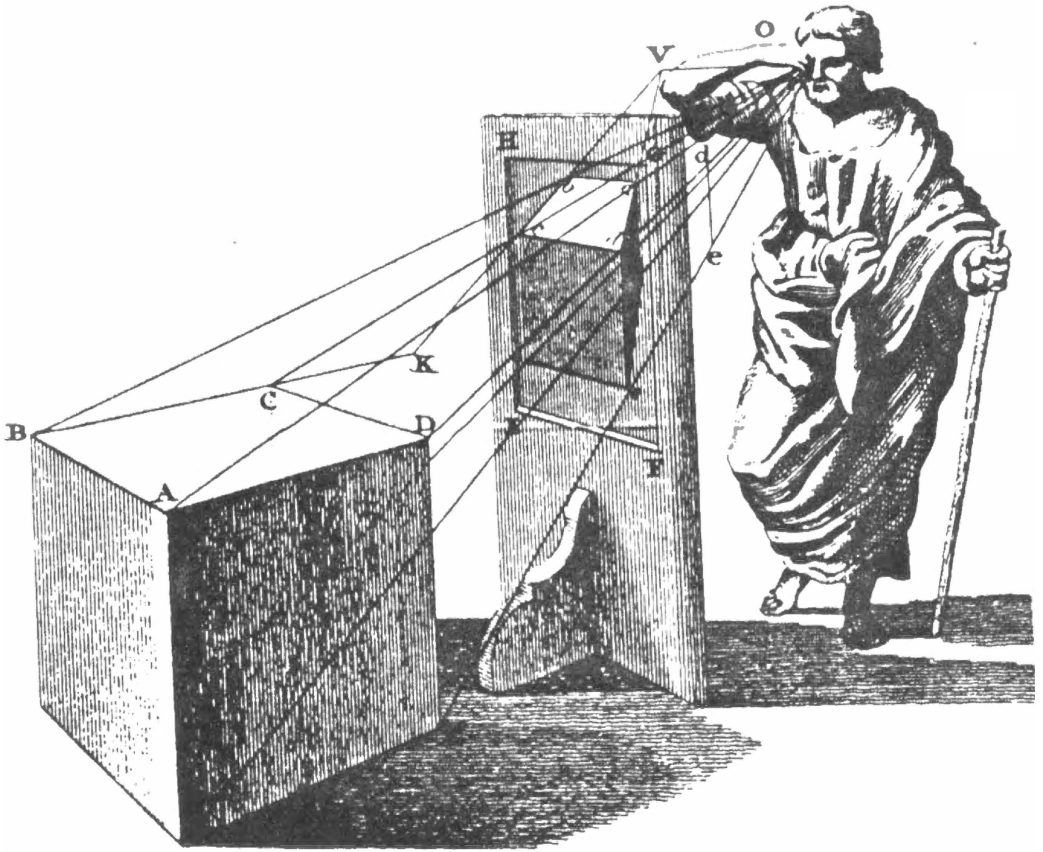
PERCEPTION OF THE FUTURE AND THE
FUTURE OF PERCEPTION*

HEINZ VON FOERSTER

University of Illinois, Urbana, Illinois

ABSTRACT

"The definition of a problem and the action taken to solve it largely depend on the view which the individuals or groups that discovered the problem have of the system to which it refers. A problem may thus find itself defined as a badly interpreted output, or as a faulty output of a faulty output device, or as a faulty output due to a malfunction in an otherwise faultless system, or as a correct but undesired output from a faultless and thus undesirable system. All definitions but the last suggest corrective action; only the last definition suggests change, and so presents an unsolvable problem to anyone opposed to change" (Herbert Brün, 1971).

Truisms have the disadvantage that by dulling the senses they obscure the truth. Almost nobody will become alarmed when told that in times of continuity the future equals the past. Only a few will become aware that from this follows that in times of socio-cultural change the future will *not* be like the past. Moreover, with a future not clearly perceived, we do not know how to act with only one certainty left: if we don't act ourselves, we shall be acted upon. Thus, if we wish to be subjects, rather than objects, what we see now, that is, our perception, must be foresight rather than hindsight.

Epidemic

My colleagues and I are, at present, researching the mysteries of cognition and perception. When, from time to time, we look through the windows of our laboratory into the affairs of this world, we become more and more distressed by what we now observe. The world appears to be in the grip of a fast-spreading disease which, by now, has assumed almost global dimensions. In the individual the symptoms of the disorder manifest themselves by a

---

progressive corruption of his faculty to perceive, with corrupted language being the pathogene, that is, the agent that makes the disease so highly contagious. Worse, in progressive stages of this disorder, the afflicted become numb, they become less and less aware of their affliction.

This state of affairs makes it clear why I am concerned about perception when contemplating the future, for:

if we can't perceive,

we can't perceive of the future

and thus, we don't know how to act now.

I venture to say that one may agree with the conclusion. If one looks around, the world appears like an anthill where its inhabitants have lost all sense of direction. They run aimlessly about, chop each other to pieces, foul their nest, attack their young, spend tremendous energies in building artifices that are either abandoned when completed, or when maintained, cause more disruption than was visible before, and so on. Thus, the conclusions seem to match the facts. Are the premises acceptable? Where does perception come in?

Before we proceed, let me first remove some semantic traps, for—as I said before—corrupt language is the pathogene of the disease. Some simple perversions may come at once to mind, as when "incursion" is used for "invasion," "protective reaction" for "aggression," "food denial" for "poisoning men, beasts, and plants," and others. Fortunately, we have developed some immunity against such insults, having been nourished with syntactic monstrosities as "X is better" without ever saying "than what." There are, however, many more profound semantic confusions, and it is these to which I want to draw your attention now.

There are three pairs of concepts in which one member of these pairs is generally substituted for the other so as to reduce the richness of our conceptions. It has become a matter of fact to confuse process with substance, relations with predicates, and quality with quantity. Let me illustrate this with a few examples out of a potentially very large catalogue, and let me at the same time show you the paralytic behavior that is caused by this conceptual dysfunction.

Process/Substance

The primordial and most proprietary processes in any man and, in fact, in any organism, namely "information" and "knowledge," are now persistently taken as commodities, that is as substance. Information is, of course, the process by which knowledge is acquired, and knowledge is the processes that integrate past and present experiences to form new activities, either as nervous activity internally perceived as thought and will, or externally per-

ceivable as speech and movement (Maturana, 1970, 1971; Von Foerster, 1969, 1971).

Neither of these processes can be "passed on" as we are told in phrases like" . . . Universities are depositories of Knowledge which is passed on from generation to generation . . .," etc., for *your* nervous activity is just *your* nervous activity and, alas, not *mine*.

No wonder that an educational system that confuses the process of creating new processes with the dispensing of goods called "knowledge" may cause some disappointment in the hypothetical receivers, for the goods are just not coming: there are no goods.

Historically, I believe, the confusion by which knowledge is taken as substance comes from a witty broadsheet printed in Nuremberg in the Sixteenth Century. It shows a seated student with a hole on top of his head into which a funnel is inserted. Next to him stands the teacher who pours into this funnel a bucket full of "knowledge," that is, letters of the alphabet, numbers and simple equations. It seems to me that what the wheel did for mankind, the Nuremberg Funnel did for education: we can now roll faster down the hill.

Is there a remedy? Of course, there is one! We only have to perceive lectures, books, slides and films, etc., not as *information* but as *vehicles* for potential information. Then we shall see that in giving lectures, writing books, showing slides and films, etc., we have not solved a problem, we just created one, namely, to find out in which context can these things be seen so that they create in their perceivers new insights, thoughts, and actions.

Relation/Predicate

Confusing relations with predicates has become a political pastime. In the proposition "spinach is green," "green" is a predicate; in "spinach is good," "good" is a relation between the chemistry of spinach and the observer who tastes it. He may refer to his relation with spinach as "good." Our 'mothers, who are the first politicians we encounter, make use of the semantic ambiguity of the syntactic operator "is" by telling us "spinach *is* good" as if they were to say "spinach is *green*."

When we grow older we are flooded with this kind of semantic distortion that could be hilarious if it were not so far reaching. Aristophanes could have written a comedy in which the wisest men of a land set out to accomplish a job that, in principle, cannot be done. They wish to establish, once and for all, all the properties that define an obscene object or act. Of course, "obscenity" is not a property residing within things, but a subject-object relationship, for if we show Mr. X a painting and he calls it obscene, we know a

lot about Mr. X but very little about the painting. Thus, when our lawmakers will finally come up with their imaginary list, we shall know a lot about them, but their laws will be dangerous nonsense.

"Order" is another concept that we are commanded to see in things rather than in our perception of things. Of the two sequences A and B,

A: 1, 2, 3, 4, 5, 6, 7, 8, 9
B: 8, 5, 4, 9, 1, 7, 6, 3, 2

sequence A is seen to be ordered while B appears to be in a mess, until we are told that B has the same beautiful order as A, for B is in alphabetical order (eight, five, four, . . .). "Everything has order once it is understood" says one of my friends, a neurophysiologist, who can see order in what appears to me at first the most impossible scramble of cells. My insistence here to recognize "order" as a subject-object relation and not to confuse it with a property of things may seem too pedantic. However, when it comes to the issue "law and order" this confusion may have lethal consequences. "Law and order" is no issue, it is a desire common to all; the issue is "which laws and what order," or, in other words, the issue is "justice and freedom."

Castration

One may dismiss these confusions as something that can easily be corrected. One may argue that what I just did was doing that. However, I fear this is not so; the roots are deeper than we think. We seem to be brought up in a world seen through descriptions by others rather than through our own perceptions. This has the consequence that instead of using language as a tool with which to express thoughts and experience, we accept language as a tool that determines our thoughts and experience.

It is, of course, very difficult to prove this point, for nothing less is required than to go inside the head and to exhibit the semantic structure that reflects our mode of perception and thinking. However, there are now new and fascinating experiments from which these semantic structures can be inferred. Let me describe one that demonstrates my point most dramatically.

The method proposed by George Miller (1967) consists of asking independently several subjects to classify on the basis of similarity of meaning a number of words printed on cards (Fig. 1). The subject can form as many classes as he wants, and any number of items can be placed in each class. The data so collected can be represented by a "tree" such that the branchpoints further away from the "root" indicate stronger agreement among the sub-

Figure 1. Example of 36 words printed on cards to be classified according to similarity in meaning.

jects, and hence suggest a measure of similarity in the meaning of the words for this particular group of subjects.

Fig. 2 shows the result of such a "cluster analysis" of the 36 words of Fig. 1 by 20 adult subjects ("root" on the left). Clearly, adults classify according to syntactic categories, putting nouns in one class (bottom tree), adjectives in another (next to bottom tree), then verbs, and finally those little words one does not know how to deal with.

The difference is impressive when the adults' results are compared with the richness of perception and imagery of children in the third and fourth grade when given the same task (Fig. 3). Miller reflects upon these delightful results:

"Children tend to put together words that might be used in talking about the same thing—which cuts right across the tidy syntactic boundaries so important to adults. Thus all twenty of the children agree in putting the verb 'eat' with the noun 'apple'; for many of them 'air' is 'cold'; the 'foot' is used to 'jump'; you 'live' in a 'house'; 'sugar' is 'sweet'; and the cluster of 'doctor,' 'needle,' 'suffer,' 'weep,' and 'sadly' is a small vignette in itself."

What is wrong with our education that castrates our power over language? Of the many factors that may be responsible I shall name only one that has a profound influence on our way of thinking, namely, the misapplication of the "scientific method."

ADULTS
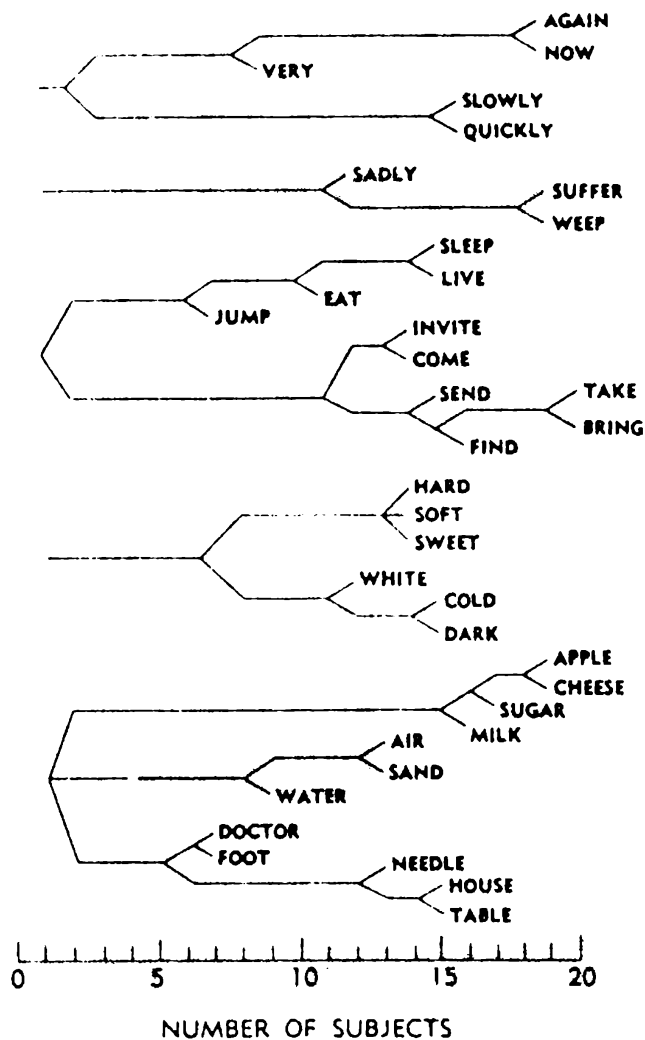
Figure 2. Cluster analysis of the 36 words of Fig. 1 classified by 20 adult subjects. Note that syntactic categories are faithfully respected, while semantic relations are almost completely ignored.

Scientific Method

The scientific method rests on two fundamental pillars:

(i) Rules observed in the past shall apply to the future. This is usually referred to as the principle of conservation of rules, and I have no doubt that you are all familiar with it. The other pillar, however, stands in the shadow of the first and thus is not so clearly visible:

(ii) Almost everything in the universe shall be irrelevant. This is usually referred to as the principle of the necessary and sufficient cause, and what it

GRADES 3 AND 4
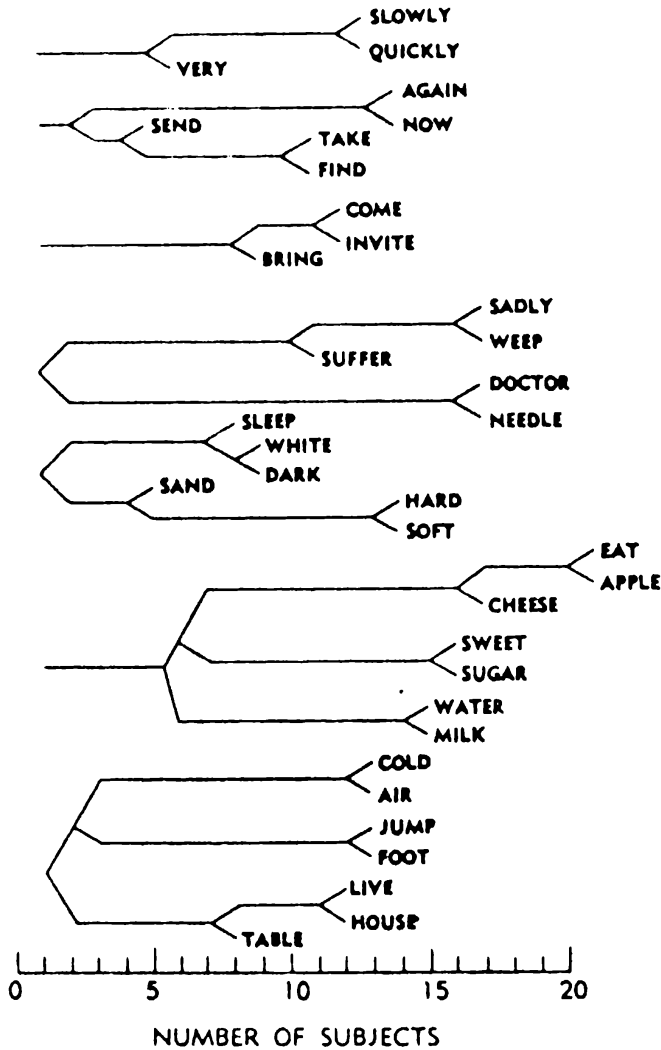
Figure 3. The same 36 words of Figs. 1 and 2 classified by children in the third and fourth grade. Note the emergence of meaningful cognitive units, while syntactic categories are almost completely ignored.

demands is at once apparent when one realizes that "relevance" is a triadic relation that relates a set of propositions $(P_1, P_2, \ldots)$ to another set of propositions $(Q_1, Q_2, \ldots)$ in the mind (M) of one who wishes to establish this relation. If P are the causes that are to explain the perceived effects Q, then the principle of necessary and sufficient cause forces us to reduce our perception of effects further and further until we have hit upon the necessary and sufficient cause that produces the desired effect: everything else in the universe shall be irrelevant.

It is easy to show that resting one's cognitive functions upon these two pillars is counter-productive in contemplating any evolutionary process, be it the growing up of an individual, or a society in transition. In fact, this was already known by Aristotle who distinguished two kinds of cause, one the "efficient cause," the other the "final cause," which provide us with two distinct explanatory frameworks for either inanimate matter, or else living organisms, the distinction being that the efficient cause *precedes* its effect while the final cause *succeeds* its effect. When striking with a match the treated surface of a matchbook, the striking is the (efficient) cause for the match to ignite. However, the cause for my striking the match is my wish to have it ignited (final cause).

Perhaps, with this distinction, my introductory remarks may appear much clearer. Of course, I had in mind the final cause when I said that if we can perceive of the future (the match being ignited), we know how to act now (strike!). This leads me immediately to draw a conclusion, namely:

At any moment we are free to act toward the future we desire.

In other words, the future will be as we wish and perceive it to be. This may come as a shock only to those who let their thinking be governed by the principle that demands that only the rules observed in the past shall apply to the future. For those the concept of "change" is inconceivable, for change is the process that obliterates the rules of the past.

Quality/Quantity ·

In order to protect society from the dangerous consequences of change, not only a whole branch of business has emerged, but also the Government has established several offices that busy themselves in predicting the future by applying the rules of the past. These are the Futurists. Their job is to confuse quality with quantity, and their products are "future scenarios" in which the qualities remain the same, only the quantities change: more cars, wider highways, faster planes, bigger bombs, etc. While these "future scenarios" are meaningless in a changing world, they have become a lucrative business for entrepreneurs who sell them to corporations that profit from designing for obsolescence.

With the diagnosis of the deficiency to perceive qualitative change, that is, a change of our subject-object and subject-subject relationships, we are very close to the root of the epidemic that I mentioned in my opening remarks. An example in neurophysiology may help to comprehend the deficiency that now occurs on the cognitive level.

Dysgnosis

The visual receptors in the retina, the cones and the rods, operate optimally only under certain conditions of illumination. Beyond or below this condition we suffer a loss in acuity or in color discrimination. However, in the vertebrate eye the retina almost always operates under these optimal conditions, because of the iris that contracts or dilates so as to admit under changing conditions of brightness the same amount of light to the receptors. Hence, the scenario "seen" by the optic nerve has always the same illumination independent of whether we are in bright sunshine or in a shaded room. How, then, do we know whether it is bright or shady?

The information about this datum resides in the regulator that compares the activity in the optic nerve with the desired standard and causes the iris to contract when the activity is too high, and to dilate when it is too small. Thus, the information of brightness does not come from inspecting the scenario—it appears always to be of similar brightness—it comes from an inspection of the regulator that suppresses the perception of change.

There are subjects who have difficulties in assessing the state of their regulator, and thus they are weak in discriminating different levels of brightness. They are called "dysphotic." They are the opposite of photographers who may be called "photic," for they have a keen sense of brightness discrimination. There are subjects who have difficulties in assessing the regulators that maintain their identity in a changing world. I shall call individuals suffering from this disorder "dysgnostic," for they have no way of knowing themselves. Since this disorder has assumed extraordinary dimensions, it has indeed been recognized at the highest national level.

As you all know, it has been observed that the majority of the American people cannot speak. This is interpreted by saying that they are "silent"; I say they are *mute*. However, as you all know very well, there is nothing wrong with the vocal tract of those who are mute: the cause of their muteness is deafness. Hence, the so-called "silent majority" is *de facto* a "deaf majority."

However, the most distressing thing in this observation is that there is again nothing wrong with their auditory system; they could hear if they wanted to: but they don't want to. Their deafness is voluntary, and in others it is their blindness.

At this point proof will be required for these outrageous propositions. *TIME Magazine* (1970) provides it for me in its study of Middle America.

There is the wife of a Glencoe, Illinois lawyer, who worries about the America in which her four children are growing up: "I want my children to live and grow up in an America as I *knew* it," [note the principle of conservation of rule where the future equals the past] "where we were proud to be

citizens of this country. I'm damned sick and tired of *listening* to all this nonsense about how awful America is." [Note voluntary deafness.]

Another example is a newspaper librarian in Pittsfield, Massachusetts, who is angered by student unrest: "Every time I see protestors, I say, 'Look at those creeps.'" [Note reduction of visual acuity.] "But then my 12-year-old son says, 'They're not creeps. They have a perfect right to do what they want.'" [Note the un-adult-erated perceptual faculty in the young.]

The tragedy in these examples is that the victims of "dysgnosis" not only do not know that they don't see, hear, or feel, they also do not want to.

How can we rectify this situation?

Trivialization

I have listed so far several instances of perceptual disorders that block our vision of the future. These symptoms collectively consitute the syndrome of our epidemic disease. It would be the sign of a poor physician if he were to go about relieving the patient of these symptoms one by one, for the elimination of one may aggrevate another. Is there a single common denominator that would identify the root of the entire syndrome?

To this end, let me introduce two concepts, they are the concepts of the "trivial" and the "non-trivial" machine. The term "machine" in this context refers to well-defined functional properties of an abstract entity rather than to an assembly of cogwheels, buttons and levers, although such assemblies may represent embodiments of these abstract functional entities.

A trivial machine is characterized by a one-to-one relationship between its "input" (stimulus, cause) and its "output" (response, effect). This invariable relationship is "the machine." Since this relationship is determined once and for all, this is a deterministic system; and since an output once observed for a given input will be the same for the same input given later, this is also a predictable system.

Non-trivial machines, however, are quite different creatures. Their input-output relationship is not invariant, but is determined by the machine's previous output. In other words, its previous steps determine its present reactions. While these machines are again deterministic systems, for all practical reasons they are unpredictable: an output once observed for a given input will most likely be not the same for the same input given later.

In order to grasp the profound difference between these two kinds of machines it may be helpful to envision "internal states" in these machines. While in the trivial machine only one internal state participates always in its internal operation, in the non-trivial machine it is the shift from one internal state to another that makes it so elusive.

One may interpret this distinction as the Twentieth Century version of Aristotle's distinction of explanatory frameworks for inanimate matter and living organisms.

All machines we construct and buy are, hopefully, trivial machines. A toaster should toast, a washing machine wash, a motorcar should predictably respond to its driver's operations. In fact, all our efforts go into one direction, to create trivial machines or, if we encounter non-trivial machines, to convert them into trivial machines. The discovery of agriculture is the discovery that some aspects of Nature can be trivialized: If I till today, I shall have bread tomorrow.

Granted, that in some instances we may be not completely successful in producing ideally trivial machines. For example, one morning turning the starter key to our car, the beast does not start. Apparently it changed its internal state, obscure to us, as a consequence of previous outputs (it may have exhausted its gasoline supply) and revealed for a moment its true nature of being a non-trivial machine. But this is, of course, outrageous and this state of affairs should be remedied at once.

While our pre-occupation with the trivialization of our environment may be in one domain useful and constructive, in another domain it is useless and destructive. Trivialization is a dangerous panacea when man applies it to himself.

Consider, for instance, the way our system of education is set up. The student enters school as an unpredictable "non-trivial machine." We don't know what answer he will give to a question. However, should he succeed in this system the answers he gives to our questions must be known. They are the "right" answers:

Q: "When was Napoleon born?"
A: "1769"
Right!
Student → Student
but
Q: "When was Napoleon born?"
A: Seven years before the Declaration of Independence."
Wrong!
Student → Non-student

Tests are devices to establish a measure of trivialization. A perfect score in a test is indicative of perfect trivialization: the student is completely predictable and thus can be admitted into society. He will cause neither any surprises nor any trouble.

Future

I shall call a question to which the answer is known an "illegitimate question." Wouldn't it be fascinating to contemplate an educational system that would ask of its students to answer "legitimate questions" that is questions to which the answers are unknown (H. Brün in a personal communication). Would it not be even more fascinating to conceive of a society that would establish such an educational system? The necessary condition for such an utopia is that its members perceive one another as autonomous, non-trivial beings. Such a society shall make, I predict, some of the most astounding discoveries. Just for the record, I shall list the following three:

1. "Education is neither a right nor a privilege: it is a necessity."
2. "Education is learning to ask legitimate questions."

A society who has made these two discoveries will ultimately be able to discover the third and most utopian one:

3. "A is better off when B is better off."

From where we stand now, anyone who seriously makes just one of those three propositions is bound to get into trouble. Maybe you remember the story Ivan Karamazov makes up in order to intellectually needle his younger brother Alyosha. The story is that of the Great Inquisitor. As you recall, the Great Inquisitor walks on a very pleasant afternoon through his town, I believe it is Salamanca; he is in good spirits. In the morning he has burned at the stakes about a hundred and twenty heretics, he has done a good job, everything is fine. Suddenly there is a crowd of people in front of him, he moves closer to see what's going on, and he sees a stranger who is putting his hand onto a lame person, and that lame one can walk. Then a blind girl is brought before him, the stranger is putting his hand on her eyes, and she can see. The Great Inquisitor knows immediately who He is, and he says to his henchmen: "Arrest this man." They jump and arrest this man and put Him into jail. In the night the Great Inquisitor visits the stranger in his cell and he says: "Look, I know who You are, troublemaker. It took us one thousand and five hundred years to straighten out the troubles you have sown. You know very well that people can't make decisions by themselves. You know very well people can't be free. *We* have to make their decisions. *We* tell them who they are to be. You know that very well. Therefore, I shall burn You at the stakes tomorrow." The stranger stands up, embraces the Great Inquisitor and kisses him. The Great Inquisitor walks out, but, as he leaves the cell, he does not close the door, and the stranger disappears in the darkness of the night.

Let us remember this story when we meet those troublemakers, and let

us keep the door open for them. We shall recognize them by an act of creation:

"Let there be vision: and there was light."

References

Brun, H. (1971). "Technology and the Composer," in Von Foerster, H., ed., *Interpersonal Relational Networks.* pp. 1/10. Cuernavaca: Centro Intercultural de Documentacion.

Maturana, H. R. (1970). "Biology of Cognition" BCL Report No. 9.0, Biological Laboratory, Department of Electrical Engineering, University of Illinois, Urbana, 93 pp.

Maturana, H. R. (1971). "Neurophysiology of Cognition," in Garvin, P., ed., *Cognition, A Multiple View,* pp. 3-23. New York: Spartan Books.

Miller, G. A. (1967). "Psycholinguistic Approaches to the Study of Communication," in Arm, D. L., ed., *Journeys in Science,* pp. 22-73. Albuquerque: Univ. New Mexico.

TIME Magazine. (1970). "The Middle Americans," (January 5).

Von Foerster, H. (1969). "What is Memory that It May Have Hindsight and Foresight as well?," in Bogoch, S., ed., *The Future of the Brain Sciences,* pp. 19-64. New York: Plenum Press.

Von Foerster, H. (1971). "Thoughts and Notes on Cognition," in Garvin, P., ed., *Cognition, A Multiple View,* pp. 25-48. New York: Spartan Books.

Aqua ☿ Phi:

Phi Gede

Phi Borb

Phi Corpus
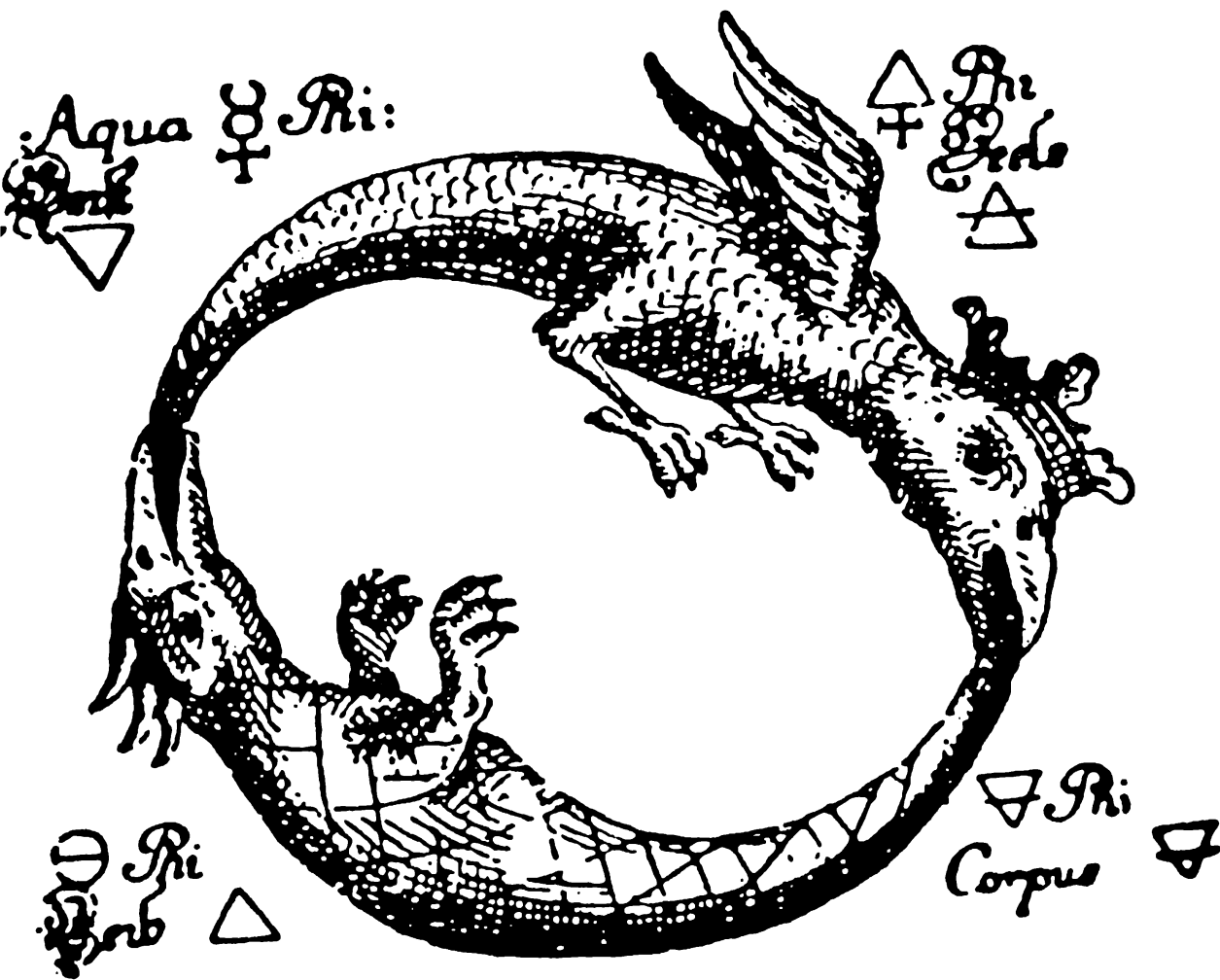
Responsibilities of Competence*

Heinz Von Foerster
Biological Computer Laboratory
University of Illinois

At our last Annual Symposium I submitted to you a theorem to which Stafford Beer referred on another occasion as "Heinz Von Foerster's Theorem Number One." As some of you may remember, it went as follows:

> "The more profound the problem that is ignored, the greater are the chances for fame and success."

Building on a tradition of a single instance, I shall again submit a theorem which, in all modesty, I shall call "Heinz Von Foerster's Theorem Number Two." It goes as follows:

> "The hard sciences are successful because they deal with the soft problems; the soft sciences are struggling because they deal with the hard problems."

Should you care to look closer, you may discover that Theorem 2 could serve as a corollary to Theorem 1. This will become obvious when we contemplate for a moment the method of inquiry employed by the hard sciences. If a system is too complex to be understood it is broken up into smaller pieces. If they, in turn, are still too complex, they are broken up into even smaller pieces, and so on, until the pieces are so small that at least one piece can be understood. The delightful feature of this process, the method of reduction, "reductionism," is that it inevitably leads to success.

Unfortunately, the soft sciences are not blessed with such favorable conditions. Consider, for instance, the sociologist, psychologist, anthropologist, linguist, etc. If they would reduce the complexity of the system of their interest, i.e., society, psyche, culture, language, etc., by breaking it up into smaller parts for further inspection they would soon no longer be able to claim that they are dealing with the original system of their choice. This is so, because these scientists are dealing with essentially nonlinear systems whose salient features are represented by the *interactions* between whatever one may call their "parts" whose properties in isolation add little, if anything, to the understanding of the workings of these systems when each is taken as a whole. Consequently, if he wishes to remain in the field of his choice the scientist who works in the soft sciences is faced with a formidable problem: he cannot afford to loose sight of the full complexity of his system, on the other hand it becomes more and more urgent that his problems be solved. This is not just to please him. By now it has become quite clear that his problems concern us all. "Corruption of our society," "psychological disturbances," "cultural erosion," the "breakdown of communication," and all the other of these "crises" of today are our problems as well as his. How can we contribute to their solution?

My suggestion is that we apply the *competences* gained in the hard sciences — and not the method of reduction — to the solution of the hard problems in the soft sciences. I hasten to add that this suggestion is not new at all. In fact, I submit that it is precisely *Cybernetics* that interfaces hard competence with the hard problems of the soft sciences. Those of us who witnessed the early development of cybernetics may well remember that before Norbert Wiener created that name of our science it was referred to as the study of "Circular-Causal and Feedback Mechanisms in Biological and Social Systems," a description it carried even years after he wrote his famous book. Of course, in his definition of Cybernetics as the science of "communication and control in the animal and the machine" Norbert Wiener went one step further in the generalization of these concepts, and today "Cybernetics" has ultimately come to stand for the science of *regulation* in the most general sense.

Since our science embraces indeed this general and all-pervasive notion, why then, unlike most of our sister sciences, do we not have a patron saint or a deity to bestow favors on us in our search for new insights, and who protects our society from evils from without as well as from within? Astronomers and physicists are looked after by Urania; Demeter patronizes agriculture; and various Muses help the various arts and sciences. But who helps Cybernetics?

One night when I was pondering this cosmic question I suddenly had an apparition. Alas, it was not one of the charming goddesses who bless the other arts and sciences. Clearly, that funny little creature sitting on my desk must be a demon. After a while he started to talk. I was right. "I am Maxwell's Demon," he said. And then he disappeared.

When I regained my composure it was immediately clear to me that nobody else but this respectable demon could be our patron, for Maxwell's Demon is *the paradigm for regulation*.

As you remember, Maxwell's Demon regulates the flow of molecules between two containers in a most *unnatural* way, namely, so that heat flows from the cold container to the hotter, as opposed to the natural course of events where without the demon's interference heat always flows from the hot container to the colder.

I am sure you also remember how he proceeds: He guards a small aperture between the two containers which he opens to let a molecule pass whenever a fast one comes from the cool side or a slow one comes from the hot side. Otherwise he keeps the aperture closed. Obviously, by this maneuver he gets the cool container becoming cooler, and the hot container getting hotter, thus apparently upsetting the Second Law of Thermodynamics. Of course, we know by now that while he succeeds in obtaining this perverse flow of heat, the Second Law remains untouched. This is because of his need for a flashlight to determine the velocity of the upcoming molecules. Were he at thermal equilibrium with one of the containers he couldn't see a thing: he is part of a black body. Since he can do his antics only as long as the battery of his flashlight lasts, we must include into the system with an active demon not only the energy of the two containers, but also that of the battery. The entropy gained by the battery's decay is not completely compensated by the negentropy gained from the increased disparity of the two containers.

The moral of this story is simply that while our demon cannot beat the Second Law, he can, by his regulatory activity, retard the degradation of the available energy, i.e., the growth of entropy, to an arbitrary slow rate.

This is indeed a very significant observation because it demonstrates the paramount importance of regulatory mechanisms in living organisms. In this context they can be seen as manifestations of Maxwell's Demon, retarding continuously the degradation of the flow of energy, that is, retarding the increase of entropy. In other words, as regulators living organisms are "entropy retarders."

Moreover, as I will show in a moment, Maxwell's Demon is not only an entropy retarder an
a paradigm for regulation, but he is also a functional isomorph of a Universal Turing Machine. T
the three concepts of regulation, entropy retardation, and computation constitute an interlaced
conceptual network which, for me, is indeed the essence of Cybernetics.

I shall now briefly justify my claim that Maxwell's Demon is not only the paradigm for reg
lation but also for computation.

When I use the term "computation" I am not restricting my self to specific operations as, f
instance, addition, multiplication, etc. I wish to interpret "computation" in the most general ser
as a mechanism, or "algorithm," for *ordering*. The ideal, or should I say the most general, repres
tation of such mechanism is, of course, a Turing Machine, and I shall use this machine to illumin
some of the points I wish to make.

There are two levels on which we can think of "ordering." The one is when we wish to mak
a description of a given arrangement of things. The other one when we wish to re-arrange things
cording to certain descriptions. It will be obvious at once that these two operations constitute in
deed for foundations for all that which we call "computation."

Let A be a particular arrangement. Then this arrangement can be computed by a universal
Turing machine with a suitable initial tape expression which we shall call a "description" of A:
D(A). The length L(A) of this description will depend on the alphabet (language) used. Hence, w
may say that a language a_1 reveals more order in the arrangement A than another language a_2,
if and only if the length $L_1(A)$ of the suitable initial tape description for computing A is shorter
than $L_2(A)$, or *mutatis mutandis*.

This covers the first level of above, and leads us immediately to the second level.

Among all suitable initial tape descriptions for an arrangement A_1 there is a shortest one:
$L^*(A_1)$. If A_1 is re-arranged to give A_2, call A_2 to be of a higher order than A_1 if and only if the
shortest initial tape description $L^*(A_2)$ is shorter than $L^*(A_1)$, or *mutatis mutandis*.

This covers the second level of above, and leads us to a final statement of perfect ordering
(computation).

Among all arrangements A_i there is one, A^*, for which the suitable initial tape description is
the shortest $L^*(A^*)$.

I hope that with these examples it has become clear that living organisms (replacing now the
Turing machine) interacting with their environment (arrangements) have several options at their d
posal: (i) they may develop "languages" (sensors, neural codes, motor organs, etc.) which "fit"
their given environment better (reveal more order); (ii) they may change their surroundings until
it "fits" their constitution; and (iii), they may do both. However, it should be noted that whateve
option they take, it will be done by computation. That these computations are indeed functional
isomorphs of our demon's activity is now for me to show.

The essential function of a Turing machine can be specified by five operations:

(i)   *Read* the input symbol x.
(ii) *Compare* x with z, the internal state of the machine.
(iii) *Write* the appropriate output symbol y.
(iv) *Change* the internal state z to the new state z'.
(v)   *Repeat* the above sequence with a new input state x'.

Similarly, the essential function of Maxwell's Demon can be specified by five operations equivalent to those above:

(i)   *Read* the velocity v of the upcoming molecule M.

(ii) *Compare* $(mv^2/2)$ with the mean energy $<mv^2/2>$ (temperature T) of, say, the cooler container (internal state T).

(iii) *Open* the aperture if $(mv^2/2)$ is greater than $<mv^2/2>$; otherwise keep it closed.

(iv) *Change* the internal state T to the new (cooler) state T'.

(v)   *Repeat* the above sequence with a new upcoming molecule M'.

Since the translation of the terms occurring in the correspondingly labeled points is obvious, with the presentation of these two lists I have completed my proof.

How can we make use of our insight that Cybernetics is the science of regulation, computation, ordering, and entropy retardation? We may, of course, apply our insight to the system that is generally understood to be the *cause célèbre* for regulation, computation, ordering, and entropy retardation, namely, the human brain.

Rather than following the physicists who order their problems according to the number of *objects* involved ("The one-body problem," "The two-body problem," "The three-body problem," etc.), I shall order our problems according to the number of *brains* involved by discussing now "The one-brain problem," "The two-brain problem," "The many-brain problem," and "The all-brain problem."

1.   The Single-Brain Problem: The Brain Sciences

It is clear that if the brain sciences do not want to degenerate into a physics or chemistry of living – or having once lived – tissue they must develop a theory of the brain: T(B). But, of course, this theory must be written by a brain: B(T). This means that this theory must be constructed so as to write itself T(B(T)).

Such a theory will be distinct in a fundamental sense from, say, physics which addresses itself to a (not quite) successful description of a "subjectless world" in which even the observer is not supposed to have a place. This leads me now to pronounce my Theorem Number Three:

"The Laws of Nature are written by man. The laws of biology must write themselves."

In order to refute this theorem it is tempting to invoke Gödel's Proof of the limits of the Entscheidungsproblem in systems that attempt to speak of themselves. But Lars Löfgren and Gotthard Günther have shown that self-explanation and self-reference are concepts that are untouched by Gödel's arguments. In other words, a science of the brain in the above sense is, I claim, indeed a legitimate science with a legitimate problem.

2.   The Two-Brain Problem: Education

It is clear that the majority of our established educational efforts is directed toward the trivialization of our children. I use the term "trivialization" exactly as used in automata theory, where a trivial machine is characterized by its fixed input-output relation, while in a non-trivial machine (Turing machine) the output is determined by the input *and* its internal state. Since our educational system is geared to generate predictable citizens, its aim is to amputate the bothersome internal states which generate unpredictability and novelty. This is most clearly demonstrated by our method of examination in which only questions are asked for which the answers are known (or defined), and are to be memorized by the student. I shall call these questions "illegitimate questions."

Would it not be fascinating to think of an educational system that de-trivializes its studen
by teaching them to ask "legitimate questions," that is, questions for which the answers are u
known?

3. The Many-Brain Problem: Society

It is clear that our entire society suffers from a severe dysfunction. On the level of the in
vidual this is painfully felt by apathy, distrust, violence, disconnectedness, powerlessness, alien
ation, and so on. I call this the "participatory crisis," for it excludes the individual from partic
pating in the social process. The society becomes the "system," the "establishment" or what h
you, a depersonalized Kafkaesque ogre of its own ill will.

It is not difficult to see that the essential cause for this dysfunction is the absence of an a
quate input for the individual to interact with society. The so-called "communication channels
the "mass media" are only one-way: they talk, but nobody can talk back. The feedback loop i
missing and, hence, the system is out of control. What cybernetics could supply is, of course, a
versally accessible social input device.

4. The All-Brain Problem: Humanity

It is clear that the single most distressing characteristic of the global system "mankind" is
demonstrated instability, and a fast approaching singularity. As long as humanity treats itself as
open system by ignoring the signals of its sensors that report about its own state of affairs, we s
approach this singularity with no breaks whatsoever. (Lately I began to wonder whether the inf
mation of its own state can reach all elements in time to act should they decide to listen rather t
fight.)

The goal is clear: we have to close the system to reach a stable population, a stable econom
and stable resources. While the problem of constructing a "population servo" and an "economic
servo" can be solved with the mental resources on this planet, for the stability of our material re
sources we are forced by the Second Law of Thermodynamics to turn to extra-planetary sources
About $2 \cdot 10^{14}$ kilowatts solar radiation are at our disposal. Wisely used, this could leave our ear
ly, highly structured, invaluable organic resources, fossilized or living, intact for the use and enjo
ment of uncounted generations to come.

If we are after fame and success we may ignore the profundity of these problems in compu-
tation, ordering, regulation, and entropy retardation. However, since we as cyberneticians sup-
posedly have the competence to attack them, we may set our goal above fame and success by qui
ly going about their solution. If we wish to maintain our scientific credibility, the first step to ta
is to apply our competence to ourselves by forming a global society which is not so much *for*
Cybernetics as it *functions* cybernetically. This is how I understand Dennis Gabor's exhortation i
an earlier issue: "Cyberneticians of the world, unite!" Without communication there is no regula-
tion; without regulation there is no goal; and without a goal the concept of "society" or "system
becomes void.

Competence implies responsibilities. A doctor must act at the scene of the accident. We can
no longer afford to be the knowing spectators at a global disaster. We must share what competen
we have through communication and cooperation in working together through the problems of o
time. This is the only way in which we can fulfill our social and individual responsibilities as cybe
neticians who should practice what they preach.
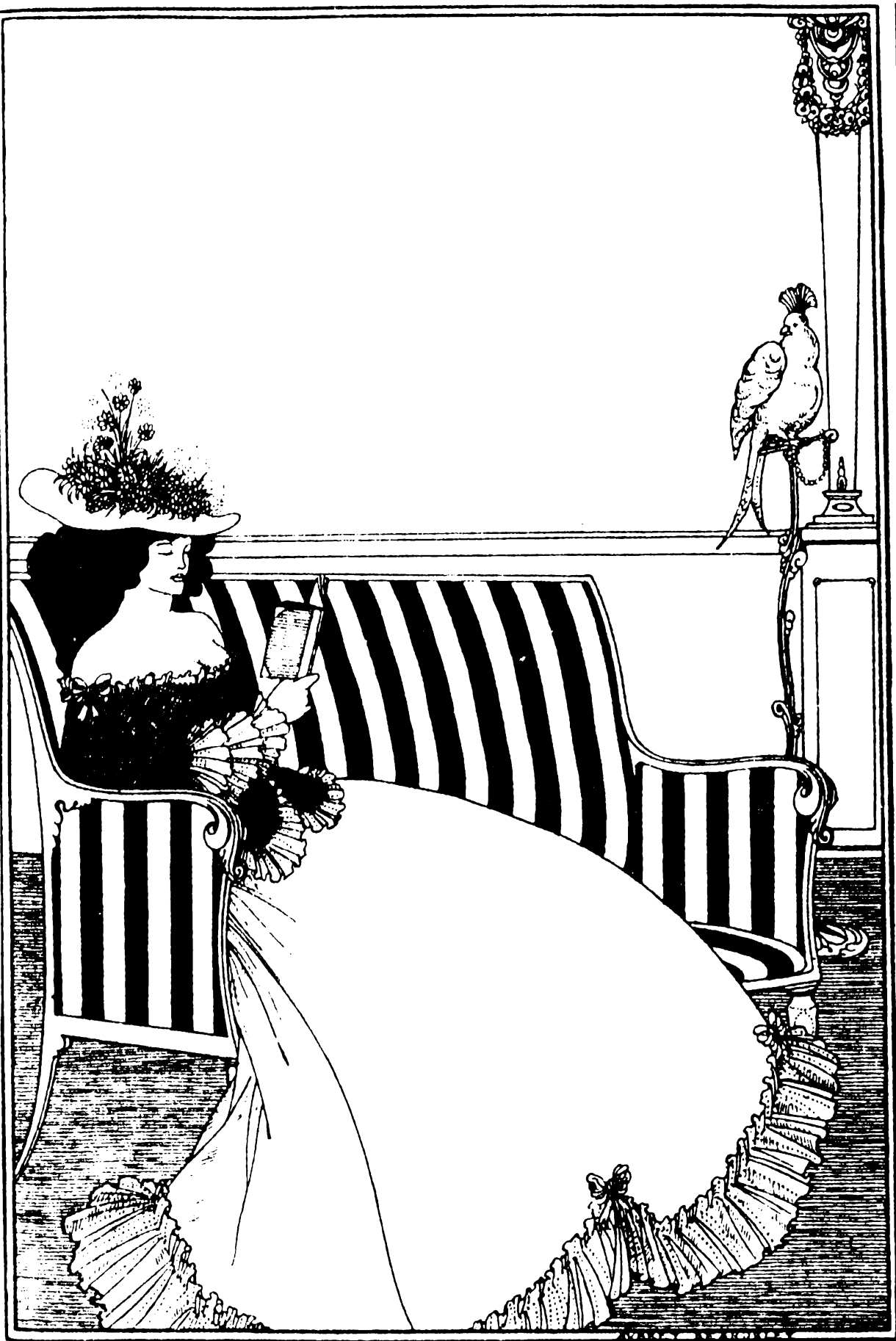
Technology:
What Will It Mean to Librarians?

[A Response]

Heinz Von Foerster

QUESTION

When a few days ago Mrs. Howe informed me of your motives for inviting me to this institute, I was very much intrigued by her persistent refusal to give me a title for my lecture. Instead, she asked me to respond to a personal question that seemed to be on most of your minds, namely, what impact technology will have on the future of your professional lives. I shall attempt to answer you as good as one who loves books and persistently asks questions.

First of all let me assure you that I cannot think of a more timely moment than this that you wish to see a climate in which your pressing professional needs may fruitfully interact with potential technologies, for I feel you should not wait and see what technology will do with you—pardon me, technologists will say "*for* you"—but rather you should tell technologists what you want to be done.

In other words, I suggest to invert Mrs. Howe's original question: "Technology: What will it mean to Librarians?" to ask now "Librarians: What will they mean to Technology?"

The answer to this question will, of course, depend on your knowledge of what can be done. However, in a more

important sense it will depend on how you, as librarians, see yourselves as part of your culture in the future, and what part you wish technology to play in your projected self-image.

TECHNOLOGY

In contrast to the general belief that we live today in an age of technology, I maintain that we are living now in an age of technocracy. By this I mean that we have, hopefully only temporarily, relinquished our responsibility to ask for a technology that will solve existent problems. Instead we have allowed existent technology to create problems it can solve. For instance, we are told that those new ceramic pots that go from freeze to flame are a spin-off from the Apollo moon program. I prefer to perceive of a world in which the Apollo moon program may be a spin-off from making such functional pans and pots.

Similarly, I can perceive of a world in which an adequate technology is created to meet the basic problems of your task, rather than one in which the solutions to some superficial problems in library science are a spin-off from the industrial computer development program. How do I see these basic problems?

LIBRARIANS

There are two functions in which I can see the librarian to serve in the social fab-

ric of the future, one as being a custodian of books, the other one as being a midwife for those who wish to give birth to new insights and ideas.

This alternative appears to make a superficial distinction, unless one realizes that approximately two million research papers in all fields of science and technology taken together are published annually, and that these papers appear distributed over some 30,000 different specialized journals. Within the present century these numbers double every ten to fifteen years, growing at rates that are more than twice the global rate of human population growth (1). This suggests that the chances for the potential user of a future library to name successfully the source for his enlightenment are diminishing at a formidable rate. Consequently, he will first shift his request for a particular book or document to a request for titles of appropriate books or documents by asking: "*Where* is the answer to my question?" Since, however, a user is primarily interested in getting an answer for his question, and only secondarily where he can find it, he will ultimately ask: "*What* is the answer to my question?"

I know that you are very well aware of the present shift from requests for documents to requests for titles, and that you are meeting these demands by rapidly developing extensive indexing languages, cross-reference file structures, and sophisticated abstracting procedures. On the other hand, I am also convinced that you are aware of the shortcomings of these strategies and of their limitations that are conceptual rather than technological. For instance, it may amuse you to know that with all these highly elaborate search techniques we shall never discover

that Max Planck created quantum mechanics: neither in the title nor in the text of his revolutionary papers[2,3] appears the term "quantum." Or one may become justifiably discouraged to establish a cross-reference matrix. Aiming at a k-th order cross-listing of D documents the matrix has D^K entries, which means adding annually to a catalogue of only second order cross-listings no less than 4 trillion entries, almost all of which will be void—but must be listed as void!

Although present day technologists are eagerly persuading you to invest large sums of money for acquiring expensive hardware as beauty spots that are to cover up these conceptual blemishes, sooner or later you must be prepared for the other shift I spoke of before, the user's shift from asking "Where is the answer . . . ?" to "What is the answer . . . ?"

This means, in other words, that the users of a future library system will not care for having access to some documents that may or may not contain the answer to their question, but will request to have a direct access to the *semantic content* of these documents, and will not care whether the answer given by this system is a verbatim citation from a particular document, or it is an equivalent paraphrase. The user wants to know the facts and does not care how they are described as long as the reply is correct and meets his needs.

CHALLENGE

This is a very severe challenge indeed, and it cannot be met by librarians standing alone. I hasten to add that this challenge cannot be met by any single science either. It will require the cooperation of broadminded experts in a wide spectrum

of sciences to construct the kind of systems that will be demanded from you.

I hope you realize that my propositions do not challenge the concept of a library as a center where knowledge can be acquired. What I do challenge, however, is the concept of the book—or its related forms of documentation—as the basic vehicle for knowledge acquisition.

If with these two statements I appear to contradict myself for a "library" is nothing else but a *bonum librorum copia*, a "wealth of good books," then it is only when insisting on this narrow definition. However, this definition shows how strongly our culture identifies the book, the carrier of the printed word, as the depository of all wisdom and knowledge, a belief that may be traced back to our Judeo-Christian heritage: "And the Lord delivered unto me two tables of stone written with the finger of God."[4] Note here that the word of God is written and can be read; but, except for Moses, it is not spoken and thus cannot be heard.

CONFUSION

Perhaps, in the extension of our worship of the Scriptures that are considered *to be* the words of God to other scriptures that are just representation of facts or ideas lies the origin of the confusion that identifies the object with the symbol that it represents. When this identification can no longer be seen as a confusion and becomes a mode of thinking, it is recognized as a symptom of schizophrenia. A patient asked: "how much is 5 × 5?" may deliver a detailed description of his home, for he happens to live on *25* East Main Street.

While this confusion between facts and their descriptions in ecstatic religious, patriotic and pathological states is to be recognized, it is recommended not to adopt it, for otherwise we do relinquish our power to judge a proposition (a description) to be either true or false. Facts are as they are: they are neither true nor false. It is only the descriptions of these facts which are either true or false.

However, for this confusion extenuating circumstances may be cited, namely, that any description is, in turn, itself a fact. But as it is with any tool that is a thing but has a purpose other than itself, so it is with a description that is a fact but has a purpose other than itself. Of course, there are those of us who collect books because of the beauty of their binding or their print; and, of course, there are the books of fiction whose fictions are the facts.

Since the user of our future library wishes to know facts, we have to tell him "100° Centigrade" if he asks for the boiling point of water, and we have to give him *Lady Chatterly's Lover* if he asks for biographical details of this charming lady.

The question now arises, what are the inner workings of such a system in which you can act the double role of a custodian of books and of a midwife for new ideas and insights, thus maintaining the concept of a library as being a place where knowledge can be acquired?

This is tantamount to asking two fundamental questions. First: "How is knowledge acquired?" and second: "How do we mechanize this process?"

COGNITION

If we wish to answer the first question we must find a solution to the problem of cognition. It is only lately that we even

begin to understand the profundity of this problem. We do not have yet the epistemology, the logic, the mathematics and the design of experiments which will give us solid ground for comprehending this enigmatic property of living things. However, the little that has been learned lately has changed substantially many of our cherished concepts, and these new insights, in turn, imply a radical departure from previous thoughts on systems that are supposed to aid their users in acquiring the knowledge that they seek.

Since I shall speak in a moment about the functional organization of such systems that, hopefully, will be yours in the future, let me briefly state the conceptual framework within which we have to search for answers.

The root of the cognitive problem is twofold: epistemological and computational. Since a living organism is an autonomous entity[5], we have to come to grips with the epistemology of "autonomy." By autonomy we mean that all decisions regarding an organism's actions are made within its skin. A living organism is a universe in itself. This implies that, unlike physics, a complete formalism for biology must close on itself. The following example consisting of two complementary propositions may illustrate this point.

(i) The interpretations of an organism's sensations determine its activity.

(ii) An organism's activity determines the interpretations of its sensations.

Such a circular explanation is usually called *circulus vitiosus*. However, by looking closer one will discover that it is this circularity that keeps the system going. For those of you who wish to know more about the legitimacy of such closed formalisms I recommend the work by Katz[6], Lofgren[7], and Brown[8]. These authors have successfully argued the logical consistency and completeness of such formalisms. The rigorous mapping of these calculi onto identifiable functional elements in living organisms, however, has as yet not been made, simply because the crucial experiments for this ethology are only in an early state of design. On the other hand, there are many observations that suggest the validity of this approach. For instance, subjects who wear for an extended period of time spectacles that optically invert the visual field report that during the first days the world they see is upside-down; gradually, however, it turns right side up again, first in the proximity within arm's reach, then regions a few steps away, later more distant places and, finally, after two or three months the whole visual field is experienced as it was without glasses. This suggests that it is the motorium that organizes the sensorium. This may go very far indeed as one adapted subject reports: the first snowfall of that year was perceived as going upward in an otherwise normal scenario.

Such observations should make us think twice before we talk of "information" as if it were a commodity outside of a perceiver's mind. The world does not contain any information: the world is as it is;[9] information about it is created in an organism through its interaction with this world. If some of the advanced document storage and retrieval systems are called Information Storage and Retrieval systems, then we fall into a dangerous

semantic trap. These systems store books, tapes, microfiches or other forms of documents because, of course, they can't store "information." It is again these books, tapes, microfiches or other documents that are retrieved which only when looked upon by human eyes may yield the desired information. By confusing *vehicles* for potential information with *information* one puts the problem of cognition nicely into one's blind spot of intellectual vision.

Let me turn now to the computational problems of cognition. These are encountered already on the most fundamental level, for instance, when we ask "what constitute the so called 'sensory modulities'?", and extend to the level of the higher mental functions, for instance, when we ask "what is memory?", "what is learning?", etc.

You may be surprised to hear that we do not yet understand the neural computations that lead to the experiences of sound, of light and color, of smell and taste, of space and shape, and so on. It was believed that receptor cells that are sensitive to specific stimuli only, say, to certain wave lengths in the electro-magnetic spectrum (the "cones" in the retina), to light intensity (the "rods"), to molecular configurations (the taste buds and organs of smell), etc., could account for these distinct experiences. This is not so, however, for none of these receptors encode into their activity the physical cause of their activity. The only message they can convey is: "There is so and so much (but not of 'what') at this point on my body." Consequently, since sensory receptors are unable to transmit the distinctions of the physical agents that caused them to respond, our experience

of the "glorious variety" of the physical world, the "what," is the result of computations on the receptors' signals.

SEMANTIC COMPUTATIONS

Let me demonstrate the notion that it is computing, rather than signalling and storing of these signals, which is at the core of cognitive processes by contemplating for a moment "memory." For it is precisely the misconception of this higher mental function to be a "data storage system" which blocks the vision for the kind of systems we shall need to meet the challenge of the future.

If memory were a data storage system it is easy to show that in order to account for what we know each of our brains should be the size of a sphere about one mile in diameter packed with nerve cells.[10] However, when of this size, the operation to recognize, for instance, the presence of a lion in its field of vision takes this brain about ten years. This might be helpful to the lion but, alas, not for the bearer of this brain.

To make this point utterly clear assume for the moment that we wish to make numerical multiplication error-free by storing the product of two numbers up to n digits in a printed table. The length of the bookshelf to accommodate this table printed on 8 1/2" × 11" double bond paper is easily calculated to be of length L:

$$L = n.10^{(2n-6)} \text{ centimeters}$$

That is, for products with factors from 1-1000 (n = 3), the shelf must accommodate a book 3 cm, or 1 1/4", thick. While this may suffice for a kid in grade school, to accommodate standard commercial need we have to go up to ten-digit numbers (n = 10). Then this table becomes

10^{15} cm long, that is a hundred times the distance of Earth to Sun, or about one lightday long. A librarian, moving with the velocity of light will, on the average, require one half a day to look up a single entry in the body of this table.

Compare the size of this store with a computer that fits into your hand and does exactly the same. (Figure 1). With on the average of thirty turns of the crank on the top the desired results involving the multiplication of two eight-digit numbers appear in the windows of the product register. Clearly, this device does not store data, it computes *on* data that are, in this example, the factors of a product. If, in this case, one wishes to speak at all of "storage," then it is only with regard to the intrinsic mechanical structure of this device that "embodies"—so to say —the principle of numerical computation.

While we do not understand in detail how the nervous system accomplishes
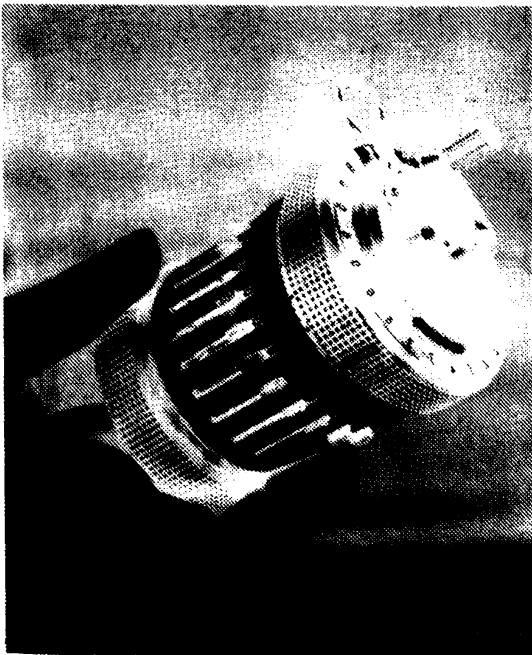
1. *"Curta," a manual digital computer accommodating products up to $10^{16} - 1$.*

these computations, our understanding of the logical relationships that are involved here permits us to represent these relationships in the structure of mechanical or electronic systems that cannot do otherwise but carry out the operations that their structure prescribes them to do.

The relationships to be considered when computing in the logico-mathematical domain are very well understood, hence the success of devices that incorporate these relationships: the large digital computer systems. However, the structure of semantic relationships which is embodied in the functional and anatomical organization of our brains, and which makes us respond to and interact with others through language and behavior, is only now being explored and slowly understood. Until recently, linguists were not too helpful in solving this problem. They were preoccupied with *syntax*, i.e., the rules by which symbols may be concatenated to form legitimate strings; but *semantics*, i.e., the rules that give meaning to those strings, was long a dirty word. This is not so any longer after it had been recognized that syntactic ambiguities are disambiguated in the semantic domain. With the steady advance in this new field of knowledge, "psycholinguistics," it becomes now possible to expand the notion of computation to include computations in the semantic domain.

Since this notion will be crucial in contemplating the computer architecture of knowledge acquisition centers of the future, let me give you an example of semantic computation which we owe to Weston.[11]

Weston contemplated the relational structure that is implicit in puzzles that

are presented by first telling a story in the form of a set of apparently disconnected statements, and then asking for particulars which seem impossible to find. Puzzle fans refer to these as of the "Smith, Robinson and Jones" variety.

Here is one analyzed by Weston:

A train is operated by three men: Smith, Robinson, and Jones. They are engineer, fireman, and brakeman, but not necessarily respectively. On the train are three businessmen of the same names, Mr. Smith, Mr. Robinson, and Mr. Jones. Consider the following facts about all concerned.

(1) Mr. Robinson lives in Detroit.
(2) The brakeman lives halfway between Chicago and Detroit.
(3) Mr. Jones earns exactly $20,000 annually.
(4) Smith beat the fireman at billiards.
(5) The brakeman's nearest neighbor, one of the passengers, earns three times as much as the brakeman, who earns $10,000 a year.
(6) The passenger whose name is the same as the brakeman's lives in Chicago.

This is all that is given. After that the following questions may be asked. For instance:

"Who is the engineer?"

Or

"What relationship holds between Jones and the passenger with the same name as the engineer?"

and so on.

These are apparently quite outrageous questions, but I might do well in reminding you that these are precisely the kind of questions that will be asked of you in the future, and it will be your task to extract the answers from a "data base" that

looks somewhat like the one given in the puzzle.

How to go about in answering these questions? Weston asked himself whether or not the situation in the problem statement can be translated into a single complex relational structure which may serve as a basis for all subsequent operations called for by the questions. He observed that in this particular case the relational structure of the problem statement is based on five binary relations amongst elements belonging to five distinct sets: the set of trainmen, TM, with the elements {Smith, Robinson, Jones}; of jobs J = {engineer, fireman, brakeman}; of passengers P; of locations L; and of salaries S, with the easily identifiable corresponding elements. The five binary relations may be called "namesake:" n; "occupation:" o; "wins over:" w; "resides in:" r; and "makes money:" m. These relations are called "binary" for they stipulate a relation amongst two "variables," for instance n(x,y), or in words: "x is the namesake of y." Similarly m(x,y) stands for "x makes the amount of y dollars per year" and so on.

With this observation it is possible indeed to represent the entire problem statement in a single relational structure (Figure 2). Sets are represented by the appropriately labeled horizontal lines; the elements as points along the corresponding "set lines," and relations by the strings of letters that bear the name of the relation, and connect the elements in question. With little effort you may "read" the problem statement from this figure, and vice versa.

This figure represents the "data base" within which all further computations may take place. First note that this data
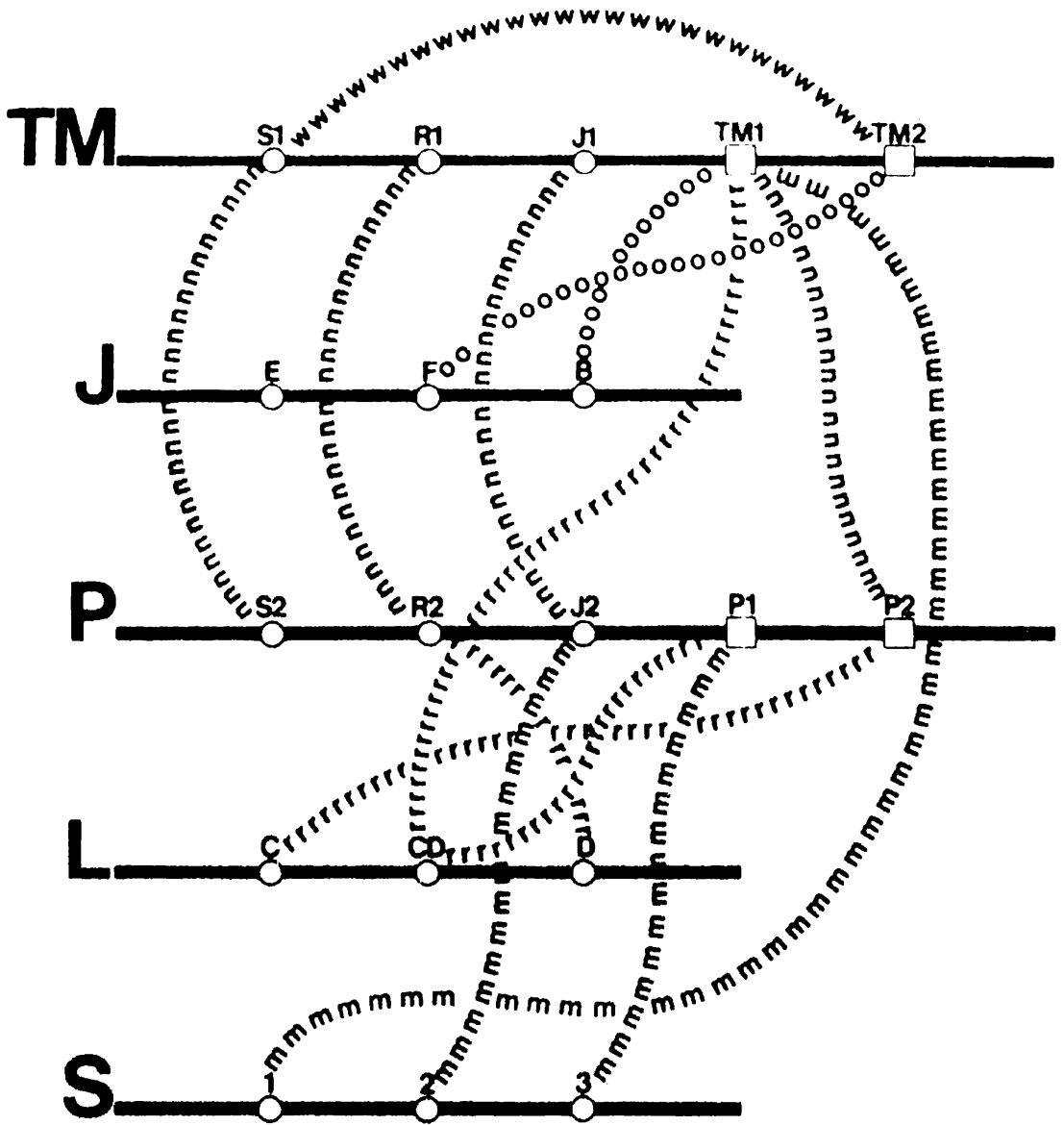
2. *Semantic Structure of the "Smith, Robinson and Jones" problem statement.*

base is equivalent to the other data base which is the problem statement, i.e., the "document," however, with the profound distinction that now one can look directly at the semantic model of this puzzle while in the document it remains obscure.

Moreover, in this representation an "algorithm," that is a computational rule, can be designed that carries out all required deductions. We may not bother with such computations, we may leave this to machines.

I hope that this example gives at least a vague idea of what is meant by "computing in the semantic domain" and what this means with regard to our future user who has now direct access to the semantic structure of the data base. Since the concept of a "document" has been lost in the distributed "wisdom" of this relational data base, he may enter it at any point he wishes and his question will be "paraphrased" to give an answer. If he is satisfied, he leaves; if not, he may ask again. "*The* Answer" is, at any rate, a myth. Answers to the question: "When was Napoleon born?" may be "Fifty-two years before he died in St. Helena," "One thousand seven hundred and seventy-nine years after Jesus Christ was born," "Seven years before the American people declared their independence," and so on. The one you prefer depends on who you are.

There are two questions that may now be raised, one is: "are there machines that can translate the documents into this kind of data base?" and the other one is: "can machines embody such a data base?" While the answer to the first question is a slow "yes," the answer to the second question is a definite "yes." Weston[12] has developed a program structure, called CYLINDER, which allows the mapping of embedded relational structures of arbitrary depth, and the computer hardware to incorporate such structures is available.

The slow "yes" to the former question, however, is not motivated by lack of confidence, knowledge or machine capacity. At that stage it is more the lack of funds than any other single cause that holds us back. This seems to be a trivial cause: we all lack funds! That here this problem is not so trivial will be seen presently when I shall give you a brief sketch of the various approaches to the mechanization of such systems with which a user may strike up a lively and enlightening conversation.

COMPUTERS FOR SEMANTICS

There are in essence two major lines of approach to the design of computer systems that respond to questions posed in the user's natural language. One approach goes under the appropriate name of "Question-Answering System," or QA for short, and historically it is the precursor of the other type I shall call "Cognitive Memory," or CM for short. While these two systems are similar in the sense that they both accept and return statements in a language that any user, in turn, may return and accept, they differ in the sense that in the QA system the data base is an unalterable "codex," to be changed only by the system's programmers when new or other data are available, the CM data structure changes after each interaction so as to include the outcome of these interactions for augmenting the richness of its structure. This distinction may be captured by saying that QA systems are "machine invariant" while CM systems are "user adaptive."

In the following I shall attempt to sketch the "anatomy and physiology" of the two species of computer systems with the explicit apologies, however, that in the remaining time I shall be unable to do justice to the complexity, sophistication and ingenuity that underlie the workings of these systems.

QA System

The basic conceptual principle adopted

Q. AM I ALLOWED TO PASS A CAR ON THE RIGHT.

A. ONLY IF YOU ARE ON A STREET OR HIGHWAY WITH TWO OR MORE
    UNOBSTRUCTED LANES IN YOUR DIRECTION.

Q. HOW OLD MUST A PERSON BE BEFORE HE CAN APPLY FOR A DRIVER*S
    LICENSE IN ILLINOIS.

A. THE MINIMUM AGE FOR A DRIVER*S LICENSE IS 18 YEARS, EXCEPT
    FOR THOSE PERSONS 16 AND 17 YEARS OF AGE WHO HAVE
    SUCCESSFULLY COMPLETED AN APPROVED DRIVER EDUCATION COURSE.

3. *Print-out of an exchange of questions and answers in the R2 system.*
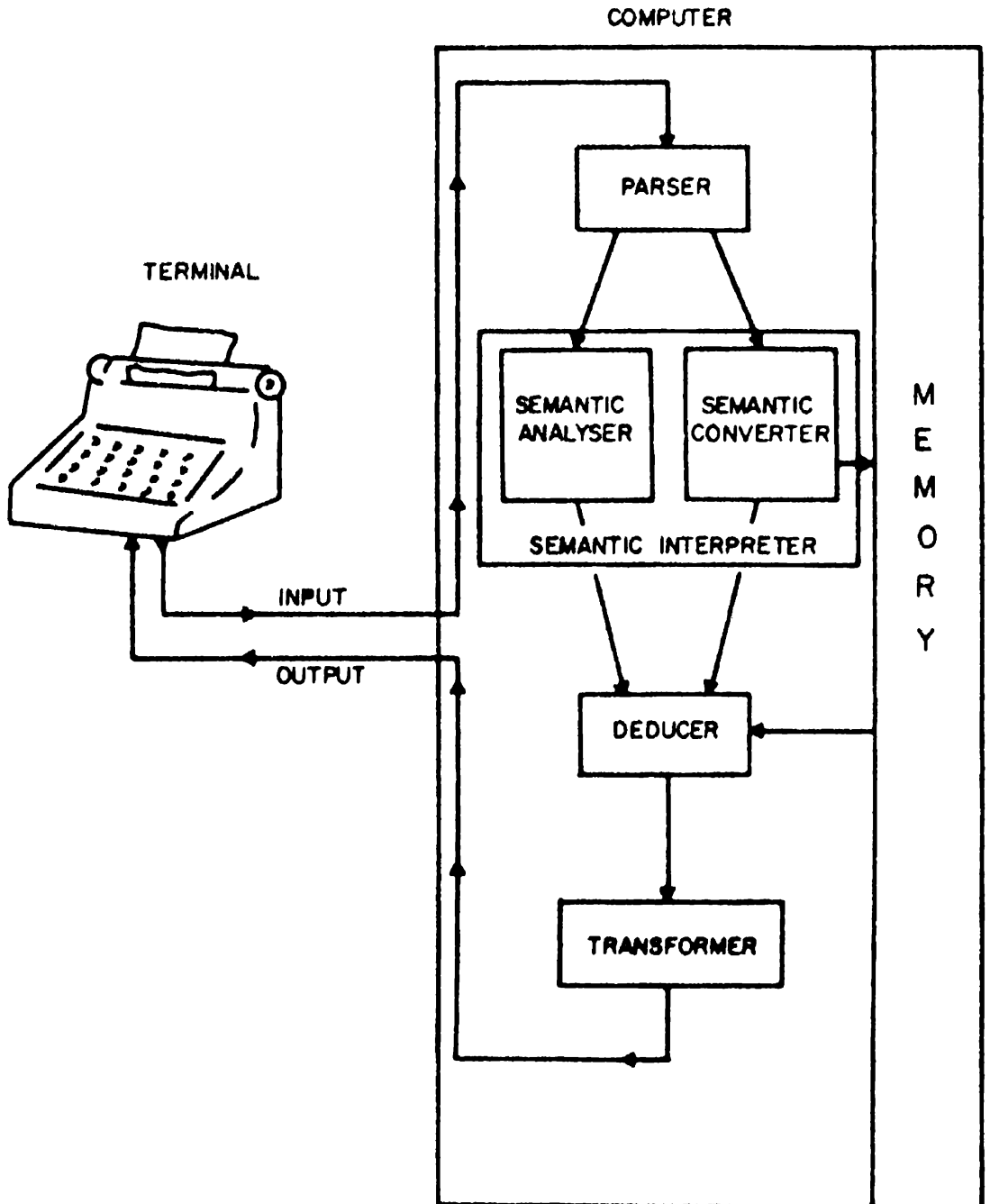
4. *Schematic representation of the R2 Question-Answering System.*

I hope that this example gives at least a vague idea of what is meant by "computing in the semantic domain" and what this means with regard to our future user who has now direct access to the semantic structure of the data base. Since the concept of a "document" has been lost in the distributed "wisdom" of this relational data base, he may enter it at any point he wishes and his question will be "paraphrased" to give an answer. If he is satisfied, he leaves; if not, he may ask again. "*The* Answer" is, at any rate, a myth. Answers to the question: "When was Napoleon born?" may be "Fifty-two years before he died in St. Helena," "One thousand seven hundred and seventy-nine years after Jesus Christ was born," "Seven years before the American people declared their independence," and so on. The one you prefer depends on who you are.

There are two questions that may now be raised, one is: "are there machines that can translate the documents into this kind of data base?" and the other one is: "can machines embody such a data base?" While the answer to the first question is a slow "yes," the answer to the second question is a definite "yes." Weston[12] has developed a program structure, called CYLINDER, which allows the mapping of embedded relational structures of arbitrary depth, and the computer hardware to incorporate such structures is available.

The slow "yes" to the former question, however, is not motivated by lack of confidence, knowledge or machine capacity. At that stage it is more the lack of funds than any other single cause that holds us back. This seems to be a trivial cause: we all lack funds! That here this problem is

not so trivial will be seen presently when I shall give you a brief sketch of the various approaches to the mechanization of such systems with which a user may strike up a lively and enlightening conversation.

COMPUTERS FOR SEMANTICS

There are in essence two major lines of approach to the design of computer systems that respond to questions posed in the user's natural language. One approach goes under the appropriate name of "Question-Answering System," or QA for short, and historically it is the precursor of the other type I shall call "Cognitive Memory," or CM for short. While these two systems are similar in the sense that they both accept and return statements in a language that any user, in turn, may return and accept, they differ in the sense that in the QA system the data base is an unalterable "codex," to be changed only by the system's programmers when new or other data are available, the CM data structure changes after each interaction so as to include the outcome of these interactions for augmenting the richness of its structure. This distinction may be captured by saying that QA systems are "machine invariant" while CM systems are "user adaptive."

In the following I shall attempt to sketch the "anatomy and physiology" of the two species of computer systems with the explicit apologies, however, that in the remaining time I shall be unable to do justice to the complexity, sophistication and ingenuity that underlie the workings of these systems.

QA System

The basic conceptual principle adopted

221

Q. AM I ALLOWED TO PASS A CAR ON THE RIGHT.

A. ONLY IF YOU ARE ON A STREET OR HIGHWAY WITH TWO OR MORE
    UNOBSTRUCTED LANES IN YOUR DIRECTION.

Q. HOW OLD MUST A PERSON BE BEFORE HE CAN APPLY FOR A DRIVER*S
    LICENSE IN ILLINOIS.

A. THE MINIMUM AGE FOR A DRIVER*S LICENSE IS 18 YEARS, EXCEPT
    FOR THOSE PERSONS 16 AND 17 YEARS OF AGE WHO HAVE
    SUCCESSFULLY COMPLETED AN APPROVED DRIVER EDUCATION COURSE.

3. *Print-out of an exchange of questions and answers in the R2 system.*

4. *Schematic representation of the R2 Question-Answering System.*

here is maintaining in essence the structure of the document as given, taking the N primitive sentences that form the data base as the axioms $A_1, A_2, A_3, \ldots A_n$, of a logical system, and treating each incoming question as a theorem that can be proven to be either true or else false. For instance, the question: "Am I allowed to do X?" will be treated as: "Prove the validity of X." If X is valid, the answer is "yes," otherwise it is "no." However, since such a terse reply is most certainly unsatisfactory to the user—at least he wants to know why—advanced QA systems point out the "reasons" for the answer by referring to the axioms in the data base which had been used to furnish the conclusion.

The printout of a typical conversation with a QA system specializing in the rules of driving on the highways of the State of Illinois is given in Figure 3.

Figure 4 sketches the internal workings of this particular QA system called "R2" for "Rules of the Road" the title of the Illinois Driver Manual whose regulations form the axioms of R2.[13] English sentences and questions are typed into the system by the user on a computer terminal keyboard. Next, these sentences are syntactically analyzed by the Parser, and the analysis submitted for disambiguation to the Semantic Interpreter which, in connection with the data base stored in the Memory performs the necessary semantic analyses and conversions that permit comparison with entries in the data base. When this is done, the computations for proving the submitted questions are initiated (Deducer), the results translated into the user's language (Transformer), and, finally, printed out on his terminal.

CM SYSTEM

While in QA systems the burden of semantic computation is given over to the analysis of each question which is then matched against a record of the original document, i.e., the description of the SRJ-puzzle on page 218 CM systems take up the burden of semantic computation already when the data base is formed, e.g., in constructing the single relational structure of the SRJ-puzzle as shown in Figure 2. While this initial complexity does not appreciably reduce the computations that are called for when entering a question, this system allows its data structure to change with each interaction by removing linkages that represent *implicit* relationships, and replacing them by new linkages that represent the relationships that are now made *explicit* by the deductions and conclusions computed on the previous data structure. This is *explication* in the proper sense of the word, and it should be noted that the better understanding of the case in question, which is provided to the user through the answer given by the system, is mirrored in the sytem by being better organized for handling future questions of this kind. Since these processes come closest to models of cognition,[9,14] "Cognitive Memory" appears to be an appropriate name for such systems. Moreover, since these processes entail specific computations whose programs may be part of the data base, such organizations in which the distinction between storage of data and of programs becomes obscure will be called "Cognitive Data Base" or CDB, for short.

Figure 5 symbolizes the inner workings of a CM, and Table I explains the symbols used. Its operations may be
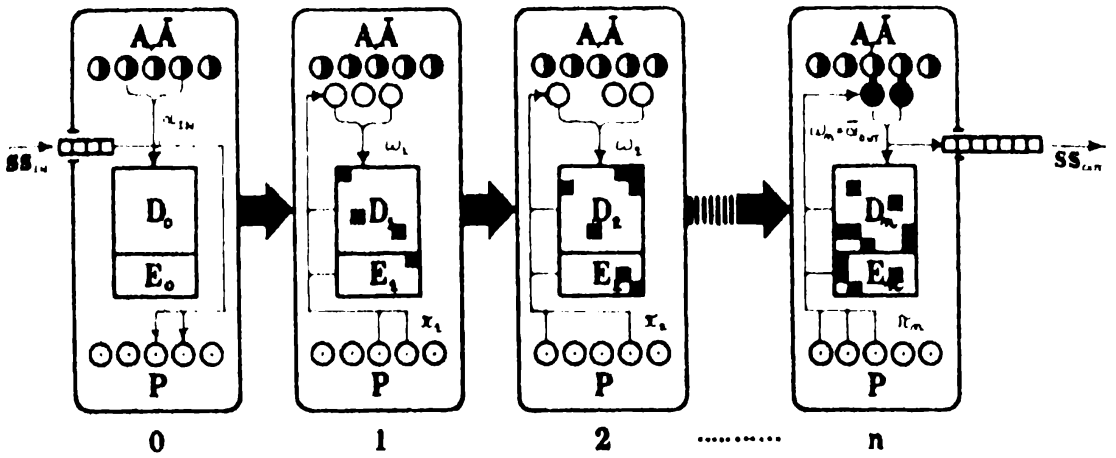
5. *Symbolic Representation of successive computational steps in a Cognitive Memory System.*

TABLE I

GRAPHIC SYMBOL	NAME	SYMBOL	RANGE OF INDEX	SET	SUB-GET	RELATION MATRIX
$\mathbb{C}$	Interface Operators SS ⟶ CDB	a_i	$i = 1 \to n_a$	$A = \{ a_i \}$	$\alpha = \{ a_i \}$	$M(\alpha)$
◖	Interface Operators CDB ⟶ SS	$\bar{a}_i = (a_i^{-1})$	$i = 1 \to n_a$	$\bar{A} = \{ \bar{a}_i \}$	$\bar{\alpha} = \{ \bar{a}_i \}$	$M(\bar{\alpha})$
⊙	Primary Operators	p_i	$i = 1 \to n_p$	$P = \{ p_i \}$	$\pi = \{ p_i \}$	$M(\pi)$
○ ●	Secondary Operators	o_i	$i = 1 \to n_o$	$O = \{ o_i \}$	$\omega = \{ o_i \}$	$M(\omega)$

D ..... Data Structure
E ..... Elementary Programs } Cognitive Data Base, CDB

T ..... Number of time intervals $(\Delta t)_o$, $t = T(\Delta t)_o$, $T = 0,1,2,$

$M(E)_T = |m(E)_{ij}|_T$ .... Matrix of the structure of relations $m(o_i , o_j)$ that prevail between all operators of a subset $E = \{ o_i \}$ which are active during the time interval T.

1. Explanation of symbols used in Figure 5 and in the description of a cognitive memory as given in the text.

sketched in the following steps:

(0) From a console, a string of symbols (the question) (SS_{IN}) is entered into the system. As a consequence, from the set Λ of interface operators an appropriate subset

$$\alpha_{IN} \equiv \{a_i\}$$

for translating this particular string into the data structure is activated. This has two consequences:

(1) (i) Certain domains of the data center become modified [$(D,E)_0 \rightarrow (D,E)_1$, and an appropriate subset of primary operators

$$\pi_1 = \{p_i\}_1$$

becomes activated.

(ii) These assemble in conjunction with the altered data structure from the elementary programs E_2, a set of new secondary operators

$$\omega_1 = \{0_i\}_1$$

which may or may not correspond to a subset of interface operators

$$\bar{A} = \{\bar{a}_i\}.$$

(2) If not, this has two consequences:

(i) Certain domains of the data center become modified [$(D,E)_1 \rightarrow (D,E)_2$], and an appropriate new subset of primary operators

$$\pi_2 = \{p_i\}_2$$

become activated.

(ii) These assemble in conjunction with the altered data structure from the elementary programs E, a set of secondary operators

$$\omega_2 = \{0_i\}_2$$

which may or may not correspond to a subset of interface operators

$$\bar{A} = \{\bar{a}_i\}.$$

(3) If not, then the process is repeated recursively with

$$(D,E)_T \rightarrow (D,E)_{T+1}$$
$$\pi_T > \pi_{T+1}$$
$$\omega_T > \omega_{T+1}$$

•
•
•

(n) until a set of secondary operators ω_n is computed that corresponds to a subset of interface operators

$$\omega_n = \{0_i\} = \{\bar{a}_i\} = \bar{\alpha}$$

(n+1) These operators $\{\bar{a}_i\}$ translate now the appropriate domain of the data structure into a string of symbols (SS_{OUT}) which is printed out at the console.

For instance, in the case of the "Smith, Robinson, Jones" puzzle, after the two questions of page 792 had been entered, the system will respond with

"Smith"

and

"They live at the same place"

respectively.

ECONOMY

Table II gives in different units of size the bulk of material that has to be handled. Along the rows one may find the number of items belonging to a smaller unit which constitute the larger unit. For instance, one article in a journal consists (on the average) of 10 pages, contains 10,000 words, and represents a block of 500,000 bits, etc. The numbers outlined by fat squares represent the primary numerical relations from which all others are derived (adjusted to a lower value).

	BITS	LETTERS	WORDS	SENTENCES	PAGES	ARTICLES	ISSUES (CHAPTERS)	BOOKS	SMALL LIBRARIES	LARGE LIBRARY
1 LETTER	8	1								
1 WORD	48	6	1							
1 SENTENCE	$6 \cdot 10^2$	75	12.5	1						
1 PAGE	$5 \cdot 10^4$	$6 \cdot 10^3$	10^3	80	1					
1 ARTICLE	$5 \cdot 10^5$	$6 \cdot 10^4$	10^4	800	10	1				
1 ISSUE (CHAPTER)	$5 \cdot 10^6$	$6 \cdot 10^5$	10^5	$8 \cdot 10^3$	100	10	1			
1 BOOK	$2 \cdot 10^7$	$3 \cdot 10^6$	10^6	$8 \cdot 10^4$	500	50	5	1		
1 SMALL LIBRARY	10^{12}	$1.5 \cdot 10^{11}$	$2.5 \cdot 10^{10}$	$2 \cdot 10^9$	$2.5 \cdot 10^7$	$2.5 \cdot 10^6$	$\approx 2.5 \cdot 10^5$	$5 \cdot 10^4$	1	
1 LARGE LIBRARY	$2.5 \cdot 10^{14}$	$6 \cdot 10^{13}$	10^{13}	$8 \cdot 10^{11}$	10^{10}	10^9	10^8	$2 \cdot 10^7$	400	1

II. Conversion of various units of documentation.

They have been chosen so as to conform with other estimates (1).

Based on these numbers one may now ask two questions:

(i) "What are the installation costs?"

(ii) "What are the operation costs?"

of computerized systems, extant or proposed, which are capable of handling the interactions with its users for data bases that range from very small collections of documents to huge libraries of the magnitude of the Library of Congress as it is today or, perhaps, in ten or thirty years.

The usual approach in answering these questions is to turn one's attention to the cost of the machines, their maintenance and service. This is the "machine oriented" attitude, by which the labor to create and use the documents is just not seen. However, as we shall see later, it is the "community oriented" attitude which will reveal the hidden costs in using any system.

Presently, however, I shall use the conventional approach in estimating installation costs. In order to have a measure of comparison, I shall discuss three systems, the well known method of using indexing languages (IL), and the two systems mentioned earlier, the question-answering system (QA) and the cognitive memory (CM).

The most easy way to arrive at an estimate of the cost of an entire system is to contemplate the incremental cost increase (Δy in dollars), when its capacity is increased by an incremental amount of "items" (Δx). It appears most convenient to take as the unit of an item a single issue of a journal or, its equivalent in size, the single chapter of a book.

While for IL systems the addition of new material means extracting new index words from already indexed documents, the QA and the CM systems require only additional components to handle the in-

creased data base. Thus we have:

(i) for IL

$$\Delta y = C_1 N \Delta x,$$

but since the number of index words N are related to the size of the data file x through the relation

$$x = C_2 2^N$$

or

$$N = C_3 \ln x$$

we have

$$\Delta y = C_4 . \ln x . \Delta x \qquad (IL)$$

(ii) for QA and CM

$$\Delta y = C_5 \Delta x \qquad (QA)$$
$$\Delta y = C_6 \Delta x \qquad (CM)$$

where the constant C_5 has to absorb the cost of processing units as well as core memory, while C_6 has in essence only to absorb core memory expenses.

The three equations above can easily be integrated to give

$$y = A_0 x \log_{10} \frac{x}{e} + A_2 \qquad (IL)'$$
$$y = B_1 x + B_2 \qquad (QA)'$$
$$y = C_1 x + C_2 \qquad (CM)'$$

where the meaning of the quantities $A_1 (= A_0 \log_{10} \frac{x}{e})$, B_1 and C_1 are the costs of the respective systems per single item, and A_2, B_2 and C_2 are the initial investments for even the smallest system $(x \to 0)$ of each kind.

From systems in operation today the six constants have been roughly estimated and are given in Table III, and the cost functions in Figure 6.

From Figure 6 it is apparent why for small libraries (up to 20,000 books, or 100,000 articles) indexing languages are the popular answer to a librarian's prob-

lem: they are much cheaper than anything else. However his headaches begin when his services are to include collections of one million or more items. In order to be still of service he should have funds to his disposal that are close to 10 million dollars. Since this will be denied to him, he may shop around among the QA or CM systems that give him the desired service and cost much less, about one million dollars. However, should his system still grow, he may be advised to get himself a CM system, for its price of two or three million dollars is "peanuts" compared with all the others.

	Number of Items x			
	10^3	10^5	10^7	
A_1	280	500	800	
B_1		30		cents/item
C_1		1		
A_2	10^3			
B_2	3.10^4			Dollars
C_2	3.10^5			

Table III. Estimate of the installation cost constants for an indexing language system (A_1, A_2), for a question-answering system (B_1, B_2), and for a cognitive memory system (C_1, C_2).
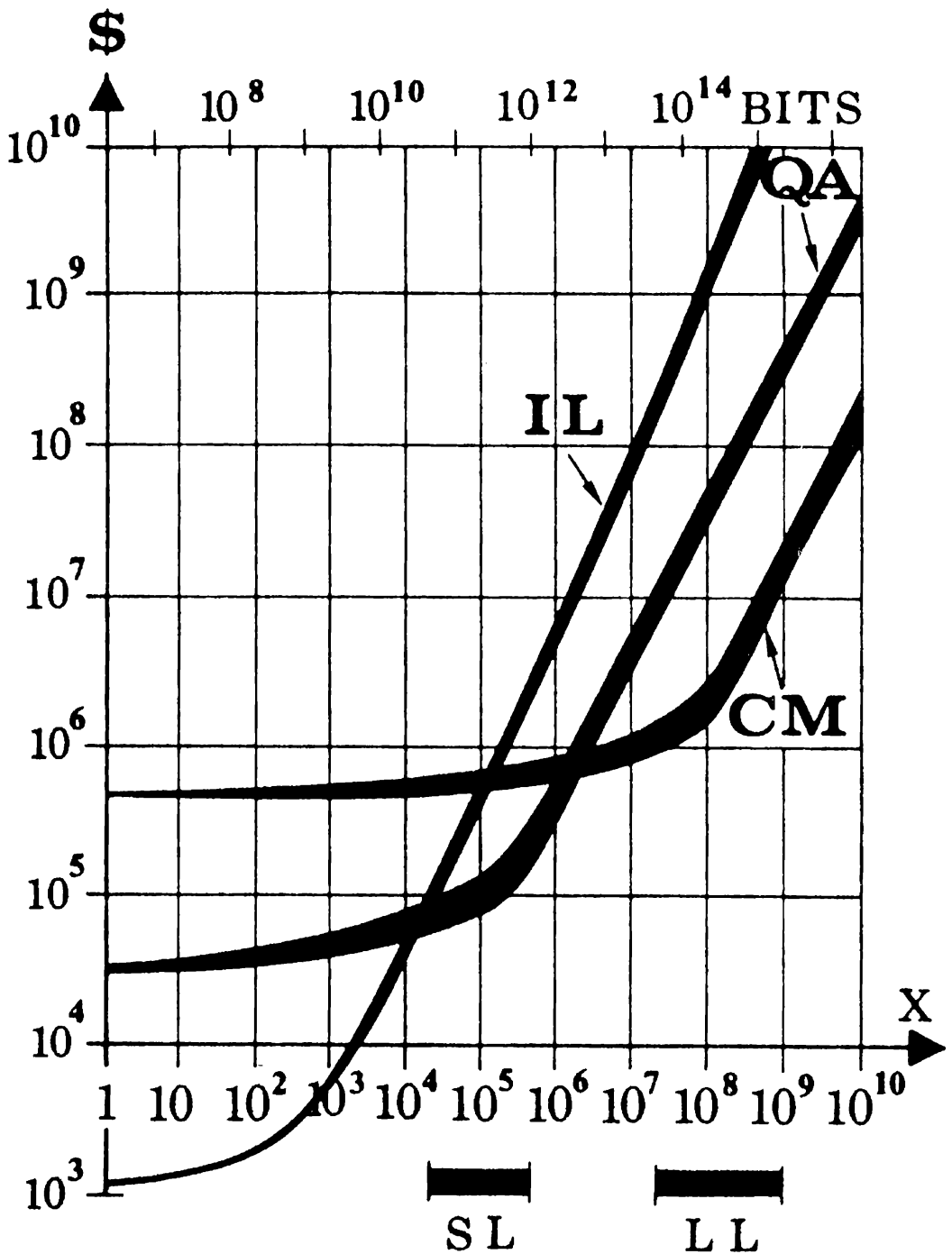
6. *Installation cost functions for an indexing language system, IL., a question-answering system, QA, and a cognitive memory system, CM. Abscissa represents the storage capacity in numbers of journal issues (or chapters of books) bottom, and in bits, top. Ordinate represents costs in dollars. (SL = Small Library; LL = Large Library.)*

Besides giving a suggestion as to the librarian's investment strategies, Figure 6 answers also the problem that I touched before, of funding for these novel systems, for it is the size of the initial investment, e.g., one half of one million dollars for a CM system that does the same as an IL system does for just a few thousand dollars, which makes this problem not a trivial affair.

After suggesting answers to the investment question, we should now turn to the cost-efficiency question. However, I shall forego this exercise, for it has been shown[15] that it is not the machine oriented estimate of costs that counts, but the community's expenses to keep such systems going that is what we have to be aware of.

In a comprehensive paper[15] Weston argues that the size of a scientific community—to be identified, say, by the members of a learned society—is limited by the amount of documents its members are capable to produce and read. When a society grows beyond this size, it begins to break up into "professional groups" with their own journals, meetings and boards, having only the name of the society and its president in common. In other words, there is an upper limit η of the percent of working time a scientist will be willing to devote for communicating with his fellow scientist may he be on the producing or the absorbing side. Say, there are N scientists in the United States, each earning (on the average) S dollars, then the price P for exchanging information through books and journals is per annum

$$P = \eta NS \ \$/year$$

I leave it to you to guess these numbers, but I venture to say that you will have difficulties to stay below an annual multi-billion dollar effort. Moreover, if the investment into these books and journals over an extended period of time is taken into consideration, the numbers assume astronomical proportions. Thus, it is a misguided strategy to invest into increased efficiency of access to these vehicles of potential information: these vehicles are too slow and too expensive. The proper strategy is to invest into the acceleration and facilitation of the development of those new vehicles whose structural principle is the maintenance of semantically related "neighborhoods," rather than the *ad hoc* recovery of these neighborhoods through some artifices from a collection of disconnected items scattered through various documents.

What I am suggesting here is to heed the time honored business principle: When the production volume is large, investments for higher production efficiency pay off. Since the volume we are dealing with here is very large indeed, it is the product NS, the national paycheck for salaries in science and technology, an investment that reduces ever so slightly the quantity η, the "communication viscosity constant" is bound to yield a substantial profit.

With computer terminals at all universities, medical centers, industrial research laboratories, etc., being connected with a centrally located full fledged Cognitive Memory, no books and no survey articles have to be written. The original findings and the arguments that lead to them can be entered directly into CM's data center, and are available in any connection and relation to other findings to a user who wishes to explore such connections and relations without being frustrated by the need of crossing the boundaries of disci-

plines, journals, books, and departments.

Again we have run against the book — or any equivalent of documentation –as the bottleneck in men's communication channels. That I dared to present this view to librarians I can only justify by my conviction that it is not the book that made you choose your profession, but it is to help others in realizing their desires of which Aristotle said:

"All men by nature desire to know."

REFERENCES

1 NAS Publication No. 1707, 336 pp., (1969). To be obtained from National Academy of Science, Printing and Publishing Office, 2101 Constitution Avenue, Washington, D. C., 20418 ($6.95).

2 Planck, M. "Über eine Verbesserung des Wien'schen Strahlungsgesetzes" in *Sitzber. Preuss. Ac. Wiss*, October, (1900).

3 Planck, M. "Über die Theorie des Energieverteilungsgesetzes normaler Spectren" in *Sitzber. Preuss. Ac. Wiss*, December, (1900).

4 Deuteronomy 9/10

5 Maturana, H. R.. "Neurophysiology of Cognition" in *Cognition: A Multiple View*, P. L. Garvin (ed.), Spartan Books, New York, p. 1-23, (1971).

6 Katz, J. J.. *The Problem of Induction and its Solution*, The University of Chicago Press, Chicago, 125 pp., (1962).

7 Löfgren, L.. "An Axiomatic Explanation of Complete Self-Reproduction," Bull. Math Biophys., *30*, 415-425, (1968)

8 Brown, G. A. *Laws of Form*, George Allen and Unwin, London, 141 pp., (1969).

9 Von Foerster, H. "Thoughts and Notes on Cognition" in *Cognition: A Multiple View*, P. L. Garvin (ed.), Spartan Books, New York, p. 25-48, (1971)

10. Von Foerster, H., A. Inselberg and P. Weston: "Memory and Inductive Inference" in *Cybernetic Problems in Bionics*, H. L. Oestreicher and D. R. Moore (eds.), Gordon and Breach, London, p. 31-68, (1968).

11 Weston, P. "To Uncover, To Deduce, To Conclude," Comp. Stud. in Humanities and Verb. Beh.

12 Weston, P. *CYLINDERS: A Relational Data Structure*, Tech Rep. No. 18, AFOSR 70-1865, Biological Computer Laboratory, Electrical Engineering Research Laboratory, Engineering Experiment Station, University of Illinois, Urbana, 78 pp., (1970)

13 Bielby, W., K. Biss, J. Schultz, F. Stahl and T. Woo: "An Answering System for Questions Posed in Natural Language" in *Cognitive Memory* Int. Rep. No. 11, Coordinated Science Laboratory, University of Illinois, Urbana, p. 6-45, (1970)

14 Von Foerster, H.: "What is Memory That It May Have Hindsight and Foresight as well" in *Future of the Brain-sciences*, S. Bogoch (ed.), Plenum Press, New York, p. 19-65, (1969).

15 Weston, P. "An End of Search, A Means of Understanding: A Preface to the Antidocument" in *Accomplishment Summary*, BCL Rep. No. 71.2, Biological Computer Laboratory, Department of Electrical Engineering, University of Illinois, Urbana, (1971).

1/11   Francisco Hoffman (1961)
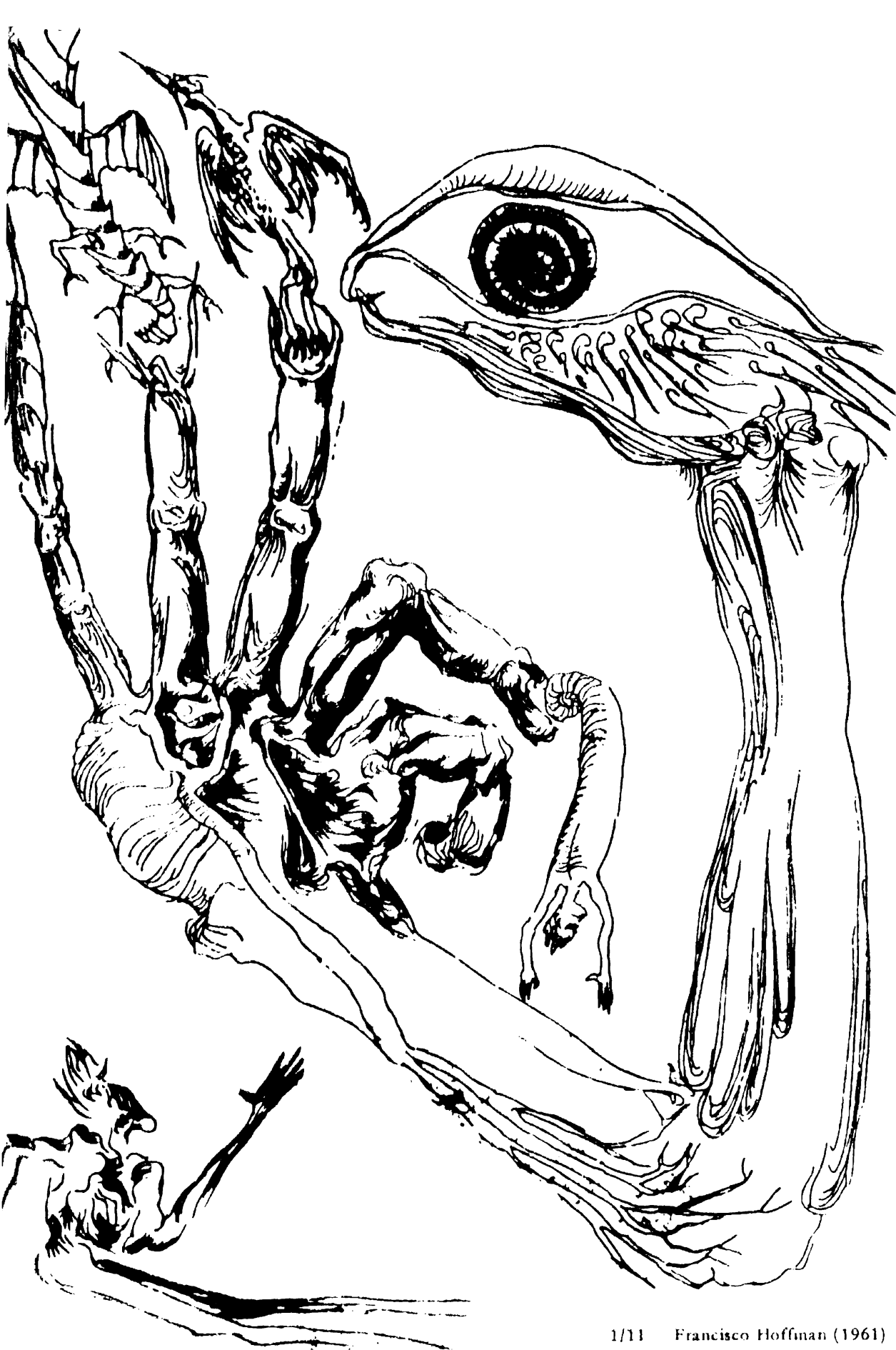
Thoughts and Notes on Cognition*

HEINZ VON FOERSTER

University of Illinois

THOUGHTS

Projecting the image of ourselves into things or functions of things in the outside world is quite a common practice. I shall call this projection "anthropomorphization." Since each of us has direct knowledge of himself, the most direct path of comprehending X is to find a mapping by which we can see ourselves represented by X. This is beautifully demonstrated by taking the names of parts of one's body and giving these names to things which have structural or functional similarities with these parts: the "head" of a screw, the "jaws" of a vise, the "teeth" of a gear, the "lips" of the cutting tool, the "sex" of electric connectors, the "legs" of a chair, a "chest" of drawers, etc.

Surrealists who were always keen to observe ambivalences in our cognitive processes bring them to our attention by pitching these ambivalences against semantic consistencies: the legs of a chair (Fig. 1[2]), a chest of drawers (Fig. 2[3]), etc.

At the turn of the century, animal psychologists had a difficult time

Fig. 1. "L'Ultra Meuble" by Kurt Seligman

in overcoming functional anthropomorphisms in a zoology populated with animals romanticized with human characteristics: the "faithful" dog, the "valiant" horse, the "proud" lion, the "sly" fox, etc. Konrad Lorenz, the great ornithologist, was chased from Vienna when he unwisely suggested controlling the population of the overbreeding, underfed, and tuberculosis carrying pigeons of the city by importing falcons which would raid the pigeons' nests for eggs. The golden heart of the Viennese could not stand the thought of "pigeon infanticide." Rather, they fed the pigeons twice as much. When Lorenz pointed out that the result of this would be twice as many underfed and tuberculosis carrying pigeons, he had to go, and fast!

Of course, in principle there is nothing wrong with anthropomorphizations, in most cases they serve as useful algorithms for determining

behavior. In trying to cope with a fox it is an advantage to know he is "sly," that is, he is a challenge to the brain rather than to the muscles.

Today, with most of us having moved to the big cities, we have lost direct contact with the animal world, and pieces of steel furniture with some functional properties, the computers, are becoming the objects of our endearments and, consequently, are bestowed now with romanticizing epithets. Since we live today, however, in an era of science and technology rather than in one of emotion and sentimentality, the endearing epithets for our machines are not those of character but of intellect. Although it is quite possible, and perhaps even appropriate to talk about a "proud IBM 360-50 system," the "valiant 1800," or the "sly PDP 8," I have never observed anyone using this style of language. Instead, we romanticize what appears to be the intellectual functions of the machines. We talk about their "memories," we say that these machines store and retrieve "information," they "solve problems," "prove theorems," etc. Apparently, one is dealing here with quite intelligent chaps, and there are even some attempts made to design an A. I. Q., an "artificial intelligence quotient" to carry over into this new field of "artificial intelligence" with efficacy and authority the miscon-
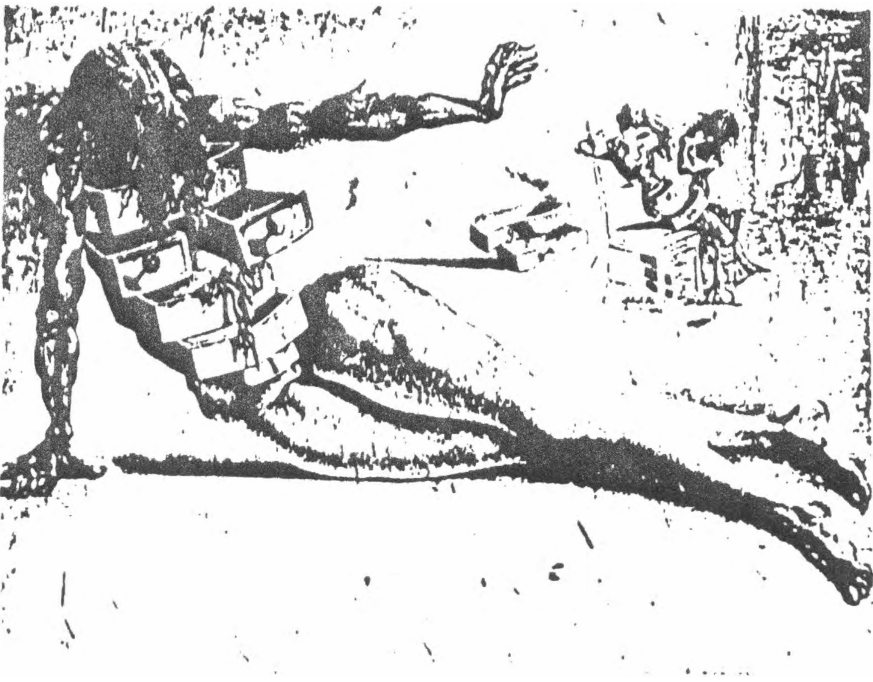
Fig. 2 - "City of Drawers" by Salvador Dali

ceptions that are still today quite popular among some prominent
behaviorists

While our intellectual relationship with these machines awaits clarifi-
cation, in the emotional sphere we seem to do all right. I wish to make
this comment as a footnote to Madeleine Mathiot's delightful observa-
tions in this volume about various degrees of "awesomeness" associated
with the referential genders "it," "he," and "she." She develops a
three-valued logical place value system in which the nonhuman "it"
carries no reference to awesomeness either in the negative (absence) or
else in the affirmative (presence), while the human "he" and "she" indeed
carry reference to awesomeness, the masculine "he" referring to its
absence, the feminine "she," of course, to its presence.

When in the early fifties at the University of Illinois ILLIAC II was
built, "it" was the referential gender used by all of us. The computer
group that now works on ILLIAC III promises that "he" will be opera-
tive soon. But ILLIAC IV reaches into quite different dimensions. The
planners say that when "she" will be switched on, the world's computing
power will be doubled.

Again, these anthropomorphisms are perfectly all right inasmuch
as they help us establish good working relations with these tools. Since
most of the people I know in our computer department are heterosexual
males, it is clear that they prefer the days and nights of their work spent
with a "she," rather than with an "it."

However, in the last decade or so something odd and distressing
developed, namely, that not only the engineers who work with these
systems gradually began to believe that those mental functions whose
names were first metaphorically applied to some machine operations
are indeed residing in these machines, but also some biologists—tempted
by the absence of a comprehensive theory of mentation—began to
believe that certain machine operations which unfortunately carried the
names of some mental processes are indeed functional isomorphs of these
operations. For example, in the search for a physiological basis of
memory, they began to look for neural mechanisms which are analogues
of electromagnetic or electrodynamic mechanisms that "freeze" tem-
poral configurations (magnetic tapes, drums, or cores) or spatial con-
figurations (holograms) of the electromagnetic field so that they may be
inspected at a later time.

The delusion, which takes for granted a functional isomorphism
between various and distinct processes that happen to be called by the
same name, is so well established in these two professions that he who

follows Lorenz's example and attempts now to "de-anthropomorphize" machines and to "de-mechanize" man is prone to encounter antagonisms similar to those Lorenz encountered when he began to "animalize" animals.

On the other hand, this reluctance to adopt a conceptual framework in which apparently separable higher mental faculties as, for example, "to learn," "to remember," "to perceive," "to recall," "to predict," etc., are seen as various manifestations of a single, more inclusive phenomenon, namely, "cognition," is quite understandable. It would mean abandoning the comfortable position in which these faculties can be treated in isolation and thus can be reduced to rather trivial mechanisms. Memory, for instance, contemplated in isolation is reduced to "recording," learning to "change," perception to "input," etc. In other words, by separating these functions from the totality of cognitive processes one has abandoned the original problem and now searches for mechanisms that implement entirely different functions that may or may not have any semblance with some processes that are, as Maturana* pointed out, subservient to the maintenance of the integrity of the organism as a functioning unit.

Perhaps the following three examples will make this point more explicit.

I shall begin with "memory." When engineers talk about a computer's "memory" they really don't mean a computer's memory, they refer to devices, or systems of devices, for recording electric signals which when needed for further manipulations can be played back again. Hence, these devices are stores, or storage systems, with the characteristic of all stores, namely, the conservation of quality of that which is stored at one time, and then is retrieved at a later time. The content of these stores is a record, and in the pre-semantic-confusion times this was also the name properly given to those thin black disks which play back the music recorded on them. I can see the big eyes of the clerk in a music shop who is asked for the "memory" of Beethoven's *Fifth Symphony*. She may refer the customer to the bookstore next door. And rightly so, for memories of past experiences do not reproduce the causes for these experiences, but—by changing the domains of quality—transform these experiences by a set of complex processes into utterances or into other forms of symbolic or purposeful behavior. When asked about the contents of my breakfast, I shall not produce scrambled eggs, I just say, "scrambled eggs." It is clear that a computer's "memory" has nothing to do with such transformations, it was never intended to have. This

*See Chapter 1, pages 3–23.

does not mean, however, that I do not believe that these machines may eventually write their own memoirs. But in order to get them there we still have to solve some unsolved epistemological problems before we can turn to the problem of designing the appropriate software and hardware.

If "memory" is a misleading metaphor for recording devices, so is the epithet "problem solver" for our computing machines. Of course, they are no problem solvers, because they do not have any problems in the first place. It is *our* problems they help us solve like any other useful tool, say, a hammer which may be dubbed a "problem solver" for driving nails into a board. The danger in this subtle semantic twist by which the responsibility for action is shifted from man to a machine lies in making us lose sight of the problem of cognition. By making us believe that the issue is how to find solutions to some well defined problems, we may forget to ask first what constitutes a "problem," what is its "solution," and—when a problem is identified—what makes us want to solve it.

Another case of pathological semantics—and the last example in my polemics—is the widespread abuse of the term "information." This poor thing is nowadays "processed," "stored," "retrieved," "compressed," "chopped," etc., as if it were hamburger meat. Since the case history of this modern disease may easily fill an entire volume, I only shall pick on the so-called "information storage and retrieval systems" which in the form of some advanced library search and retrieval systems, computer based data processing systems, the nationwide Educational Resources Information Center (ERIC), etc., have been seriously suggested to serve as analogies for the workings of the brain.

Of course, these systems do not store information, they store books, tapes, microfiche or other sorts of documents, and it is again these books, tapes, microfiche or other documents that are retrieved which only if looked upon by a human mind may yield the desired information. Calling these collections of documents "information storage and retrieval systems" is tantamount to calling a garage a "transportation storage and retrieval system." By confusing *vehicles* for potential information with *information*, one puts again the problem of cognition nicely into one's blind spot of intellectual vision, and the problem conveniently disappears. If indeed the brain were seriously compared with one of these storage and retrieval systems, distinct from these only by its quantity of storage rather than by quality of process, such a theory would require a demon with cognitive powers to zoom through this huge system in order to extract from its contents the information that is vital to the owner of this brain.

Dificile est satiram non scribere. Obviously, I have failed to overcome this difficulty, and I am afraid that I will also fail in overcoming the other difficulty, namely, to say now what cognition *really* is. At this moment, I even have difficulties in relating my feelings on the profoundness of our problem, if one cares to approach it in its full extension. In a group like ours, there are probably as many ways to look at it as there are pairs of eyes. I am still baffled by the mystery that when Jim, a friend of Joe, hears the noises that are associated with reading aloud from the black marks that follow

ANN IS THE SISTER OF JOE

or just sees these marks—knows that indeed Ann is the sister of Joe, and, *de facto*, changes his whole attitude toward the world, commensurate with his new insight into a relational structure of elements in this world.

To my knowledge, we do not yet understand the "cognitive processes" which establish this insight from certain sensations. I shall not worry at this moment whether these sensations are caused by an interaction of the organism with objects in the world or with their symbolic representations. For, if I understood Dr. Maturana correctly, these two problems, when properly formulated, will boil down to the same problem, namely, that of cognition *per se*.

In order to clarify this issue for myself, I gathered the following notes which are presented as six propositions labeled $n = 1 \longrightarrow 6$. Propositions numbered n.1, n.2, n.3, etc., are comments on proposition numbered n. Propositions numbered n.m1, n.m2, etc., are comments on proposition n.m, and so on.

Here they are.

NOTES

1 A living organism, Ω, is a bounded, autonomous unit whose functional and structural organization is determined by the interaction of its contiguous elementary constituents.

1.1 The elementary constituents, the cells, are, in turn, bounded, functional, and structural units, however, they are not necessarily autonomous.

1.11 Autonomy of cells is progressively lost with increasing differentiation in organisms of ascending complexity which, on the other hand, provides the appropriate "organic environment" for these units to maintain their structural and functional integrity.

1.2 A living organism, Ω, is bounded by a closed orientable surface. Topologically this is equivalent to a sphere with an even number $2p$ of holes which are connected in pairs by tubes. The number p is called the genus of the surface.

1.21 Should the histological distinction between ectoderm and endoderm be maintained, then a surface of genus $p = (s + t)/2$ is equivalent to a sphere with s surface holes which are connected through a network of tubes with t T-branches. Ectoderm is then represented by the surface of the sphere, endoderm by the lining of the tubes (Fig. 3).

1.3 Any closed orientable surface is metrizable. Hence, each point on this surface can be labeled by the two coordinates α, β, of a geodesic coordinate system that may be chosen to cover the surface conveniently. One of the properties of a geodesic coordinate system is that it is locally
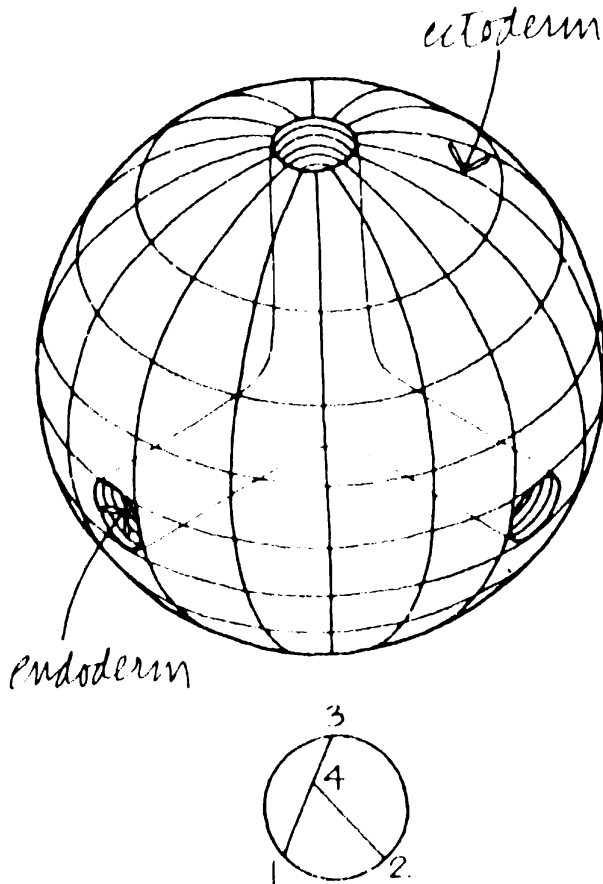
Fig. 3.—Closed orientable surface of genus $p = 2$, ($s = 3$, $t = 1$, $[s + t]/2 = 4/2 = 2$)

Cartesian. Surface coordinates α, β, will be referred to as the "proprietory coordinates," denoted by the single symbol ξ.

1.31 If to the vicinity of each surface point ξ the Gaussian curvature γ is given, then the totality of triples α, β, γ, determines the shape Γ of the surface ($\gamma = \gamma[\alpha,\beta]$).

1.32 Since a living organism is bounded by a closed orientable surface, an appropriate geodesic coordinate system can be drawn on the surface of this organism at an arbitrary "rest state," and each surface element (ectodermal or endodermal cell) can be labeled according to the proprietory coordinates ξ of its location.

1.33 A cell c_ξ so labeled shall carry its label under subsequent distortions of the surface (continuous distortions), and even after transplants to locations ξ' (discontinuous distortions).

1.331 The geodesic coordinates on the surface of the organism can be mapped onto a topologically equivalent unit sphere ($R = 1$) so that to each point ξ and its vicinity on the organism corresponds precisely one point λ and its vicinity on the unit sphere. Consequently, each cell c_ξ on the surface of the organism has an image c_λ on the surface of the unit sphere.

1.332 It is clear that surface distortions of the organism, even transplants of cells from one location to another, are not reflected by any changes on the surface of this sphere. The once established map remains invariant under such transformations, hence this sphere will be referred to as the "representative body sphere" (Fig. 3, or appropriate modifications with $p > 2$).

1.34 Since the volume enclosed by a closed orientable surface is metrizable, all that has been said (1.3 $\longrightarrow$ 1.332) for surface points ξ and cells c_ξ holds for volume points ζ and cells c_ζ with representative cells c_μ in the body sphere.

1.4 The organism, Ω, is supposed to be embedded in an "environment" with fixed Euclidean metric, with coordinates a, b, c, or x for short, in which its position is defined by identifying three environmental points x_1, x_2, and x_3 with three surface points ξ_1, ξ_2, and ξ_3 of the organism. Conversely, the representative body sphere is embedded in a "representative environment" with variable non-Euclidean metric, and with the other conditions *mutatis mutandis*.

1.41 The two pairs of figures (Figs. 4a and 4b and Figs. 5a and 5b) illustrate the configuration of the proprietory space, ξ, of an organism (fish-like creature) as seen from an Euclidean environment (4a and 5a), and the configuration of the non-Euclidean environment as seen from the unit sphere (4b, 5b) for the two cases in which the organism is at rest (4a, 4b) and in motion (5a, 5b).
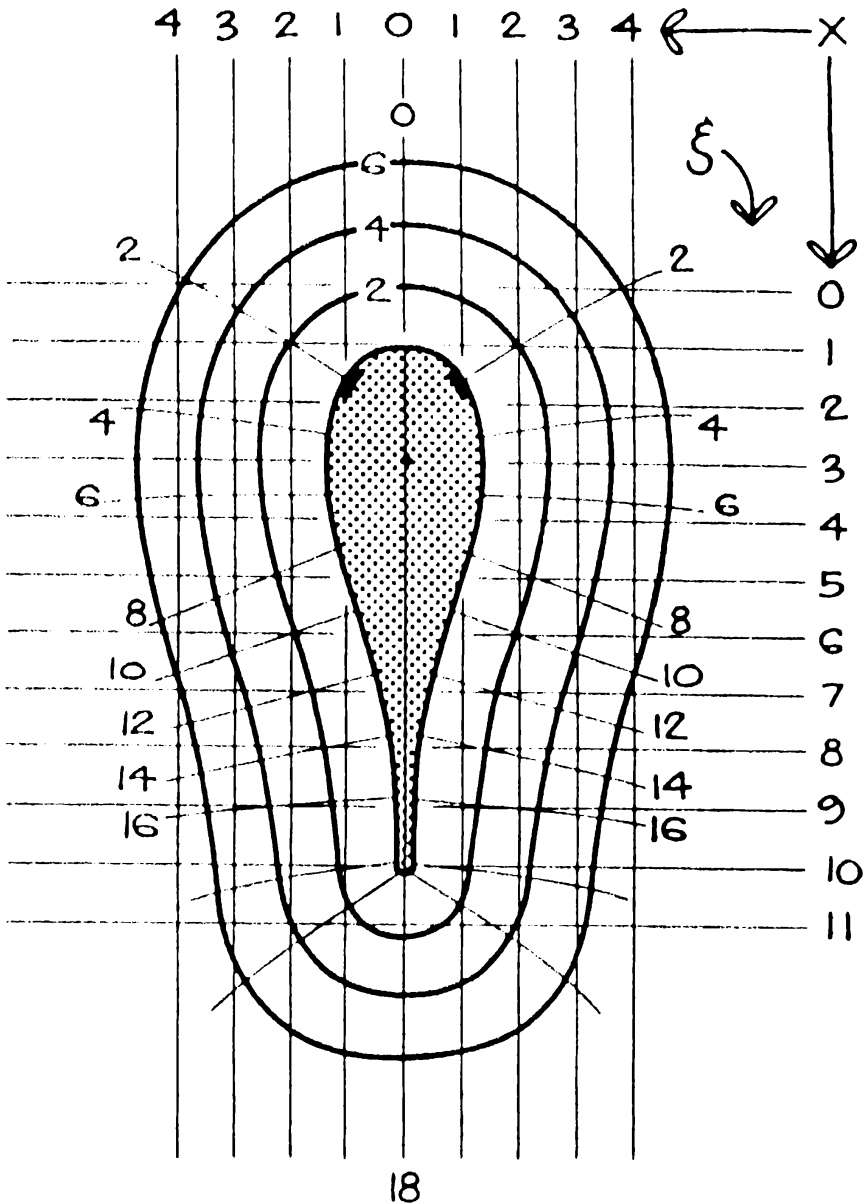
FIG. 4a.—Geodesics of the proprietory coordinate system of an organism Ω at rest [δ𝑙 = 0] embedded in an environment with Euclidean metric

2 Phylogenetically as well as ontogenetically the neural tube develops from the ectoderm. Receptor cells r_i are differentiated ectodermal cells c_i. So are the other cells deep in the body which participate in the transmission of signals (neurons) n_i, the generation of signals (proprio-
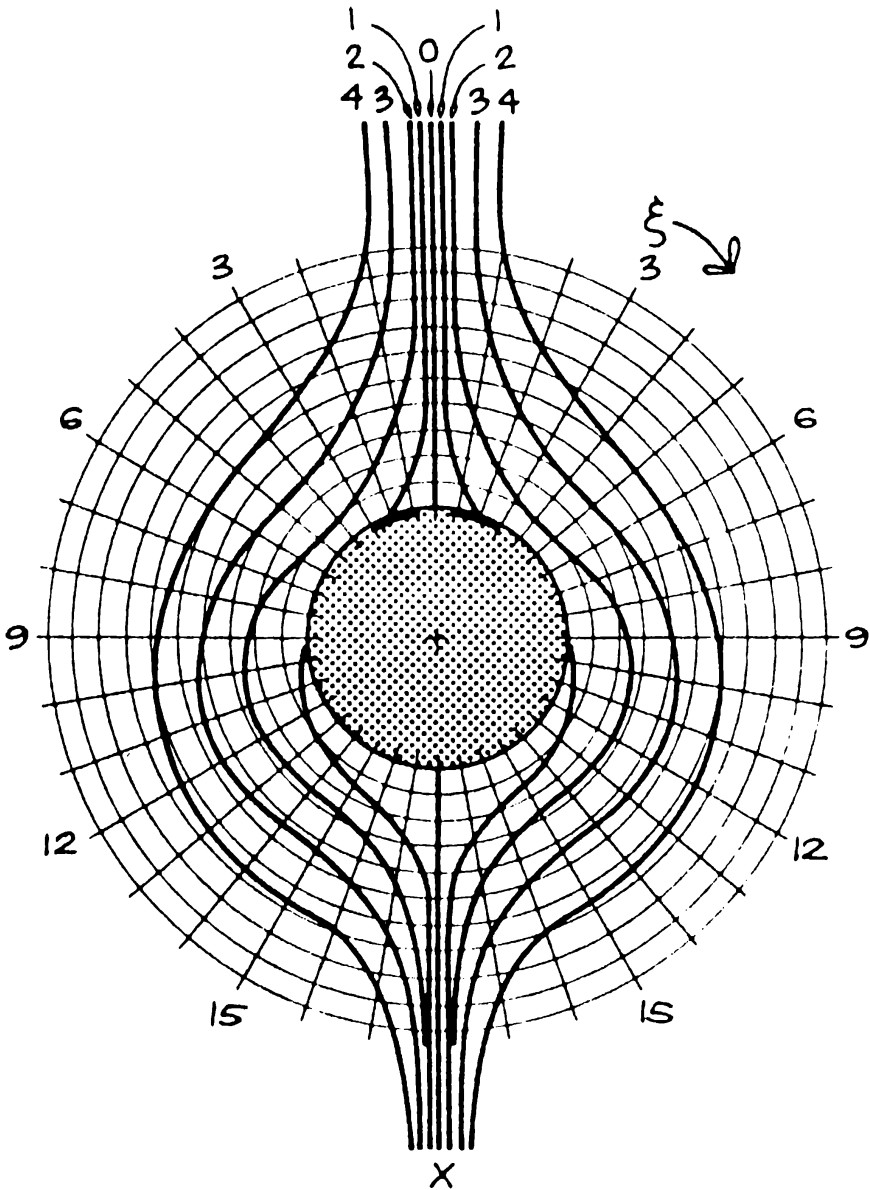
Fıc. 4b.—Geodesics (circles, radii) of the proprietory coordinate system with respect to the representative body sphere embedded in an environment with non-Euclidean metric corresponding to the organism at rest (Fig. 4a)

ceptors) p_ξ, as well as those (effectors) e_ξ which cause by their signaling specialized fibers (muscles) m_ξ to contract, thus causing changes $\delta\Gamma$ in the shape of the organism (movement).

2.1 Let A be an agent of amount A distributed in the environment
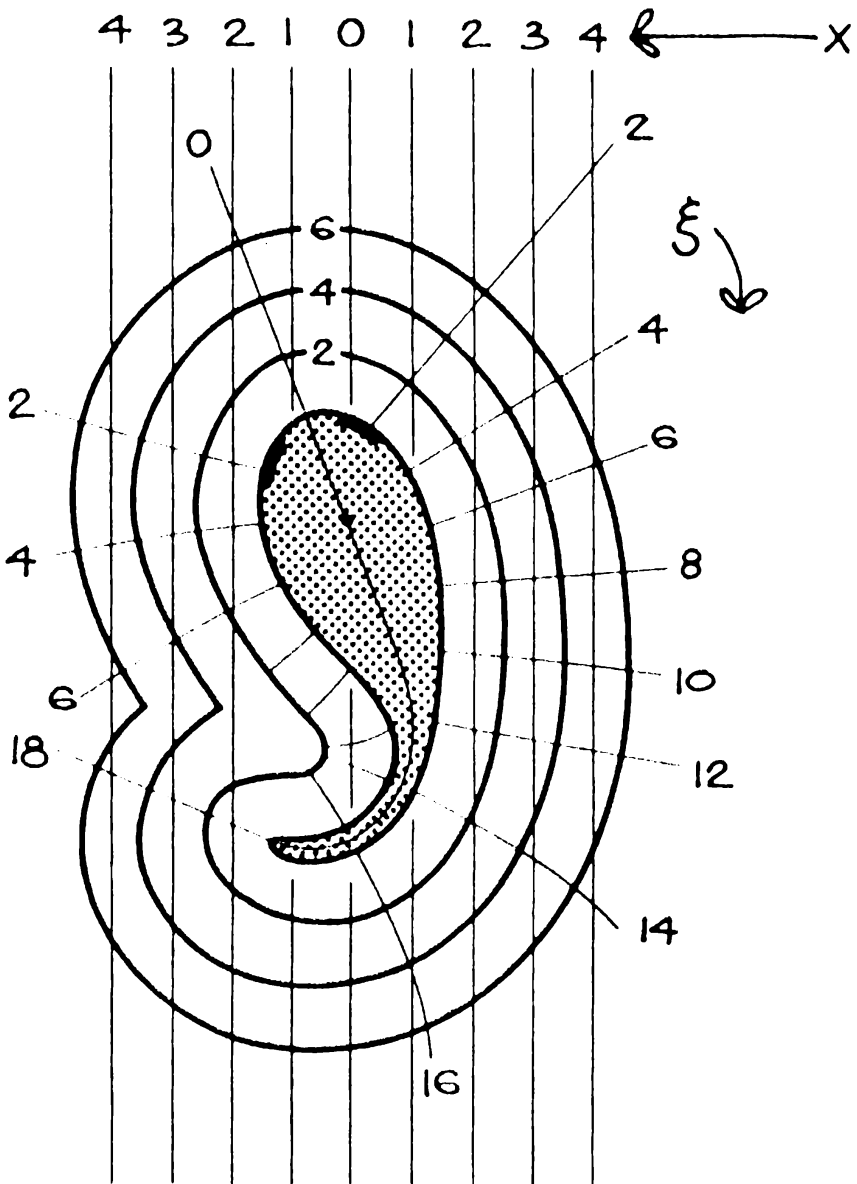
Fig. 5a.—Geodesics of the proprietory coordinate system of an organism Ω in motion ($\delta\Gamma \neq 0$) embedded in an environment with Euclidean metric

and characterized by a distribution function of its concentration (intensity) over a parameter p:

$$S(x,p) = \frac{d^2A}{dx\,dp} \quad \left(\frac{da}{dp}\right)_x$$
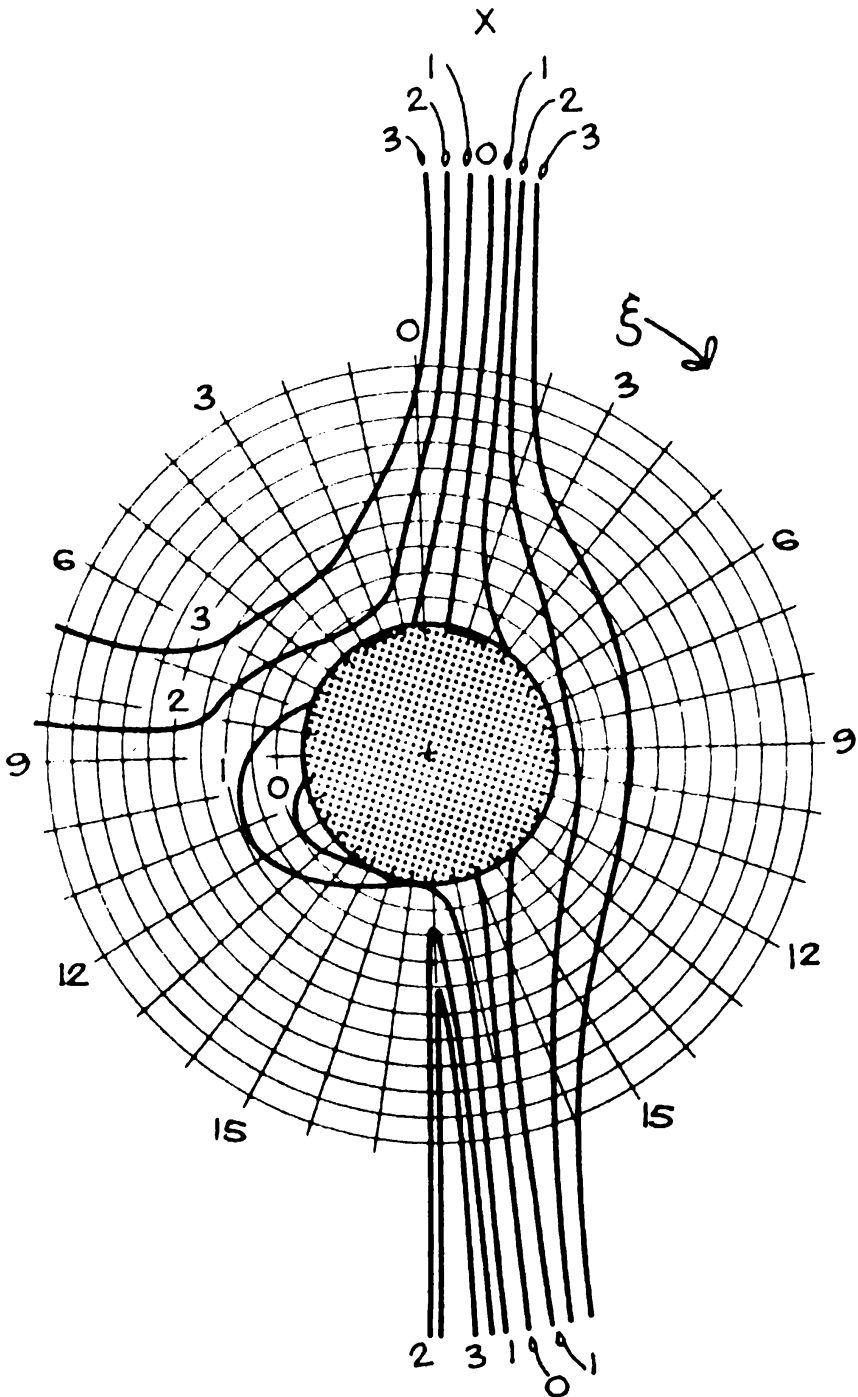
Fig. 5b.—Geodesics (circles, radii) of the proprietory coordinate system with respect to the representative body sphere embedded in an environment with non-Euclidean metric corresponding to the organism in motion (Fig. 5a)

with

$$\int_0^\infty S(x,p)\,dp = a_x = \frac{dA}{dx}.$$

2.11 Let $s(\xi,p)$ be the (specific) sensitivity of receptor r_ξ with respect to parameter p, and ρ_ξ be its response activity:

$$s(\xi,p) = k\frac{d\rho}{da} = k\left(\frac{d\rho}{dp}\right)_\xi \Big/ \left(\frac{da}{dp}\right)_\xi$$

with

$$\int_0^\infty s(\xi,p)\,dp = 1,$$

and k being a normalizing constant.

2.12 Let x and ξ coincide. The response ρ_ξ of receptor r_ξ to its stimulus S_ξ is now

$$\rho_\xi = \int_0^\infty S(\xi,p) \cdot s(\xi,p)\,dp = F(a_\xi).$$

2.2 This expression shows that neither the modality of the agent, nor its parametric characteristic, nor reference to environmental point x is encoded in the receptor's response, solely some clues as to the presence of a stimulant for receptor r_ξ are given by its activity ρ_ξ.

2.21 Since all receptors of an organism respond likewise, it is clear that organisms are incapable of deriving any notions as to the "variety of environmental features," unless they make reference to their own body by utilizing the geometrical significance of the label ξ of receptor r_ξ which reports: "so and so much ($\rho = \rho_1$) is at *this* place on my body ($\xi = \xi_1$)."

2.22 Moreover, it is clear that any notions of a "sensory modality" cannot arise from a "sensory specificity," say a distinction in sensitivity regarding different parameters p_1 and p_2, or in different sensitivities s_1 and s_2 for the same parameter p, for all these distinctions are "integrated out" as seen in expression 2.12. Consequently, these notions can only arise from a distinction of the body oriented loci of sensation ξ_1 and ξ_2. (A pinch applied to the little toe of the left foot is felt not in the brain but at the little toe of the left foot. Dislocating one eyeball by

gently pushing it aside displaces the environmental image of this eye with respect to the other eye.)

2.23 From this it becomes clear that all inferences regarding the environment of Ω must be computed by operating on the distribution function ρ_ξ. (It may also be seen that these operations ω_{ij} are in some sense coupled to various sensitivities $s_i[\xi, p_i]$.)

2.24 This becomes even more apparent if a physical agent in the environment produces "actions at a distance."

2.241 Let $g(r, p)$ be the environmental distribution of sources of the agent having parametric variety (p); let R be the distance between any point r in the environment and a fixed point x_0, and let $\Phi(R)$ be the distance function by which the agent loses its intensity. Moreover let the point ξ_0 on the body of an organism coincide with r_0, then the stimulus intensity for receptor r_{ξ_0} is (Fig. 6):
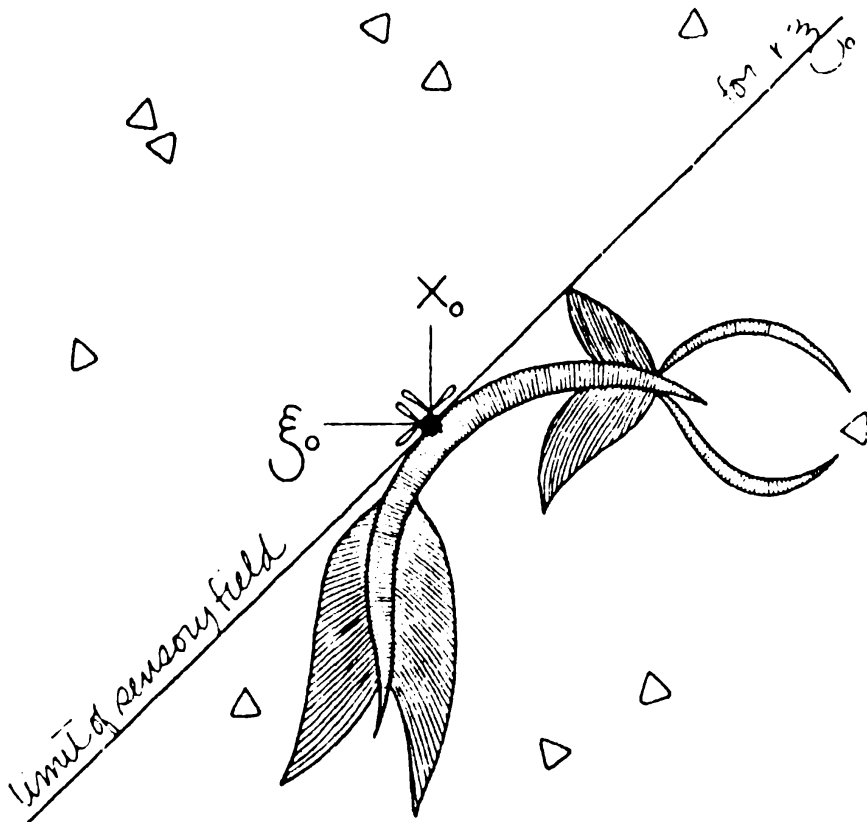
Fig. 6.— Geometry of the sensory field for a specific sensor r_{ξ_0} susceptible to an agent g distributed over environmental space

$$S(\mathbf{x}_{\xi_0}, p)_{\text{position}} = \int_{\substack{\text{sensory} \\ \text{field}}} \Phi(\mathbf{x} - \mathbf{x}_0) g(\mathbf{x}, p) \, d\mathbf{x}$$

and its response is

$$\rho_{\xi_0, \text{position}} = F(S_{\xi_0, \text{position}})$$

(compare with 2.12)

    2.242 Again this expression shows suppression of all spatial clues, save for self reference expressed by the bodily location ξ_0 of the sensation *and* by the position of the organism as expressed by the limits of the integral which can, of course, only be taken over the sensory field "seen" by receptor cell r_{ξ_0} (Fig. 6).

    2.25 Since ρ_ξ gives no clues as to the kind of stimulant (p), it must be either ξ, the place of origin of the sensation, or the operation $\omega(\rho_\xi)$, or both that establish a "sensory modality."

    2.251 In some cases it is possible to compute the spatial distribution of an agent from the known distribution of its effects along a closed surface of given shape. For instance, the spatial distribution of an electrical potential V_x has a unique solution by solving Laplace's equation

$$\Delta V = 0$$

for given values V_ξ along a closed orientable surface (electric fish). Other examples may be cited.

    2.252 In some other cases it is possible to compute the spatial distribution of an agent from its effects on just two small, but distinct regions on the body. For instance, the (Euclidean, 3-D) notion of "depth" is computed by resolving the discrepancy of having the "same scene" represented as different images on the retinas of the two eyes in binocular animals (Fig. 7). Let $L(x,y)$ be a postretinal network which computes the relation "x is left of y." While the right eye reports object "a" to be to the left of "b," ($L_r[a,b]$), the left eye gives the contradictory report of object "b" being to the left of "a," ($L_l[b,a]$). A network B which takes cognizance of the different origin of signals coming from cell groups $\{r_\xi\}_r$ and $\{r_\xi\}_l$ to the right and left side of the animal's body computes with $B(L_r, L_l)$ a new "dimension," namely, the relation $B_s(a,b)$. "a is behind b with respect to *me*" (subscript s = "self").

    2.253 The results of these computations always imply a relation
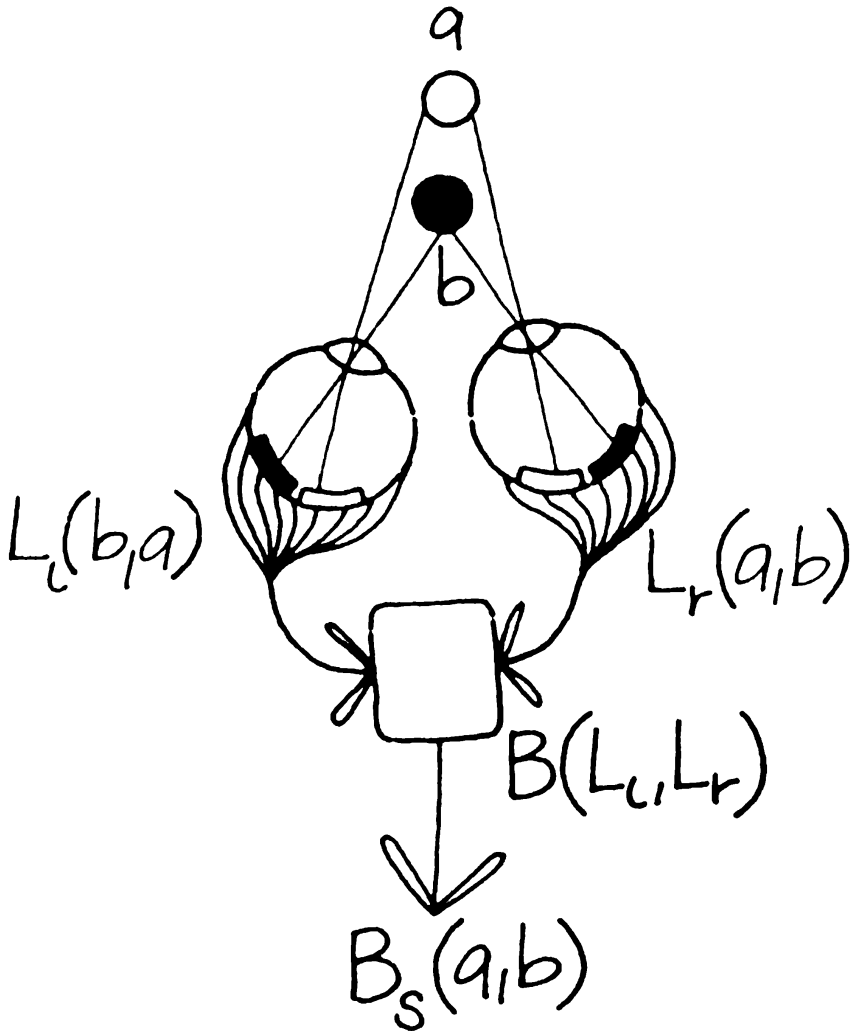
Fic. 7.—Computation of "depth" by resolving a sensory discrepancy in binocular vision; (L.) networks computing the relation "x is left of y"; (B) networks computing the relation "x is behind y"

(geometrical or otherwise) of the organism to its environment, as indicated by the relative notion of "behind"; or to itself, as indicated by the absolute notions of "my left eye" or "my right eye." This is the origin of "self-reference."

2.254 It is clear that the burden of these computations is placed upon the operations ω which compute on the distribution functions ρ_ξ.

2.26 For any of such operations to evolve, it is necessary that

changes in sensation δp_ζ are compared with causes of these changes that are controlled by the organism.

3 In a stationary environment, anisotropic with respect to parameters p_i, a movement $\delta\Gamma$ of the organism causes a change of its sensations δp_ζ. Hence we have

$$(\text{movement}) \longrightarrow (\text{change in sensation})$$

but not necessarily

$$(\text{change in sensation}) \longrightarrow (\text{movement}) .$$

3.1 The terms "change in sensation" and "movement" refer to experiences by the organism. This is evidenced by the notation employed here which describes these affairs p_ζ, μ_ζ, purely in proprietary coordinates ξ and ζ. (Here μ_ζ has been used to indicate the activity of contractile elements m_ζ. Consequently μ_ζ is an equivalent description of $\delta\Gamma$: $\mu_\zeta \longrightarrow \delta\Gamma$.)

3.11 These terms have been introduced to contrast their corresponding notions with those of the terms "stimulus" and "response" of an organism which refer to the experiences of one who *observes* the organism and not to those of the organism itself. This is evidenced by the notation employed here which describes these affairs S_x, $\delta\Gamma_x$ in terms of environmental coordinates x. This is correct insofar as for an observer O the organism Ω is a piece of environment.

3.111 From this it is clear that "stimulus" cannot be equated with "change in sensation" and likewise "response" not with "movement." Although it is conceivable that the complex relations that undoubtedly hold between these notions may eventually be established when more is known of the cognitive processes in both the observer and the organism.

3.112 From the *non sequitur* established under proposition 3, it follows *a fortiori*:

$$\text{not necessarily: (stimulus)} \longrightarrow (\text{response}) .$$

3.2 The presence of a perceptible agent of weak concentration may cause an organism to move toward it (approach). However, the presence of the same agent in strong concentration may cause this organism to move away from it (withdrawal).

3.21 This may be transcribed by the following schema:

$$S_0 \overset{+}{\rightarrow} S_\xi \searrow$$
$$\qquad \qquad \searrow \mu_\xi \longrightarrow \omega(\rho_\xi) \qquad \nearrow \mu_\xi^! \dashrightarrow \delta\Gamma^{++}$$
$$S_0 \overset{-}{\rightarrow} S_\xi \nearrow \qquad \qquad \searrow \mu_\xi \longrightarrow \delta\Gamma^-$$
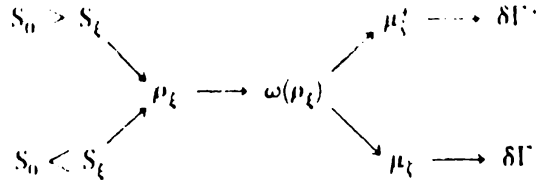
where the (+) and (–) denote approach and withdrawal respectively

    3.211 This schema is the minimal form for representing

        a) "environment" [S]

        b) "internal representation of environment" ($\omega|\rho_\xi|$)

        c) "description of environment" ($\delta\Gamma^+, \delta\Gamma^-$).

4 The logical structure of descriptions arises from the logical structure of movements; "approach" and "withdrawal" are the precursors for "yes" and "no."

    4.1 The two phases of elementary behavior, "approach" and "withdrawal," establish the operational origin of the two fundamental axioms of two-valued logic, namely, the "law of the excluded contradiction": $X \& \bar{X}$ (not: X *and* not-X); and the "law of the excluded middle": $X \vee \bar{X}$ (X *or* not-X); (Fig. 8).

    4.2 We have from Wittgenstein's *Tractatus*,[4] proposition 4.0621:

      it is important that the signs "*p*" and "non *p*" can say the same thing. For it shows that nothing in reality corresponds to the sign "non."

The occurrence of negation in a proposition is not enough to characterize its sense (non-non-*p* = *p*).

    4.21 Since nothing in the environment corresponds to negation, negation as well as all other "logical particles" (inclusion, alternation, implication, etc.) must arise within the organism as a consequence of perceiving the relation of itself with respect to its environment.

    4.3 Beyond being logical affirmative or negative, descriptions can be true or false.

    4.31 We have from Susan Langer, *Philosophy in a New Key*[5]

The use of signs is the very first manifestation of mind. It arises as early in biological history as the famous "conditioned reflex," by which a concomitant of a stimulus takes over the stimulus-function. The concomitant becomes a *sign* of the condition to which the reaction is really appropriate. This is the real beginning of mentality, for here is the birthplace of *error*, and herewith of *truth*.

    4.32 Thus, not only the logical structure of descriptions but also their truth values are coupled to movement.

    4.4 Movement, $\delta\Gamma$, is internally represented through operations on peripheral signals generated by:
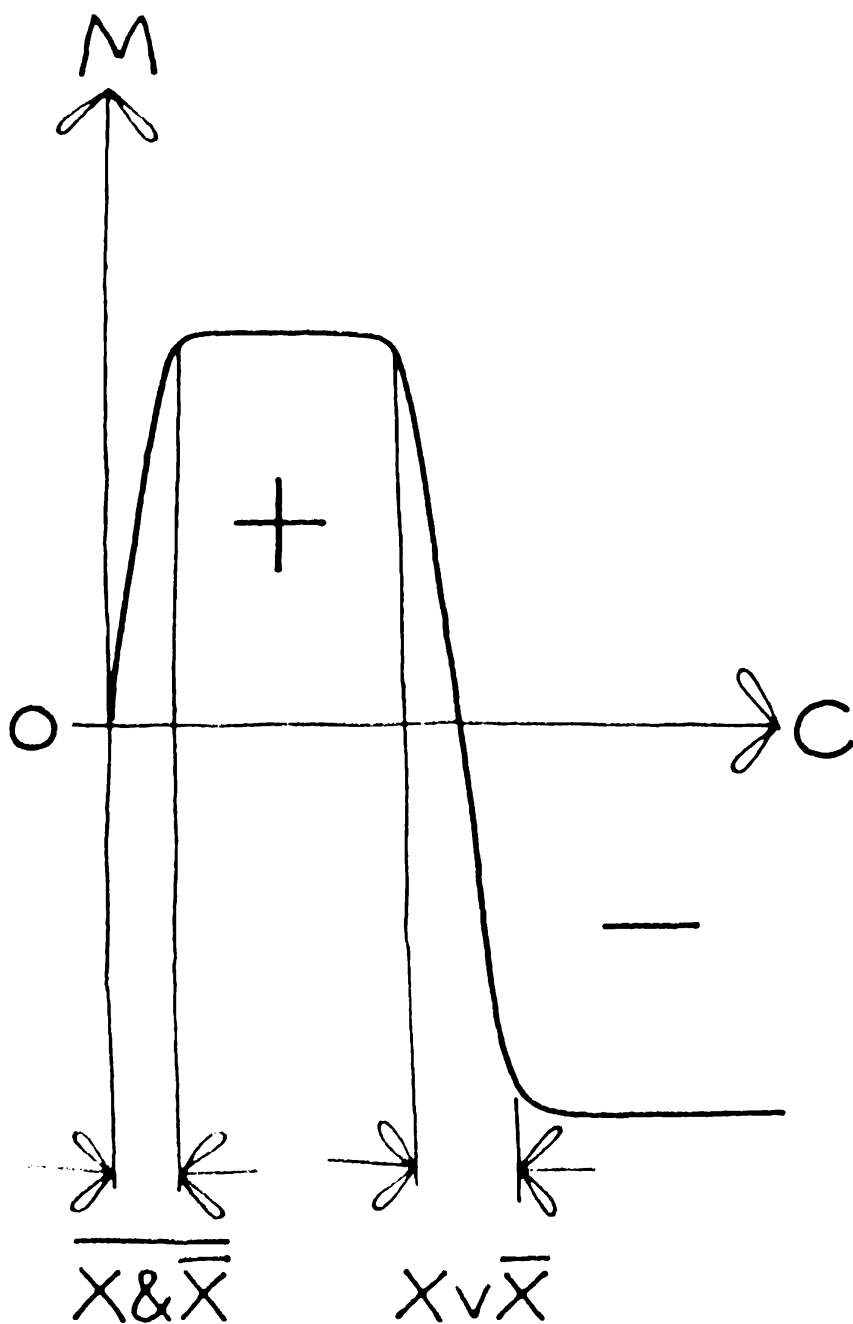
FIG. 8. — The laws of "excluded contradiction" (X & $\bar{X}$) and of "excluded middle" ($X \lor \bar{X}$) in the twilight zones between no motion ($M = 0$) and approach ($+$), and between approach ($+$) and withdrawal ($-$) as a function of the concentration (C) of a perceptible agent.

a) proprioceptors, p_ξ:

$$\delta\Gamma \longrightarrow \pi_\xi \longrightarrow \omega(\pi_\xi),$$

b) sensors, r_ξ:

$$\delta\Gamma \dashrightarrow \delta\rho_\xi \longrightarrow \omega(\rho_\xi);$$

and movement is initiated by operations on the activity ν_ξ of central elements n_ξ.

c) intent:

$$\omega(\nu_\xi) \dashrightarrow \mu_\xi \longrightarrow \delta\Gamma.$$

4.41 Since peripheral activity implies central activity

$$\rho_\xi \longrightarrow \nu_\xi \longleftarrow \pi_\xi$$

we have

$$
\begin{array}{ccc}
\omega(\nu_\xi) & \longrightarrow & \delta\Gamma \\
\uparrow & & \downarrow \\
\delta\Gamma & \longleftarrow & \omega(\nu_\xi).
\end{array}
$$

4.411 From this it is seen that a conceptualization of descriptions of (the internal representation of) the environment arises from the conceptualization of potential movements. This leads to the contemplation of expressions having the form

$$\omega^{(n)}(\delta\Gamma_1, \omega^{(n-1)}[\delta\Gamma_2, \omega^{(n-2)}(\ldots [\rho_\xi])]),$$

that is "descriptions of descriptions of descriptions . . ." or, equivalently, "representations of representations of representations. . . ."

5 The information associated with an event E is the formation of operations ω which control this event's internal representation $\omega(\rho_\xi)$ or its description $\delta\Gamma$.

5.1 A measure of the number of choices of representations ($\omega_i[E]$) or of descriptions ($\delta\Gamma_i[E]$) of this event—or of the probabilities p_i of their occurrence—is the "amount of information" of this event with respect to the organism Ω. ($H[E,\Omega] = -\log_2 p_i$, that is, the negative mean value* of all the $[\log_2 p_i]$)

---

* The mean value of a set of quantities, x_i, whose probability of occurrence of p_i is given by $x_i = \Sigma x_i \cdot p_i$.

5.11 This shows that information is a relative concept. And so is *H*.

5.2 The class of different representations $\omega \equiv (\omega_i[E])$ of an event E determines an equivalence class for different events $(E_i[\omega]) \equiv E$. Hence, a measure of the number of events (E_i) which constitute a cognitive unit, a "category E"[6]—or of the probabilities p_i of their occurrence—is again the "amount of information", *H*, received by an observer upon perceiving the occurrence of one of these events.

5.21 This shows that the amount of information is a number depending on the choice of a category, that is, of a cognitive unit.

5.3 We have from a paper by Jerzy Konorski[7]:

> It is not so, as we would be inclined to think according to our introspection, that the receipt of information and its utilization are two separate processes which can be combined one with the other in any way; on the contrary, information and its utilization are inseparable constituting, as a matter of fact, one single process.

5.31 These processes are the operations ω, and they are implemented in the structural and functional organization of nervous activity.

5.4 Let ν_i be the signals traveling along single fibers, *i*, and $\nu^{(1)}$ be the outcome of an interaction of *N* fibers $(i = 1, 2 \ldots N)$:

$$\nu^{(1)} = F^{(1)}(\nu_1, \nu_2, \ldots \nu_N) \equiv F^{(1)}([\nu_i]) .$$

5.41 It is profitable to consider the activity of a subset of these fibers as determiner for the functional interaction of the remaining ones ("inhibition" changes the functional interaction of "facilitatory" signals). This can be expressed by a formalism that specifies the functions computed on the remaining fibers:

$$\nu^{(1)} = f^{(1)}_{[\nu_i]}([\nu_i]), \qquad j \neq i.$$

The correspondence between the values v of the row vector (ν_i) and the appropriate functions $f_v^{(1)}$ constitutes a functional for the class of functions $f^{(1)}_{[\nu_i]}$.

5.411 This notation makes it clear that the signals themselves may be seen as being, in part, responsible for determining the operations being performed on them.

5.42 The mapping that establishes this correspondence is usually interpreted as being the "structural organization" of these operations, while the set of functions so generated as being their "functional organization."

5.421 This shows that the distinction between structural and

functional organization of cognitive processes depends on the observer's point of view.

5.43 With N fibers being considered, there are 2^N possible interpretations (the set of all subsets of N) of the functional and structural organization of such operations. With all interpretations having the same likelihood, the "uncertainty" of this system regarding its interpretability is $H = \log_2 2^N = N$ bits.

5.5 Let $\nu_i^{(1)}$ be the signals traveling along single fibers, i, and $\nu^{(2)}$ be the outcome of an interaction of N_1 such fibers ($i = 1, 2, \ldots N_1$):

$$\nu^{(2)} = F^{(2)}([\nu_i^{(1)}])$$

or, recursively from 5.4:

$$\nu^{(k)} = F^{(k)}(F^{(k-1)}[F^{(k-2)}(\ldots F^{(1)}[\nu_i])]).$$

5.51 Since the $F^{(k)}$ can be interpreted as functionals $f_{[\nu_i]}^{(k)}$, this leads to a calculus of recursive functionals for the representation of cognitive processes ω.

5.511 This becomes particularly significant if $\nu_i^{(k-t)}$ denotes the activity of fiber, i, at a time interval t prior to its present activity $\nu_i^{(k)}$. That is, the recursion in 5.5 can be interpreted as a recursion in time.

5.52 The formalism of recursive functionals as employed in 5.5 for representing cognitive processes, ω, is isomorphic to the lexical definition structure of nouns. Essentially, a noun signifies a class, $\text{cl}^{(1)}$, of things. When defined, it is shown to be a member of a more inclusive class, $\text{cl}^{(2)}$, denoted by a noun which, in turn, when defined is shown to be a member of a more inclusive class, $\text{cl}^{(3)}$, and so on [pheasant $\longrightarrow$ bird $\longrightarrow$ animal $\longrightarrow$ organism $\longrightarrow$ thing]:

$$\text{cl}^{(n)} = (\text{cl}_{i_{n-1}}^{(n-1)} [\text{cl}_{i_{n-2}}^{(n-2)} (\ldots [\text{cl}_{i_1}^{(1)}])])$$

where the notation (e_i) stands for a class composed of elements e_i, and subscripted subscripts are used to associate these subscripts with the corresponding superscripts.

5.521 The highest order n^* in this hierarchy of classes is always represented by a single, undefined term "thing," "entity," "act," etc., which refers to basic notions of being able to perceive at all.

5.6 Cognitive processes create descriptions of, that is information, about the environment.

6 The environment contains no information. The environment is as it is.

REFERENCES

1. G. Gunther, "Cybernetic Ontology and Transjunctional Operations," *Self-Organizing Systems*, ed. M. C. Yovits *et al.* (Washington, D.C.: Spartan Books, 1962).
2. A. H. Barr, Jr. (ed.), *Fantastic Art, Dada, Surrealism* (3rd. ed.; New York: The Museum of Modern Art, 1947), p. 156.
3. M. Jean, *Histoire de la painture Surrealiste* (Edition du Seuil; Paris: 1959), p. 284.
4. L. Wittgenstein, *Tractatus Logico Philosophicus* (New York: Humanities Press, 1961). ·
5. S. Langer, *Philosophy in a New Key* (New York: New American Library, 1951).
6. H. Quastler, "A Primer in Information Theory," *Information Theory in Biology*, ed. H. P. Yockey *et al.* (New York: Pergamon Press, 1958), pp. 3–49.
7. J. Konorski, "The Role of Central Factors in Differentiation," *Information Processing in the Nervous System*, 3, eds. R. W. Gerard and J. W. Duyff (Amsterdam: Excerpta Medica Foundation, 1962), pp. 318–29.
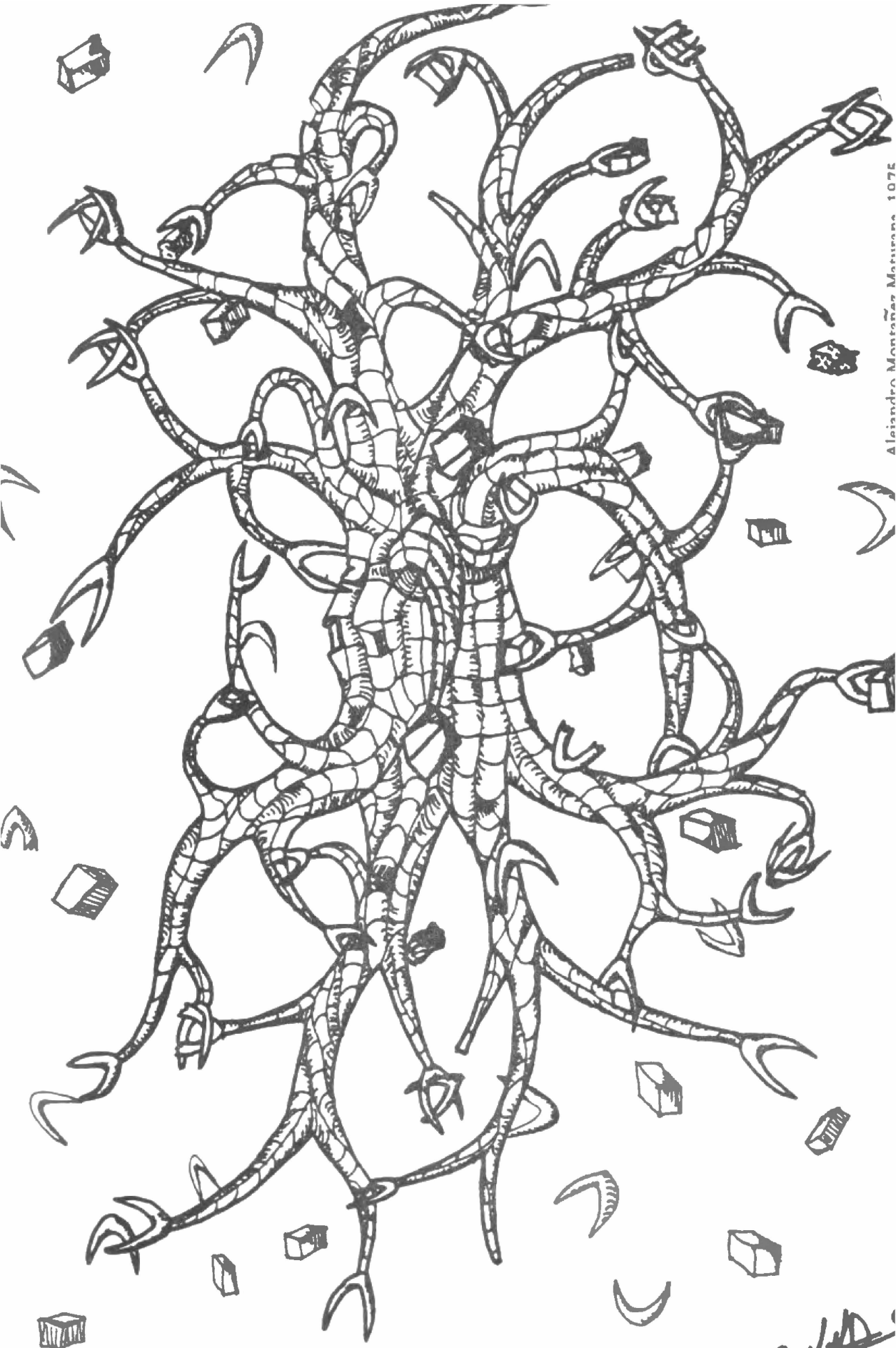
NOTES ON AN EPISTEMOLOGY FOR LIVING THINGS*

I. PROBLEM

While in the first quarter of this century physicists and cosmologists were forced to revise the basic notions that govern the natural sciences, in the last quarter of this century biologists will force a revision of the basic notions that govern science itself. After that "first revolution" it was clear that the classical concept of an "ultimate science," that is an objective description of the world in which there are no subjects (a "subjectless universe"), contains contradictions.

To remove these one had to account for an "observer" (that is at least for one subject): (i) Observations are not absolute but relative to an observer's point of view (i.e., his coordinate system: Einstein); (ii) Observations affect the observed so as to obliterate the observer's hope for prediction (i.e., his uncertainty is absolute: Heisenberg).

After this, we are now in the possession of the truism that a description (of the universe) implies one who describes (observes it). What we need now is the description of the "describer" or, in other words, we need a theory of the observer. Since it is only living organisms which would qualify as being observers, it appears that this task falls to the biologist. But he himself is a living being, which means that in his theory he has not only to account for himself, but also for his writing this theory. This is a new state of affairs in scientific discourse for, in line with the traditional viewpoint which separates the observer from his observations, reference to this discourse was to be carefully avoided. This separation was done by no means because of excentricity or folly, for under certain circumstances inclusion of the observer in his descriptions may lead to paradoxes, to wit the utterance "I am a liar."

In the meantime however, it has become clear that this narrow restriction not only creates the ethical problems associated with scientific activity, but also cripples the study of life in full context from molecular to social organizations. Life cannot be studied *in vitro*, one has to explore it *in vivo*.

In contradistinction to the classical problem of scientific inquiry that pos-

tulates first a description-invariant "objective world" (as if there were such a thing) and then attempts to write its description, now we are challenged to develop a description-invariant "subjective world," that is a world which includes the observer: *This is the problem.*

However, in accord with the classic tradition of scientific inquiry which perpetually asks "How?" rather than "What?," this task calls for an epistemology of "How do we know?" rather than "What do we know?"

The following notes on an epistemology of living things address themselves to the "How?" They may serve as a magnifying glass through which this problem becomes better visible.

II.   INTRODUCTION

The twelve propositions labeled 1, 2, 3, . . . 12, of the following 80 Notes are intended to give a minimal framework for the context within which the various concepts that will be discussed are to acquire their meaning. Since Proposition Number 12 refers directly back to Number 1, Notes can be read in a circle. However, comments, justifications, and explanations, which apply to these propositions follow them with decimal labels (e.g., "5.423") the last digit ("3") referring to a proposition labeled with digits before the last digit ("5.42"), etc. (e.g., "5.42" refers to "5.4," etc.).

Although Notes may be entered at any place, and completed by going through the circle, it appeared advisable to cut the circle between propositions "11" and "1," and present the notes in linear sequence beginning with Proposition 1.

Since the formalism that will be used may for some appear to obscure more than it reveals, a preview of the twelve propositions* with comments in prose may facilitate reading the notes.

1.   The environment is experienced as the residence of objects, stationary, in motion, or changing. **
Obvious as this proposition may look at first glance, on second thought one may wonder about the meaning of a "changing object." Do we mean the change of appearance of the same object as when a cube is rotated, or a person turns around, and we take it to be the same object (cube, person, etc.); or when we see a tree growing, or meet an old schoolmate after a decade or two, are they

---

*In somewhat modified form .

**Propositions appear in italics.

different, are they the same, or are they different in one way and the same in another? Or when Circe changes men into beasts, or when a friend suffers a severe stroke, in these metamorphoses, what is invariant, what does change? Who says that these were the same persons or objects?

From studies by Piaget [1] and others [2] we know that "object constancy" is one of many cognitive skills that are acquired in early childhood and hence are subject to linguistic and thus cultural bias.

Consequently, in order to make sense of terms like "biological invariants," "cultural universals," etc., the logical properties of "invariance" and "change" have first to be established.

As the notes procede it will become apparent that these properties are those of descriptions (representations) rather than those of objects. In fact, as will be seen, "objects" do owe their existence to the properties of representations.

To this end the next four propositions are developed.

2.   *The logical properties of "invariance" and "change" are those of representations. If this is ignored, paradoxes arise.*
Two paradoxes that arise when the concepts "invariance" and "change" are defined in a contextual vacuum are cited, indicating the need for a formalization of representations.

3.   *Formalize representations R, S, regarding two sets of variables $\{x\}$ and $\{t\}$, tentatively called "entities" and "instants" respectively.*
Here the difficulty of beginning to talk about something which only later makes sense so that one can begin talking about it, is pre-empted by "tentatively," giving two sets of as yet undefined variables highly meaningful names, viz, "entities" and "instants," which only later will be justified.

This apparent deviation from rigor has been made as a concession to lucidity. Striking the meaningful labels from these variables does not change the argument.

Developed under this proposition are expressions for representations that can be compared. This circumvents the apparent difficulty to compare an apple with itself before and after it is peeled. However, little difficulties are encountered by comparing the peeled apple as it is *seen now* with the unpeeled apple as it is *remembered* to have been before.

With the concept "comparison," however an operation ("computation") on representations is introduced, which requires a more detailed analysis. This is done in the next proposition. From here on the term "computation" will be consistently applied to all operations (not necessarily numerical) that transform, modify, re-arrange, order, etc., either symbols (in the "abstract" sense) or their physical manifestations (in the "concrete" sense). This is done to enforce a feeling for the realizability of these operations in the structural and functional organization of either grown nervous tissue or else constructed machines.

4.   Contemplate relations, "Rel," between representations, R, and S.
However, immediately a highly specific relation is considered, viz, an "Equivalence Relation" between two representations. Due to the structural properties of representations, the computations necessary to confirm or deny equivalence of representations are not trivial. In fact, by keeping track of the computational pathways for establishing equivalence, "objects" and "events" emerge as *consequences* of branches of computation which are identified as the processes of abstraction and memorization.

5.   Objects and events are not primitive experiences. Objects and events are representations of relations.
Since "objects" and "events" are not primary experiences and thus cannot claim to have absolute (objective) status, their interrelations, the "environment," is a purely personal affair, whose constraints are anatomical or cultural factors. Moreover, the postulate of an "external (objective) reality" disappears to give way to a reality that is determined by modes of internal computations [3].

6.   Operationally, the computation of a specific relation is a representation of this relation.
Two steps of crucial importance to the whole argument forwarded in these notes are made here at the same time. One is to take a computation for a representation; the second is to introduce here for the first time"recursions." By recursion is meant that on one occasion or another a function is substituted for its own argument. In the above Proposition 6 this is provided for by taking the computation of a relation between *representations* again as a representation.

While taking a computation for a representation of a relation may not cause conceptual difficulties (the punched card of a computer program which controls the calculations of a desired relation may serve as a adequate metaphor), the adoption of recursive expressions appears to open the door for all kinds of logical mischief.

However, there are means to avoid such pitfalls. One, e.g., is to devise a notation that keeps track of the order of representations, e.g., "the represen-

tation of a representation of a representation'' may be considered as a third order representation, $R^{(3)}$. The same applies to relations of higher order, n: $Rel^{(n)}$.

After the concepts of higher order representations and relations have been introduced, their physical manifestations are defined. Since representation and relations are computations, their manifestations are "special purpose computers" called "representors" and "relators" respectively. The distinction of levels of computation is maintained by referring to such structures as n-th order representors (relators). With these concepts the possibility of introducing "organism" is now open.

7.   A living organism is a third order relator which computes the relations that maintain the organism's integrity.
The full force of recursive expressions is now applied to a recursive definition of living organisms first proposed by H. R. Maturana [4] [5] and further developed by him and F. Varela in their concept of "autopoiesis" [6].

As a direct consequence of the formalism and the concepts which were developed in earlier propositions it is now possible to account for an interaction between the internal representation of an organism of himself with one of another organism. This gives rise to a theory of communication based on a purely connotative "language." The surprising property of such a theory is now described in the eighth proposition.

8.   A formalism necessary and sufficient for a theory of communication must not contain primary symbols representing communicabilia (e.g., symbols, words, messages, etc.).
Outrageous as this proposition may look at first glance, on second thought however it may appear obvious that a theory of communication is guilty of circular definitions if it assumes communicabilia in order to prove communication.

The calculus of recursive expressions circumvents this difficulty, and the power of such expressions is exemplified by the (indefinitely recursive) reflexive personal pronoun "I." Of course the semantic magic of such infinite recursions has been known for some time, to wit the utterance "I am who I am" [7].

9.   Terminal representations (descriptions) made by an organism are manifest in its movements; consequently the logical structure of descriptions arises from the logical structure of movements.
The two fundamental aspects of the logical structure of descriptions, namely

their sense (affirmation or negation), and their truth value (true or false), are shown to reside in the logical structure of movement: approach and withdrawal regarding the former aspect, and functioning or dysfunctioning of the conditioned reflex regarding the latter.

It is now possible to develop an exact definition for the concept of "information" associated with an utterance. "Information" is a relative concept that assumes meaning only when related to the cognitive structure of the observer of this utterance (the "recipient").

10. The information associated with a description depends on an observer's ability to draw inferences from this description.
Classical logic distinguishes two forms of inference: deductive and inductive [8]. While it is in principle possible to make infallible deductive inferences ("necessity"), it is in principle impossible to make infallible inductive inferences (" chance"). Consequently, chance and necessity are concepts that do not apply to the world, but to our attempts to create (a description of) it.

11. The environment contains no information; the environment is as it is.

12. Go back to Proposition Number 1.

REFERENCES

1.   Piaget, J.: The Construction of Reality in the Child. Basic Books, New York, (1954).

2.   Witz, K. and J. Easley: *Cognitive Deep Structure and Science Education* in Final Report, Analysis of Cignitive Behavior in Children; Curriculum Laboratory, University of Illinois, Urbana, (1972).

3.   Castaneda, C.: A Separate Reality. Simon and Schuster, New York, (1971).

4.   Maturana, H.: *Neurophysiology of Cognition* in Cognition: A Multiple View, P. Garvin (ed.), Spartan Books, New York, pp. 3-23, (1970).

5.   Maturana, H.: Biology of Cognition, BCL Report No. 9.0, Biological Computer Laboratory, Department of Electrical Engineering, University of Illinois, Urbana, 95 pp., (1970).

6.   Maturana, H. and F. Varela: Autopoiesis. Faculdad de Ciencias, Universidad de Chile, Santiago, (1972).

7.   Exodus, 3, 14.

8.   Aristotle: Metaphysica. Volume VIII of The Works of Aristotle, W.D. Ross (ed., tr.), The Clarendon Press, Oxford, (1908).

II.   NOTES

1.   The environment is experienced as the residence of objects, stationary, in motion, or changing.
1.1   "Change" presupposes invariance, and "invariance" change.

2.   The logical properties of "invariance" and "change" are those of representations. If this is ignored paradoxes arise.
2.1   The paradox of "invariance:"
         THE DISTINCT BEING THE SAME
      But it makes no sense to write $x_1 = x_2$ (why the indices?).
      And $x = x$ says something about "=" but nothing about x.
2.2   The paradox of "change:"
         THE SAME BEING DISTINCT
      But it makes no sense to write $x \neq x$.

3.   Formalize the representations R, S, . . . regarding two sets of variables x_i and t_j (i, j = 1, 2, 3. . .), tentatively called "entities" and "instants" respectively.
3.1   The representation R of an entity x regarding the instant t_1 is distinct from the representation of this entity regarding the instant t_2:
         $$R(x(t_1)) \neq R(x(t_2))$$
3.2   The representation S of an instant t regarding the entity x_1 is distinct from the representation of this instant regarding the entity x_2:
         $$S(t(x_1)) \neq S(t(x_2))$$
3.3   However, the comparative judgment ("distinct from") cannot be made without a mechanism that computes these distinctions.
3.4   Abbreviate the notation by
         $$R(x_i(t_j)) \rightarrow R_{ij}$$
         $$S(t_k(x_l)) \rightarrow S_{kl}$$
(i, j, k, l = 1, 2, 3, . . .)

4.   Contemplate relations Rel_μ between the representations R and S:
         $$Rel_\mu(R_{ij}, S_{kl})$$
(μ = 1, 2, 3, . . .)
4.1   Call the relation which obliterates the distinction $x_i \neq x_l$ and $t_j \neq t_k$ (i.e., i = l; j = k) the "Equivalence Relation" and let it be represented by:
         $$Equ(R_{ij}, S_{ji})$$
4.11 This is a representation of a relation between two representations and reads:
      "The representation R of an entity x_i retarding the instant t_j is equivalent to the representation S of an instant t_j regarding the entity x_i."

4.12 A possible linguistic metaphor for the above representation of the equi-valence relation between two representations is the equivalence of "thing acting" (most Indo-European languages) with "act thinging" (some African languages) (cognitive duality). For instance:

"The horse gallops" $\rightleftharpoons$ "The gallop horses"

4.2 The computation of the equivalence relation 4.1 has two branches:

4.21 One computes equivalences for x only

$$Equ(R_{ij}, S_{ki}) = Obj(x_i)$$

4.211 The computations along this branch of equivalence relation are called "abstractions:" *Abs.*

4.212 The results of this branch of computation are usually called "objects" (entities), and their invariance under various transformations $(t_j, t_k, \ldots)$ is in-dicated by giving each object a distinct but invariant label N_i ("Name"):

$$Obj(x_i) \rightarrow N_i$$

4.22 The other branch computes equivalences for t only:

$$Equ(R_{ij}, S_{jl}) \equiv Eve(t_j)$$

4.221 The computations along this branch of equivalence relation are called "memory:" *Mem.*

4.222 The results of this branch of computation are usually called "events" (instants), and their invariance under various transformations $(x_i, x_h, \ldots)$ is in-dicated by associating with each event a distinct but invariant label T_j ("Time"):

$$Eve(t_j) \rightarrow T_j$$

4.3 This shows that the concepts "object," "event," "name," "time," "ab-straction," "memory," "invariance," "change," generate each other.

From this follows the next proposition:

5.   *Objects and events are not primitive experiences. "Objects" and "Events" are representations of relations.*

5.1 A possible graphic metaphor for the complementarity of "object" and "event" is an orthogonal grid that is mutually supported by both (Fig. 1).
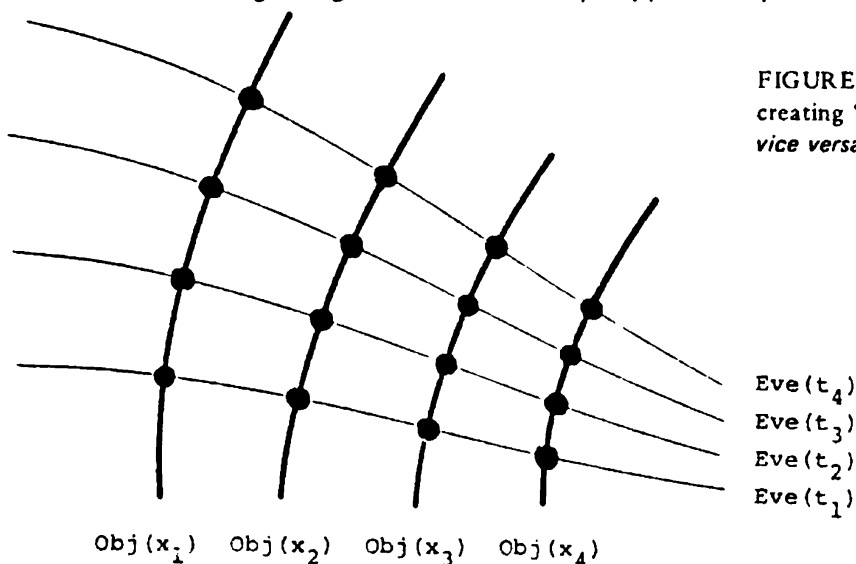
FIGURE 1. "Objects" creating "Events" and *vice versa.*

Eve(t_4)
Eve(t_3)
Eve(t_2)
Eve(t_1)

Obj(x_1) Obj(x_2) Obj(x_3) Obj(x_4)

5.2   "Environment" is the representation of relations between "objects" and "events"

$$Env(\text{Obj, Eve})$$

5.3   Since the computation of equivalence relations is not unique, the results of these computations, namely, "objects" and "events" are likewise not unique.

5.31   This explains the possibility of an arbitrary number of different, but internally consistent (language determined) taxonomies.

5.32   This explains the possibility of an arbitrary number of different, but internally consistent (culturally determined) realities.

5.4   Since the computation of equivalence relations is performed on primitive experiences, an external environment is not a necessary prerequisite of the computation of a reality.

6.   *Operationally, the computation Cmp*(Rel) *of a specific relation is a representation of this relation.*

$$R = Cmp(\text{Rel})$$

6.1   A possible mathematical metaphor for the equivalence of a computation with a representation is, for instance, Wallis' computational algorithm for the infinite product:

$$2 \cdot \frac{2}{1} \cdot \frac{2}{3} \cdot \frac{4}{3} \cdot \frac{4}{5} \cdot \frac{6}{5} \cdot \frac{6}{7} \cdot \cdot \cdot \cdot$$

Since this is one of many possible definitions of π (3.14159...), and π is a number, we may take π as a (numerical) representation of this computation.

6.2   Call representations of computations of relations "second order representations." This is clear when such a representation is written out fully:

$$R = Cmp(\text{Rel}(R_{ij}, S_{kl})),$$

where R_{ij} and S_{kl} are, of course, "first order representations" as before (3.3).

6.21   From this notation it is clear that first order representations can be interpreted as zero-order relations (note the double indices on S and R).

6.22   From this notation it is also clear that higher order (n-th order) representations and relations can be formulated.

6.3   Call a physical mechanism that computes an n-th order representation (or an n-th order relation) an "n-th order representor" $RP^{(n)}$ (or "n-th order relator" $RL^{(n)}$) respectively.

6.4   Call the externalized physical manifestation of the result of a computation a "terminal representation" or a "description."

6.5   One possible mechanical metaphor for relator, relation, objects, and descriptions, is a mechanical desk calculator (the relator) whose internal structure (the arrangement of wheels and pegs) is a representation of a relation commonly called "addition:" Add (a, b; c). Given two objects, a = 5, b = 7, it computes a terminal representation (a description), c, of the relation between these two objects in digital, decadic, form:

$$\text{Add } (5, 7; 12)$$

6.51 Of course, a machine with a different internal representation (structure) of the same relation Add (a, b; c), may have produced a different terminal representation (description), say, in the form of prime products, of this relation between the same objects:

$$\text{Add } (5, 7; 2^2 \cdot 3^1)$$

6.6. Another possible mechanical metaphor for taking a computation of a relation as a representation of this relation is an electronic computer and its program. The program stands for the particular relation, and it assembles the parts of the machine such that the terminal representation (print-out) of the problem under consideration complies with the desired form.

6.61 A program that computes programs is called a "meta-program." In this terminology a machine accepting meta-programs is a second-order relator.

6.7 These metaphors stress a point made earlier (5.3), namely, that the computations of representations of objects and events is not unique.

6.8 These metaphors also suggest that my nervous tissue which, for instance, computes a terminal representation in the form of the following utterance: "These are my grandmother's spectacles" neither resembles my grandmother nor her spectacles; nor is there a "trace" to be found of either (as little as there are traces of "12" in the wheels and pegs of a desk calculator, or of numbers in a program). Moreover, my utterance "These are my grandmother's spectacles" should neither be confused with my grandmother's spectacles, nor with the program that computes this utterance, nor with the representation (physical manifestation) of this program.

6.81 However, a relation between the utterance, the objects, and the algorithms computing both, is computable (see 9.4).

7.   *A living organism Ω is a third-order relator ($\Omega = RL^{(3)}$) which computes the relations that maintain the organism's integrity [1] [2]:*

$$\Omega \text{ Equ } [R(\Omega(\text{Obj})), S(\text{Eve}(\Omega))]$$

This expression is recursive in Ω .

7.1   An organism is its own ultimate object.

7.2   An organism that can compute a representation of this relation is self-conscious.

7.3   Amongst the internal representations of the computation of objects Obj(x_j) within one organism Ω may be a representation Obj(Ω^*) of another organism Ω^*. Conversely, we may have in Ω^* a representation Obj*(Ω) which computes Ω .

7.31   Both representations are recursive in Ω, Ω^* respectively. For instance, for Ω :

$$\text{Obj}^{(n)} \ (\Omega^{*(n-1)}(\text{Obj}^{*(n-1)}(\Omega^{(n-2)}(\text{Obj}^{(n-2)}(\ldots\Omega^*)))))).$$

7.32   This expression is the nucleus of a theory of communication.

8.   *A formalism necessary and sufficient for a theory of communication must not contain primary symbols representing "communicabilia" (e.g., symbols, words, messages, etc.).*

8.1 This is so, for if a "theory" of communication were to contain primary communicabilia, it would not be a theory but a technology of communication, taking communication for granted.

8.2 The nervous activity of one organism cannot be shared by another organism.

8.21 This suggests that indeed nothing is (can be) "communicated."

8.3 Since the expression in 7.31 may become cyclic (when $Obj^{(k)} = Obj^{(k-2i)}$), it is suggestive to develop a teleological theory of communication in which the stipulated goal is to keep $Obj(\Omega^*)$ invariant under perturbations by Ω^*.

8.31 It is clear that in such a theory such questions as: 'Do you see the color of this object as I see it?" become irrelevant.

8.4 Communication is an observer's interpretation of the interaction between two organisms Ω_1, Ω_2.

8.41 Let $Evs_1 \equiv Evs(\Omega_1)$, and $Evs_2 \equiv Evs(\Omega_2)$, be *sequences* of events $Eve(t_j)$, ($j = 1, 2, 3, \ldots$) with regard to two organisms Ω_1 and Ω_2 respectively; and let *Com* be an observer's (internal) representation of a relation between these sequences of events:

$$OB(Com(Evs_1, Evs_2))$$

8.42 Since either Ω_1 or Ω_2 or both can be observers ($\Omega_1 = OB_1$; $\Omega_2 = OB_2$) the above expression can become recursive in either Ω_1 or in Ω_2 or in both.

8.43 This shows that "communication" is an (internal) representation of a relation between (an internal representation of) oneself with somebody else.

$$R(\Omega^{(n+1)}, Com(\Omega^{(n)}, \Omega^*))$$

8.44 Abbreviate this by

$$C(\Omega^{(n)}, \Omega^*).$$

8.45 In this formalism the reflexive personal pronoun "I" appears as the indefinitely applied) recursive operator

$$Equ[\Omega^{(n+1)} C(\Omega^{(n)}, \Omega^{(n)})]$$

or in words:

> "I am the observed relation between myself and observing myself."

8.46 "I" is a relator (*and* representor) of infinite order.

9. *Terminal representations (descriptions) made by an organism are manifest in its movements; consequently, the logical structure of descriptions arises from the logical structure of movements.*

9.1 It is known that the presence of a perceptible agent of weak concentration may cause an organism to move toward it (approach). However, the presence of the same agent in strong concentration may cause this organism to move away from it (withdrawal).

9.11 That is "approach" and "withdrawal" are the precursors for "yes" or "no."

9.12 The two phases of elementary behavior, "approach" and "withdrawal," establish the operational origin of the two fundamental axioms of two-valued logic, namely, the "law of the excluded contradiction:"

$$\overline{x \ \& \ \bar{x}},$$

in words: "not: x *and* not-x;"
and the law of the excluded middle:

$$x \vee \bar{x},$$

in words: "x *or* not-x;" (see Fig. 2).
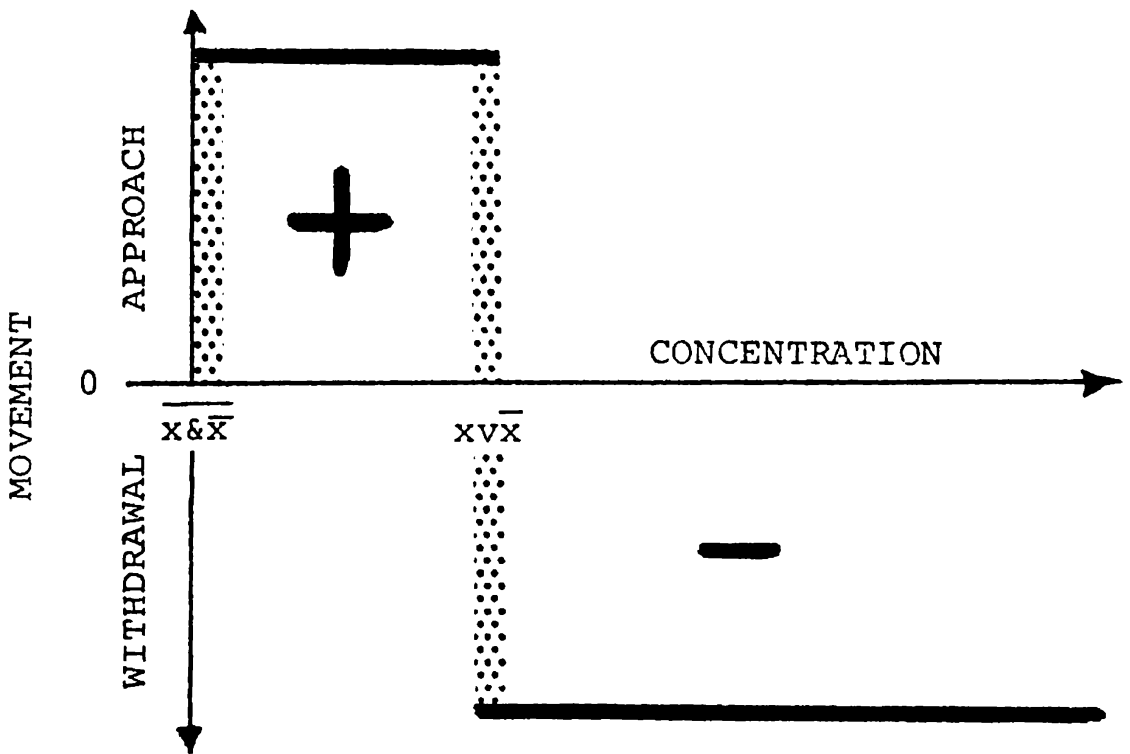
FIGURE 2. The laws of "excluded contradiction" $(\overline{x \ \& \ \bar{x}})$ and of "excluded middle" $(x \vee \bar{x})$ in the twilight zones between no motion (M = 0) and approach (+), and between approach (+) and withdrawal (−) as a function of the concentration (C) of a perceptible agent.

9.2 We have from Wittgenstein's Tractatus [3], proposition 6.0621: ". . . it is important that the signs "p" and "non-p" *can* say the same thing. For it shows that nothing in reality corresponds to the sign "non."
The occurence of negation is a proposition is not enough to characterize its sense (non-non-p = p)."

9.21 Since nothing in the environment corresponds to negation, negation as well as all other "logical particles" (inclusion, alternation, implication, etc.) must arise within the organism itself.

9.3 Beyond being logical affirmative or negative, descriptions can be true or false.

9.31 We have from Susan Langer, Philosophy in a New Key [4] :

"The use of signs in the very first-manifestation of mind. It arises as early in biological history as the famous 'conditioned reflex,' by which a concomitant of a stimulus takes over the stimulus-function. The concomitant becomes a *sign* of the condition to which the reaction is really appropriate. This is the real beginning of mentality, for here is the birthplace of *error*, and herewith of *truth*."

9.32 Thus, not only the sense (yes or no) of descriptions but also their truth values (true or false) are coupled to movement (behavior).

9.4 Let D* be the terminal representation made by an organism Ω^*, and let it be observed by an organism Ω; let Ω's internal representation of this description be $D(\Omega,D^*)$; and, finally, let Ω's internal representation of his environment be $E(\Omega, E)$. Then we have:

The domain of relations between D and E which are computable by Ω represents the "information" gained by Ω from watching Ω^*:

$$Inf(\Omega,D^*) \equiv \text{Domain } Rel_\mu(D,E)$$

$(\mu = 1, 2, 3. . .m)$

9.41 The logarithm (of base 2) of the number m of relations Rel_μ computable by Ω (or the negative mean value of the logarithmitic probabilities of their occurance $\langle \log_2 p_i \rangle = \Sigma p_i \log_2 p_i$; i = 1 → m) is the "amount of information, H" of the description D* with respect to Ω :

$$H(D^*, \Omega) = \log_2 m$$
$$m$$

$$(\text{or } H(D^*, \Omega) = -\Sigma p_i \log_2 p_i)$$
$$1$$

9.42 This shows that information is a relative concept. And so is H.

9.5 We have from a paper by Jerzy Konorski [5] :

". . .It is not so, as we would be inclined to think according to our introspection, that the receipt of information and its utilization are two separate processes which can be combined one with the other in any way; on the contrary, information and its utilization are inseparable constituting, as a matter of fact, one single process."

10. *The information associated with a description depends on an observer's ability to draw inferences from this description.*

10.1 "Necessity" arises from the ability to make infallible deductions.

10.2 "Chance" arises from the inability to make infallible inductions.

11. *The environment contains no information. The environment is as it is.*

12. *The environment is experienced as the residence of objects, stationary, in motion, or changing (Proposition 1).*

REFERENCES

1.   Maturana, H.: *Neurophysiology of Cognition* in Cognition: A Multiple View, P. Garvin (ed.), Spartan Books, New York, pp. 3-23, (1970).

2.   Maturana, H.: Biology of Cognition, BCL Report No. 9.0, Biological Computer Laboratory, Department of Electrical Engineering, University of Illinois, Urbana, 95 pp., (1970).

3.   Wittgenstein, L.: Tractatus Logico Philosophicus, Humanities Press, New York, (1961).

4.   Langer, S.: Philosophy in a New Key, New American Library, New York, (1951).

5.   Konorski, J.: *The Role of Central Factors in Differentiation* in Information Processing in the Nervous System, R.W. Gerard and J.W. Duyff (eds.), Excerpta Medica Foundation, Amsterdam, 3, pp. 318-329, (1962).

OBJECTS: TOKENS FOR (EIGEN-)BEHAVIORS*

A seed, alas, not yet a flower, for Jean Piaget to his 80th birthday from Heinz von Foerster with admiration and affection.

I shall talk about notions that emerge when the organization of sensori-motor interactions (and also that of central processes (cortical-cerebellar-spinal, cortico-thalamic-spinal, etc.)) is seen as being essentially of circular (or more precisely of recursive) nature. Recursion enters these considerations whenever the changes in a creature's sensations are accounted for by its movements ($s_i = S(m_k)$), and its movements by its sensations ($m_k = M(s_j)$). When these two accounts are taken together, then they form "recursive expressions," that is, expressions that determine the states (movements, sensations) of the system (the creature) in terms of these very states ($s_i = S(M(s_j)) = SM(s_j)$; $m_k = M(S(m_l)) = MS(m_l)$).

One point that with more time, effort and space could be made rigorously and not only suggestively as it has been made here, is that what is referred to as "objects" (GEGEN-STAENDE = "against-standers") in an observer-excluded (linear, open) epistemology, appears in an observer-included (circular, closed) epistemology as "tokens for stable behaviors" (or, if the terminology of Recursive Function Theory is used, as "tokens for Eigen-functions").

Of the many possible entries into this topic the most appropriate one for this occasion appears to me the (recursive) expression that forms the last line on page 63 of J. Piaget's *L'Equilibration des Structures Cognitives* (1975):

$$\text{Obs.O} \rightarrow \text{Obs.S} \rightarrow \text{Coord.S} \rightarrow \text{Coord.O} \rightarrow \text{Obs.O} \rightarrow \text{etc.}$$

This is an observer's account of an interaction between a subject S and an object (or a set of objects) O. The symbols used in this expression (defined on page 59 *op. cit.*) stand for (see also Fig. 1):

Obs.S	"observables relatifs a l'action du sujet"
Obs.O	"observables relatifs aux objets"

Coord.S   "coordinations inferentielles des actions (ou operations) du sujet"

Coord.O   "coordinations inferentielles entre objets"

"etc."   "the (syntactic) injunction to iterate (with no limits specified) the sequence of these operations (HVF)"
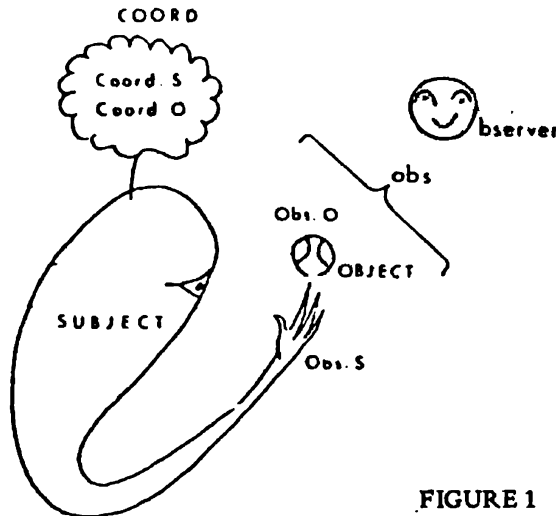
FIGURE 1

For the sake of brevity (lucidity?) I propose to compress the symbolism of before even further, compounding all that is observed (i.e. Obs.O and Obs.S) into a single variable

$$obs,$$

and compounding coordinating operations that are performed by the subject (i.e. Coord.S and Coord.O) into a single operator

$$COORD.$$

COORD transforms, rearranges, modifies etc., the forms, arrangements, behaviors, etc., observed at one occasion (say, initially obs_0, and call it the "primary argument") into those observed at the next occasion, obs_1. Express the outcome of this operation through the equality: [1]

---

[1] By replacing the arrow "→", whose operational meaning is essentially to indicate a one-way (semantic) connectedness (e.g., "goes to," "implies," "invokes," "leads to," etc.) between adjacent expressions, with the equality sign provides the basis for a calculus. However, in order that legitimate use of this sign can be made, the variables "obs_1" must belong to the same domain. The choice of domain is, of course left to the observer who may wish to express his observations in form of, for instance, numerical values, of vectors representing arrangements or geometrical configurations, or his observations of behaviors in form of mathematical functions (e.g., "equations of motion," etc.), or by logical propositions (e.g., McCulloch-Pitts' "TPE's" 1943 (i.e., Temporal Propositional Expressions), etc.).

$$obs_1 = COORD(obs_0).$$

While some relational fine structure is (clearly) lost in this compression, gained, however, may be an easiness by which the progression of events, suggested on the last lined page of 62 *op. cit.* and copied here can now be watched.

Obs. S(n) ⟶ Coord. S(n) ⟷ Obs. O(n) ⟵ Coord. O(n)

Obs. S(n+1) ⟶ Coord. S(n+1) ⟶ Obs. O(n+1) ⟶ Coord. O(n+1)

Obs. S(n+2) ⟶ Coord. S(n+2) ⟷ Obs. O(n+2) ⟵ Coord. O(n+2)

etc.   etc.
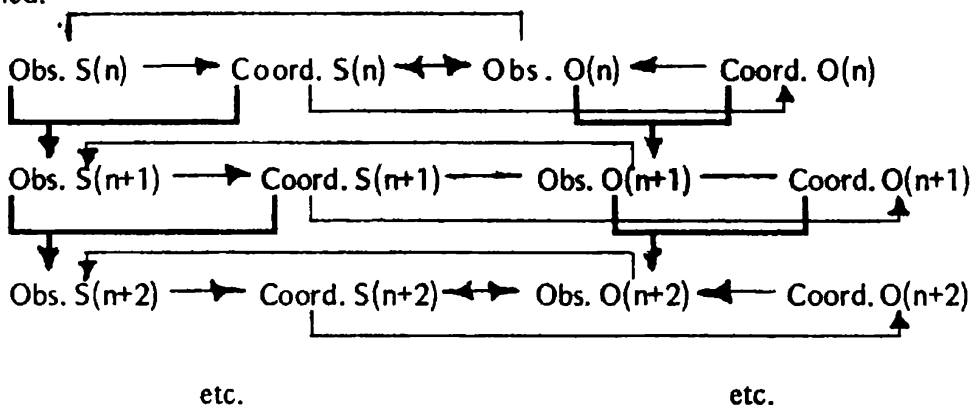
Allow the operator COORD to operate on the previous outcome to give

$$obs_2 = COORD(obs_1) = COORD(COORD(obs_0)) \qquad (2)$$

and (recursively) after n steps

$$obs_n = COORD(COORD(COORD(\ldots\ldots\ldots (obs_0)\ )\ )..), \qquad n \text{ times} \qquad (3)$$

n times

or by notational abbreviation

$$obs_n = COORD^{(n)} (obs_0). \qquad (4)$$

By this notational abbreviation it is suggested that also functionally

n times

can be replaced by

$$obs_o \text{———} \boxed{COORD^{(n)}} \text{———} obs_n$$

* * * * *
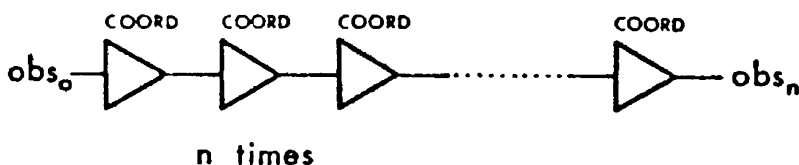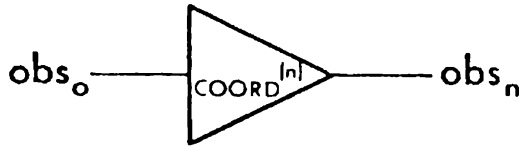
Let n grow without limit $(n \to \infty)$:

$$obs_\infty = \lim_{n \to \infty} COORD^{(n)}(obs_o) \tag{5}$$

or:

$$obs_\infty = COORD(COORD(COORD(COORD. \ldots \tag{6}$$

Contemplate the above expression (6) and note:

(i)   that the independent variable obs_o, the "primary argument" has dis-appeared (which may be taken as a signal that the simple connection between independent and dependent variables is lost in indefinite recursions, and that such expressions take on a different meaning).

(ii)   that, because obs_∞ expresses an indefinite recursion of operators COORD onto operators COORD, any indefinite recursion within that expression can be replaced by obs_∞:

$$obs_\infty =$$

COORD(COORD(COORD(COORD(. . . . .
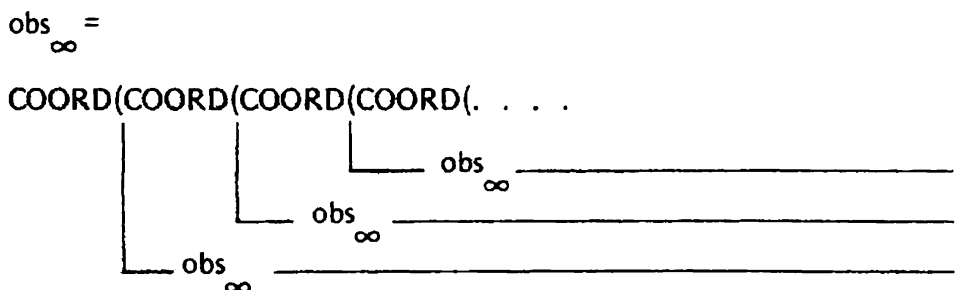
(iii)   Hence:

$$obs_\infty = obs_\infty \tag{7.0}$$
$$obs_\infty = COORD(obs_\infty) \tag{7.1}$$
$$obs_\infty = COORD(COORD(obs_\infty)) \tag{7.2}$$
$$obs_\infty = COORD(COORD(COORD(obs_\infty))) \tag{7.3}$$
etc.

Note that while in this form the *horror infinitatis* of expression (6) has disappeared (all expressions in COORD are finite), a new feature has emerged, namely, that the dependent variable obs_∞ is, so to say, "self-depending" (or "self-defining," or "self-reflecting," etc., through the operator COORD).

Should there exist values $obs_{\infty\,i}$ that satisfy equations (7), call these values

"Eigen-Values"
$$obs_{\infty\,i} \equiv Obs_i \qquad\qquad\qquad (8)$$

(or "Eigen-Functions," "Eigen-Operators," "Eigen-Algorithms," "Eigen-Behaviors," etc., depending on the domain of obs) and denote these "Eigen-Values by capitalizing the first letter. (For examples see Appendix A).

Contemplate expressions of the form (7) and note:

(i)   that Eigenvalues are discrete (even if the domain of the primary argument obs_0 is continuous).

This is so because any infinitesimal perturbation $\pm\,\epsilon$ from an Eigenvalue Obs_i (i.e., $Obs_i \pm \epsilon$) will disappear, as did all other values of obs, except those for which $obs = Obs_i$, and obs will be brought either back to Obs_i (*stable* Eigenvalue), or to another Eigenvalue Obs_j (*instable* Eigenvalue Obs_i).

In other words, Eigenvalues represent *equilibria*, and depending upon the chosen domain of the primary argument, these equilibria may be equilibrial values ("Fixed Points"), functional equilibria, operational equilibria, structural equilibria, etc.

(ii)   that Eigenvalues Obs_i and their corresponding operators COORD stand to each other in a complementary relationship, the one implying the other, and vice versa; there the Obs_i represent the externally observable manifestations of the (introspectively accessible) cognitive computations (operations) COORD.

(iii)   that Eigenvalues, because of their self-defining (or self-generating) nature imply topological "closure" ("circularity") (see Figures 2 and 3):

This state of affairs allows a symbolic re-formulation of expression (5);

$$\lim_{n \to \infty} COORD^{(n)} \equiv COORD \;\underline{\quad}$$

that is, the snake eating its own tail:
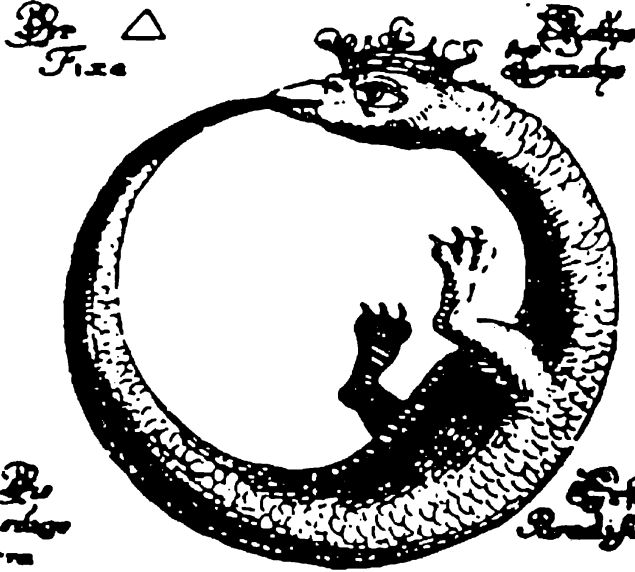
FIGURE 2

cognition computing its own cognitions.
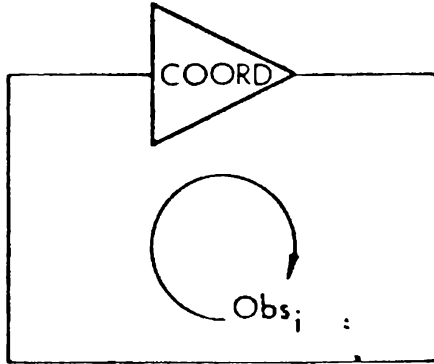
FIGURE 3

* * * * *

Let there be, for a given operator COORD, at least three Eigenvalues

$$Obs_1 , Obs_2 , Obs_3 ,$$

and let there be an (albegraic) composition "*" such that

$$Obs_1 * Obs_2 = Obs_3 , \tag{10}$$

then the coordinating operations COORD appear to coordinate the whole (i.e., the composition of the parts) as a composition of the apparent coordinations of the parts (see proof in Appendix B):

$$COORD(Obs_1 * Obs_2) = COORD(Obs_1) * COORD(Obs_2). \tag{11}$$

In other words, the coordination of compositions (i.e., the whole) corresponds to the composition of coordinations.

This is the condition for what may be called the "principle of cognitive continuity" (e.g., breaking pieces of chalk produces pieces of chalk).

This may be contrasted with the "principle fo cognitive diversity" which arises when the Obs_1 and the composition "*" are *not* the Eigenvalues and compositions complementing the coordination COORD':

$$COORD' (Obs_1 * Obs_2) \neq COORD'(Obs_1) * COORD'(Obs_2), \qquad (12)$$

and which says that the whole is neither more nor is it less than the sum of its parts: it is *different*. Moreover, the formalism in which this sentiment appears (expression (12)) leaves little doubt that it speaks neither of "wholes," nor of "parts" but of a subject's distinction drawn between two states of affairs which by an (other) observer may be seen as being not qualitatively, but only quantitatively distinct.

* * * * *

Eigenvalues have been found ontologically to be discrete, stable, separable and composable, while ontogenetically to arise as equilibria that determine themselves through circular processes. Ontologically, Eigenvalues and objects, and likewise, ontogenetically, stable behavior and the manifestation of a subject's "grasp" of an object cannot be distinguished. In both cases "objects" appear to reside exclusively in the subject's own experience of his sensori-motor coordinations; that is, "objects" appear to be exclusively subjective! Under which conditions, then, do objects assume "objectivity?"

Apparently, only when a subject, S_1 , stipulates the existence of another subject, S_2 , not unlike himself, who, in turn, stipulates the existence of still another subject, not unlike himself, who may well be S_1.

In this atomical social context each subject's (observer's) experience of his own sensori-motor coordination can now be referred to by a token of this experience, the "object," which, at the same time, may be taken as a token for the externality of communal space.

With this I have returned to the topology of closure
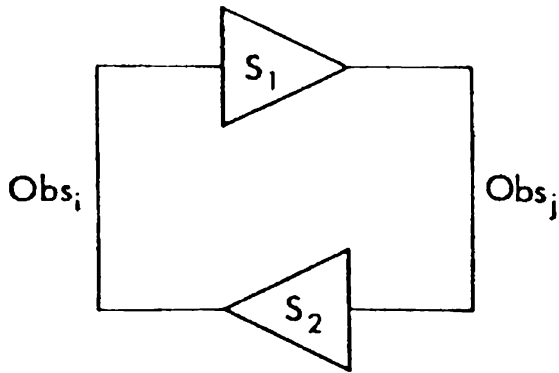
where equilibrium is obtained when the Eigenbehaviors of one participant generate (recursively) those for the other (see, for instance, Appendix Example A 2); where one snake eats the tail of the other as if it were its own, and where cognition computes its own cognitions through those of the other: here is the origin of ethics.
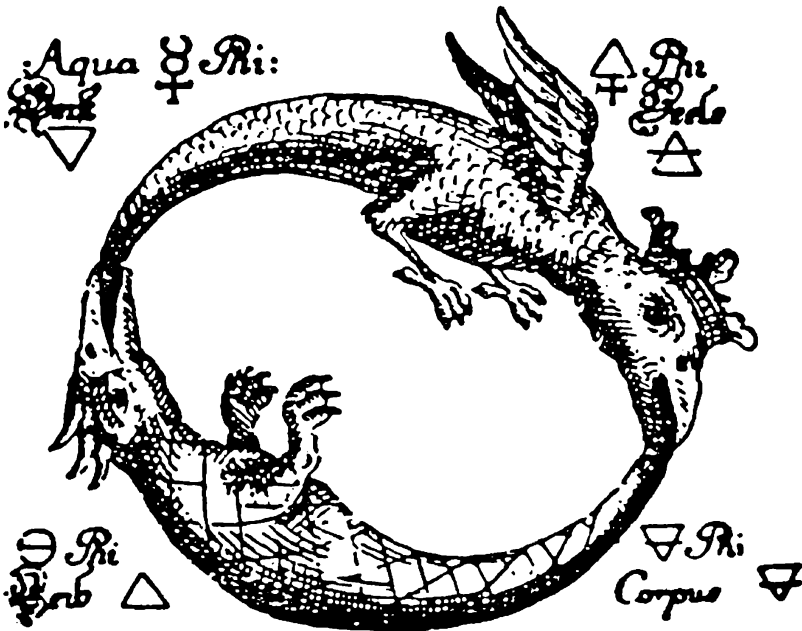
APPENDIX A

Examples:

A 1.   Consider the operator (linear transform) Op_1:

$$Op_1 = \text{``divide by two and add one''}$$

and apply it (recursively) to x_o, x_1, etc., (whose domains are the real numbers).

Choose an initial x_0, say $x_0=4$.

$$x_1 = Op_1(4) = \frac{4}{2} + 1 = 2 + 1 = 3;$$
$$x_2 = Op_1(3) = 2.500;$$
$$x_3 = Op_1(2.500) = 2.250;$$
$$x_4 = Op_1(2.250) = 2.125;$$
$$x_5 = Op_1(2.125) = 2.063;$$
$$x_6 = Op_1(2.063) = 2.031;$$
$$x_{11} = Op_1(x_{10}) = 2.001;$$
$$x_\infty = Op_1(x_\infty) = 2.000$$

Choose another initial value; say $x_0 = 1$

$$x_1 = Op_1(1) = 1.500;$$
$$x_2 = Op_1(1.500) = 1.750;$$
$$x_3 = Op_1(1.750) = 1.875;$$
$$x_8 = Op_1(x_7) = 1.996;$$
$$x_{10} = Op_1(x_9) = 1.999;$$
$$x_\infty = Op_1(x_\infty) = 2.000$$

And indeed:

$$\frac{1}{2} \cdot 2 + 1 = 2$$

$$Op_1(2) = 2$$

i.e., "2" is the (only eigenvalue of Op_1.

A 2.   Consider the operator Op_2:

$$Op_2 = \exp(\cos\quad).$$

There are three eigenvalues, two of which imply each other ("bi-stability"), and the third one being instable:

$$Op_2(2.4452...) = 0.4643...$$
$$Op_2(0.4643...) = 2.4452... \quad \text{stable}$$
$$Op_2(1.3029...) = 1.3092... \quad \text{instable}$$

This means that:

$$Op_2^{(2)} (2.4452...) = 2.4452 \text{ stable}$$

$$Op_2^{(2)} (0.4643...) = 0.4643 \text{ stable}$$

A 3.   Consider the differential operator Op_3:

$$Op_3 = \frac{d}{dx}.$$

The eigenfunction for this operator is the exponential function "exp:"

$$Op_3(\exp) = \exp$$

i.e.,

$$\frac{de^x}{dx} = e^x$$

The generalizations of this operator are, of course, all differential equation, integral equations, integro-differential equations, etc., which can be seen at once when these equations are re-written in operator form, say:

$$F(Op_3^{(n)}, Op_3^{(n-1)} ...., f) = 0$$

Of course, these operators, in turn, may be eigenvalues (eigen-operators) of "meta-operators" and so on. This suggests that COORD, for instance, may itself be treated as an eigen-operator, stable within bounds, and jumping to other values whenever the boundary conditions exceed its former stable domain:

$$Op(COORD_i) = COORD_i.$$

One may be tempted to extend the concept of a meta-operator to that of a "meta-meta-operator" that computes the "eigen-meta-operators," and so on and up a hierarchy without end. However, there is no need to invoke this escape as Warren S. McCulloch has demonstrated years ago in his paper (1945): "A Heterarchy of Values Determined by the Topology of Nervous Nets."

It would go too far in this presentation to demonstrate the construction of heterarchies of operators based on their composability.

A 4.   Consider the (self-referential) proposition:

"THIS SENTENCE HAS. . . LETTERS"

and complete it by writing into the appropriate space the word for the number (or if there are more than one, the numbers) that make this proposition true.

Proceeding by trial and error (comparing what this sentence says (abscissa) with what it is (ordinate)):
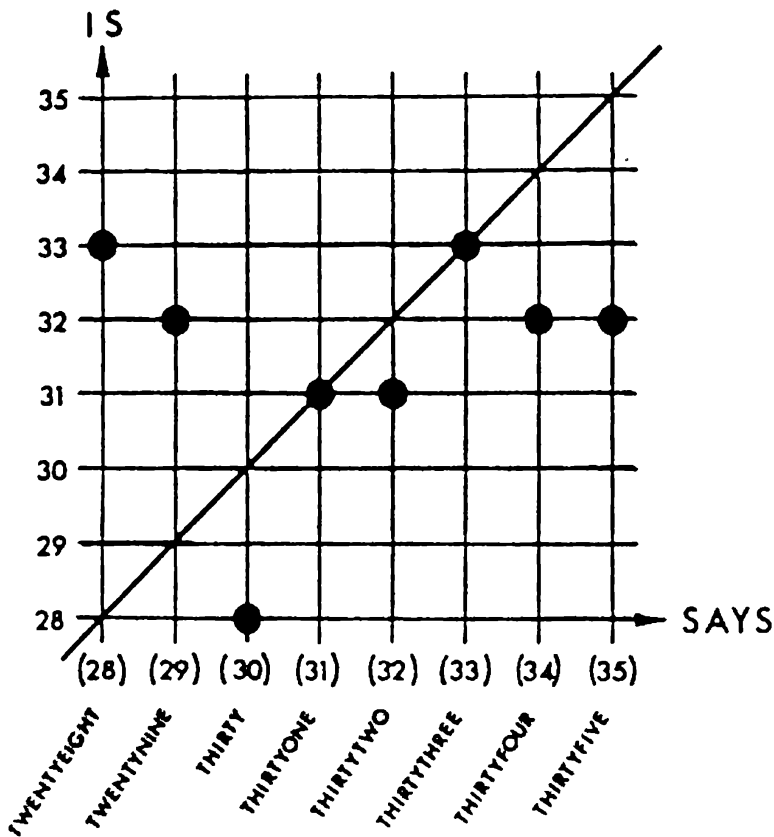
one finds two eigenvalues "thirty-one" and "thirty-three." Apply the proposition above to itself: "This sentence has thirty-one letters' has thirty-one letters." Note that, for instance, the proposition: "this sentence consists of . . . letters" has only one eigenvalue (thirty-nine); while the proposition: "This sentence is composed of. . . letters" has none!

APPENDIX B

B 1. Proof of Expression (11):

$$COORD(Obs_1 * Obs_2) = COORD(Obs_3) =$$
$$Obs_3 = Obs_1 * Obs_2 = COORD(Obs_1) * COORD(Obs_2)$$

Q.E.D.

The apparent distributivity of the operator COORD over the composition "*" should not be misconstrued as "*" being a linear composition. For instance, the fixed points $u_i = \exp(2 \pi \lambda i)$, (for $i = 0, 1, 2, 3 . . .$) that complement the operator $Op(u)$:

$$Op(u) = u \tan \left(\frac{\pi}{4} \pm \frac{1}{\lambda} \ell n\ u \right),$$

with λ an arbitrary constant, compose multiplicatively:

$$Op(u_i \cdot u_j) = Op(u_i) \cdot Op(u_j).$$

etc.

REFERENCES

McCulloch, W.S., *A Heterarchy of Values Determined by the Topology of Nervous Nets*, Bulletin of Mathematical Biophysics, 1945, 7: 89-93.

McCulloch, W.S. and Pitts, W.H., *A Logical Calculus of the Ideas Immanent in Nervous Activity*, Bulletin of Mathematical Biophysics, 1943, 5: 115-133.

Piaget, T., L'Equilibration des Structures Cognitives. Presses Univ. France, Paris, 1975.
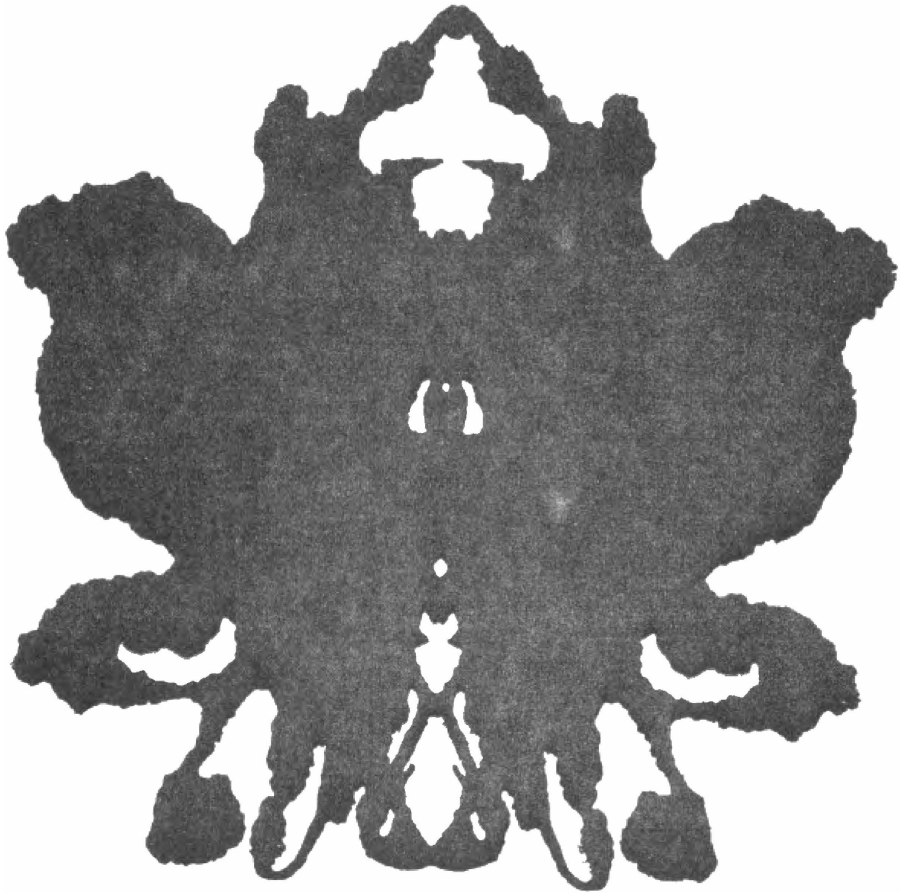
On Constructing a Reality*

Abstract: "Draw a distinction!"[1]

The Postulate: I am sure you remember the plain citizen Jourdain in Moliere's "Bourgeois Gentilhomme" who, nouveau riche, travels in the sophisticated circles of the French aristocracy, and who is eager to learn. On one occasion with his new friends they speak about poetry and prose, and Jourdain discovers to his amazement and great delight that whenever he speaks, he speaks prose. He is overwhelmed by this discovery: "I am speaking Prose! I have always spoken Prose! I have spoken Prose throughout my whole life!"

A similar discovery has been made not so long ago, but it was neither of poetry nor prose—it was the environment that was discovered. I remember when, perhaps ten or fifteen years ago, some of my American friends came running to me with the delight and amazement of having just made a great discovery: "I am living in an Environment! I have always lived in an Environment! I have lived in an Environment throughout my whole life!"

However, neither M. Jourdain nor my friends have as yet made another discovery, and that is when M. Jourdain speaks, may it be prose or poetry, it is he who invents it, and likewise when we perceive our environment, it is we who invent it.

Every discovery has a painful and a joyful side: painful, while struggling with a new insight; joyful, when this insight is gained. I see the sole purpose of my presentation to minimize the pain and maximize the joy for those who have not yet made this discovery; and for those who have made it, to let them know they are not alone. Again, the discovery we all have to make for ourselves is the following postulate: *the environment as we perceive it is our invention.*

The burden is now upon me to support this outrageous claim. I shall proceed by first inviting you to participate in an experiment; then I shall report a clinical case and the results of two other experiments. After this I will give an interpretation, and thereafter a highly compressed version of the neurophysiological basis of these experiments and my postulate of before. Finally, I shall attempt to suggest the significance of all that to aesthetical and ethical considerations.

I. *Blindspot.* Hold next page with your right hand, close your left eye and fixate asterisk of Fig. 1 with your right eye. Move the book slowly back and forth along line of vision until at an appropriate distance, from about 12 to 14 inches, the round black spot

---

Figure 1

disappears. Keeping the asterisk well focused, the spot should remain invisible even if the figure is slowly moved parallel to itself in any direction.

This localized blindness is a direct consequence of the absence of photo receptors (rods or cones) at that point of the retina, the "disc", where all fibers, leading from the eye's light sensitive surface, converge to form the optic nerve. Clearly, when the black spot is projected onto the disc, it cannot be seen. Note that this localized blindness is not perceived as a dark blotch in our visual field (seeing a dark blotch would imply "seeing"), but this blindness is not perceived at all, that is, neither as something present, nor as something absent: whatever is perceived is perceived "blotch-less".

II. *Scotoma.* Well localized occipital lesions in the brain, e.g., injuries from high velocity projectiles, heal relatively fast without the patient's awareness of any perceptible loss in his vision. However, after several weeks motor dysfunction in the patient becomes apparent, e.g., loss of control of arm or leg movements of one side or the other, etc. Clinical tests, however, show that there is nothing wrong with the motor system, but that in some cases there is substantial loss of a large portion of the visual field (*scotoma*) (Fig. 2).[2] A successful therapy consists of blindfolding the patient over a period of one to two months until he regains control over his motor system by shifting his "attention" from 'non-existent' visual clues regarding his posture to 'fully operative' channels that give direct postural clues from 'proprioceptive' sensors em-
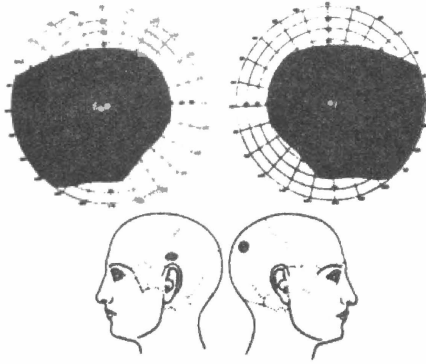
Figure 2

bedded in muscles and joints. Note again the ab-
sence of perception of "absence of perception", and
also the emergence of perception through sen-
sor-motor interaction. This prompts two meta-
phors: "Perceiving is Doing"; and "If I don't see I
am blind, I am blind; but if I see I am blind, I see".

III. *Alternates.* A single word is spoken once into
a tape recorder and the tape smoothly spliced,
without a click, into a loop. The word is repetitively
played back with a high rather than low volume.
After one or two minutes of listening, from 50 to
150 repetitions, the word clearly perceived so far
abruptly changes into another meaningful and
clearly perceived word: an "alternate". After 10 to
30 repetitions of this first alternate, a sudden switch
to a second alternate is perceived, and so on.[3] The
following is a small selection of the 758 alternates
reported from a population of about 200 subjects
who were exposed to a repetitive playback of the
single word *Cogitate: agitate; annotate; arbitrate; ar-
tistry; back and forth; brevity; ca d'etait; candidate; can't
you see; can't you stay; cape cod you say; card estate; cardio
tape; car district; catch a tape; cavitate; cha cha che; cogi-
tate; computate; conjugate; conscious state; counter tape;
count to ten; count to three; count yer tape; cut the steak;
entity; fantasy; God to take; God you say; got a date; got
your pay; got your tape; gratitude; gravity; guard the tit;
gurgitate; had to take; kinds of tape; majesty; marmalade.*

IV. *Comprehension*. Literally defined: *con* ➔ together; *prehendere* ➔ to seize, grasp. Into the various stations of the auditory pathways in a cat's brain, micro-electrodes are implanted which allow a recording, "Electroencephalogram", from the nerve cells first to receive auditory stimuli, Cochlea Nucleus: CN, up to the Auditory Cortex.[4] The so prepared cat is admitted into a cage that contains a food box whose lid can be opened by pressing a lever. However, the lever-lid connection is operative only when a short single tone (here C_6, that is about 1000 Hz) is repetitively presented.[*] The cat has to learn that C_6 "means" food. Figures 3 to 6 show the pattern of nervous activity at eight ascending auditory stations, and at four consecutive stages of this learning process.[4] The cat's behavior associated with the recorded neural activity is for Fig. 3: "Random search"; Fig 4: "Inspection of lever"; Fig.
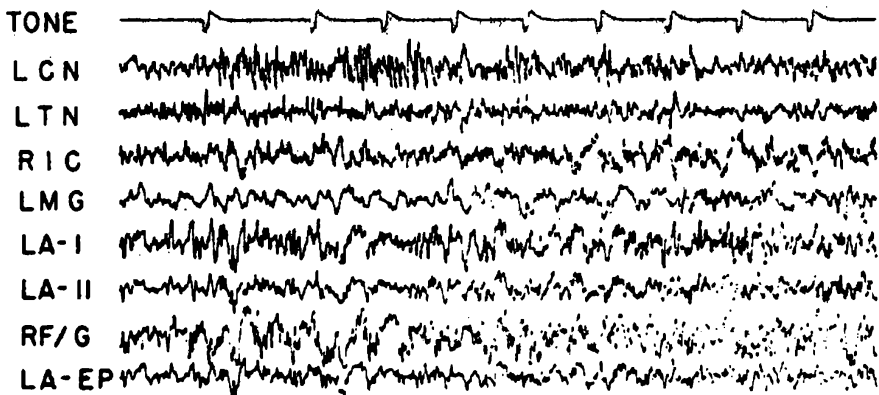
FIGURE 3. Trial 1 (no behavioral evidence of learning)

[*]"Hz" means 1 cycle per second, is the unit for oscillations named after Heinrich Hertz who generated the first radio signals.
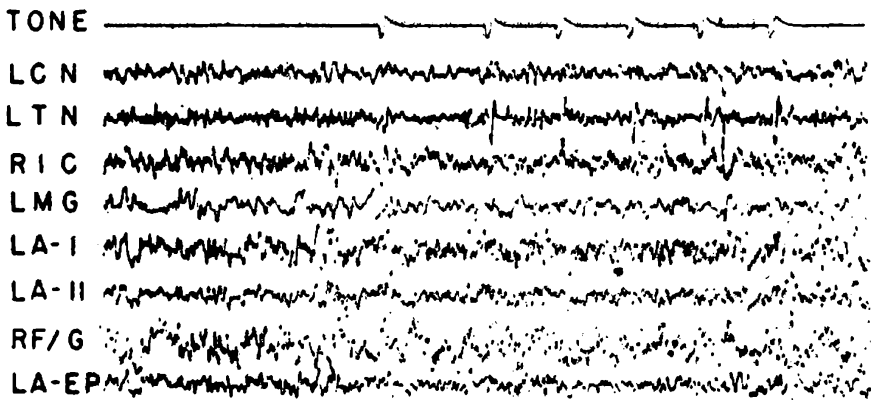
FIGURE 4. Trial 13 (begins to wait for tones)

FIGURE 5. Trial 4/20 (hypothesizes)

FIGURE 6. Trial 6/9 (understands)

5: "Lever pressed at once"; and for Fig 6: "Walking straight toward level (full comprehension)". Note that no tone is perceived as long as this tone is uninterpretable (Figs. 3, 4; pure noise), but the whole system swings into action with the appearance of the first "beep" (Figs. 5, 6; noise becomes signal) when sensation becomes comprehensible, when *our* perception of "beep", "beep", "beep", is in the *cat's* perception "food", "food", "food".

Interpretation. In these experiments I have cited instances in which we see or hear what is not "there", or in which we do not see or hear what is "there", unless coordination of sensation and movement allows us to "grasp" what appears to be there. Let me strengthen this observation by citing now the "Principle of Undifferentiated Encoding":

> The response of a nerve cell does *not* encode the physical nature of the agents that caused its response. Encoded is only "how much" at this point on my body, but not "what".

Take, for instance, a light sensitive receptor cell in the retina, a "rod", which absorbs the electro-magnetic radiation originating from a distant source. This absorption causes a change in the electro-chemical potential in the rod which will ultimately give rise to a periodic electric discharge of some cells higher up in the post-retinal networks with a period that is commensurate with the intensity of the radiation absorbed, but without a clue that it was electro-magnetic radiation that caused the rod to discharge. The same is true for any other sensory receptor, may it be the taste buds, the touch receptors, and all the other receptors that are associated with the sensations of smell, heat and cold, sound, etc.: they are all "blind" as to the quality of their stimulation, responsive only as to their quantity.

Although surprising, this should not come as a surprise, for indeed "out there" there is no light and no color, there are only electro-magnetic waves; "out there" there is no sound and no music, there are only periodic variations of the air pressure; "out

there" there is no heat and no cold, there are only moving molecules with more or less mean kinetic energy, and so on. Finally, for sure, "out there" there is no pain.

Since the physical nature of the stimulus—its *quality*—is not encoded into nervous activity, the fundamental question arises as to how does our brain conjure up the tremendous variety of this colorful world as we experience it any moment while awake, and sometimes in dreams while asleep. This is the "Problem of Cognition", the search for an understanding of the cognitive processes.

The way in which a question is asked determines the way in which an answer may be found. Thus, it is upon me to paraphrase the "Problem of Cognition" in such a way that the conceptual tools that are today at our disposal may become fully effective. To this end let me paraphrase (→) "cognition" in the following way:

COGNITION → computing a reality

With this I anticipate a storm of objections. First, I appear to replace one unknown term, "cognition", with three other terms, two of which, "computing" and "reality", are even more opaque than the definiendum, and with the only definite word used here being the indefinite article "a". Moreover, the use of the indefinite article implies the ridiculous notion of other realities besides "the" only and one reality, our cherished Environment; and finally I seem to suggest by "computing" that everything, from my wristwatch to the Galaxies, is merely computed, and is not "there". Outrageous!

Let me take up these objections one by one. First, let me remove the semantic sting that the term "computing" may cause in a group of women and men who are more inclined toward the humanities than to the sciences. Harmlessly enough, computing (from *com-putare*) literally means to reflect, to contemplate (*putare*) things in concert (*com-*), without any explicit reference to numerical quantities. Indeed, I shall use this term in this most general

sense to indicate any operation, not necessarily numerical, that transforms, modifies, re-arranges, or orders observed physical entities, "objects", or their representations, "symbols". For instance, the simple permutation of the three letters A,B,C, in which the last letter now goes first: C,A,B, I shall call a computation. Similarly, the operation that obliterates the commas between the letters: CAB; and likewise the semantic transformation that changes CAB into TAXI, and so on.

I shall now turn to the defense of my use of the indefinite article in the noun-phrase "a reality". I could, of course, shield myself behind the logical argument that solving for the general case, implied by the "a", I would also have solved any specific case denoted by the use of "the". However, my motivation lies much deeper. In fact, there is a deep hiatus that separates the "The" -school-of- thought from the "A"-school-of-thought in which respectively the distinct concepts of "confirmation" and "correlation" are taken as explanatory paradigms for perceptions. The "The-School": My sensation of touch is *confirmation* for my visual sensation that here is a table. The "A-School": My sensation of touch in *correlation* with my visual sensation generate an experience which I may describe by "here is a table".

I am rejecting the THE-position on epistemological grounds, for in this way the whole Problem of Cognition is safely put away in one's own cognitive blind spot: even its absence can no longer be seen.

Finally one may rightly argue that cognitive processes do not compute wristwatches or galaxies, but compute at best *descriptions* of such entities. Thus I am yielding to this objection and replace my former paraphrase by:
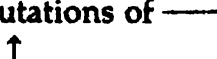
COGNITION → computing descriptions
of a reality.

Neurophysiologists, however, will tell us that a description computed on one level of neural activi-

ty, say a projected image on the retina, will be operated on again on higher levels, and so on, whereby some motor activity may be taken by an observer as a "terminal description", for instance the utterance: "here is a table".[5] Consequently, I have to modify this paraphrase again to read:

COGNITION → computing descriptions of ⌐
$\uparrow$ ⌐____⌐

where the arrow turning back suggests this infinite recursion of descriptions of descriptions . . . etc. This formulation has the advantage that one unknown, namely, "reality" is successfully eliminated. Reality appears only implicit as the operation of recursive descriptions. Moreover, we may take advantage of the notion that computing descriptions is nothing else but computations. Hence:

COGNITION → computations of ⌐
$\uparrow$ ⌐____⌐

In summary, I propose to interpret cognitive processes as never ending recursive processes of computation, and I hope that in the following *tour de force* of neurophysiology I can make this interpretation transparent.

Neurophysiology

I. *Evolution.* In order that the principle of recursive computation is fully appreciated as being the underlying principle of all cognitive processes— even of life itself, as one of the most advanced thinkers in biology assures me—it may be instructive to go back for a moment to the most elementary—or as evolutionists would say, to very "early"—manifestations of this principle.[6] These are the "independent effectors", or independent sensory-motor units, found in protozoa and metazoa distributed over the surface of these animals (Fig. 7). The triangular portion of this unit, protruding with its tip from the surface, is the sensory part, the onion-shaped portion the contractile motor part. A change in the chemical concentration of an agent in the immediate vicinity of the sensing tip,

and "perceptible" by it, causes an instantaneous contraction of this unit. The resulting displacement of this or any other unit by change of shape of the animal or its location may, in turn, produce perceptible changes in the agent's concentration in the vicinity of these units which, in turn, will cause their instantaneous contraction, . . . etc. Thus, we have the recursion:

→ change of sensation → change of shape ─┐
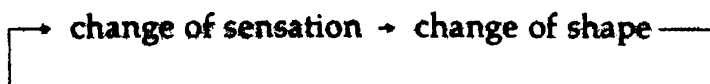
Separation of the sites of sensation and action appears to have been the next evolutionary step (Figure 8). The sensory and motor organs are now connected by thin filaments, the "axons" (in essence degenerated muscle fibers having lost their contractility), which transmit the sensor's perturbations to its effector, thus giving rise to the concept of a "signal": see something here, act accordingly there.

The crucial step, however, in the evolution of the complex organization of the mammalian central nervous system (CNS) appears to be the appearance of an "internuncial neuron", a cell sandwiched between the sensory and the motor unit (Fig. 9). It
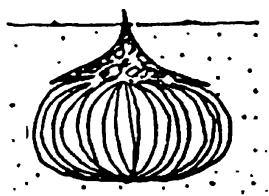
Figure 7
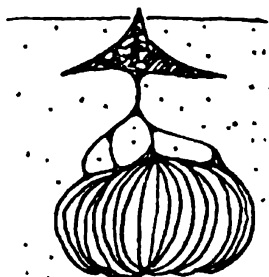
Figure 8
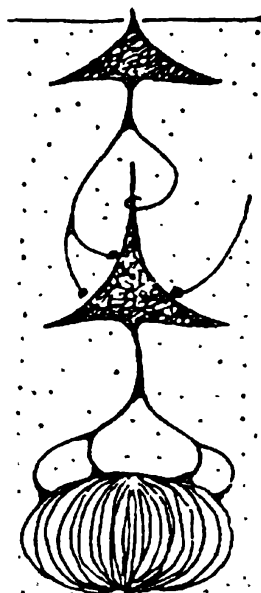
Figure 9

is, in essence, a sensory cell, but specialized so as to respond only to a universal "agent", namely, the electrical activity of the afferent axons terminating in its vicinity. Since its present activity may affect its subsequent responsivity, it introduces the element of computation in the animal kingdom, and gives these organisms the astounding latitude of non-trivial behaviors. Having once developed the genetic code for assembling an internuncial neuron, to add the genetic command "repeat" is a small burden indeed. Hence, I believe, it is now easy to comprehend the rapid proliferation of these neurons along additional vertical layers with growing horizontal connections to form those complex interconnected structures we call "brains".

II. *Neuron.* The neuron, of which we have more than ten billion in our brain, is a highly specialized single cell with three anatomically distinct features (Fig. 10): (a) the branch-like ramifications stretching up and to the side, the "dendrites"; (b) the bulb in the center housing the cell's nucleus, the "cell body"; and (c), the "axon", the smooth fiber stretching downward. Its various bifurcations terminate on dendrites of another (but sometimes [recursively] on the same) neuron. The same membrane which envelopes the cell body forms also the tubular sheath for dendrites and axon, and causes the inside of the cell to be electrically charged against the outside with about one tenth of a volt. If in the dendritic region this charge is sufficiently perturbed, the neuron "fires" and sends this perturbation along its axons to their terminations, the synapses.

III. *Transmission.* Since these perturbations are electrical, they can be picked up by "microprobes", amplified and recorded. Fig. 11 shows three examples of periodic discharges from a touch receptor under continuous stimulation, the low frequency corresponding to a weak, the high frequency to a strong stimulus. The magnitude of the discharge is clearly everywhere the same, the pulse frequency representing the stimulus intensity, but
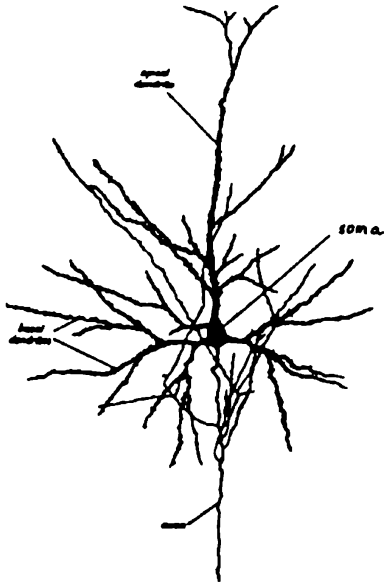
Figure 11

Figure 10

the intensity only.

IV. *Synapse.* Fig. 12 sketches a synaptic junction. The afferent axon (Ax), along which the pulses travel, terminates in an end bulb (EB) which is separated from the spine (sp) of a dendrite (D) of the target neuron by a minute gap (sy), the "synaptic gap" (Note the many spines that cause the rugged appearance of the dendrites in Fig. 10). The chemical composition of the "transmitter substances"
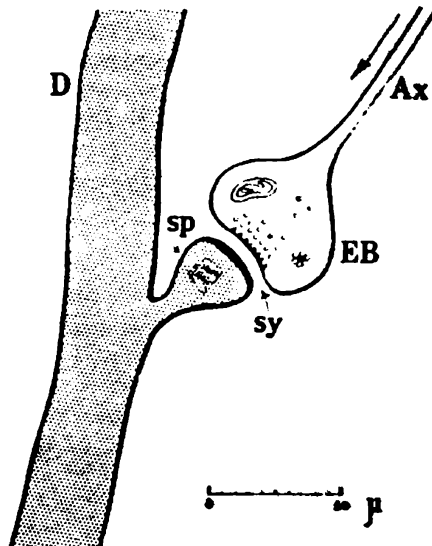
Figure 12

filling the synaptic gap is crucial in determining the effect an arriving pulse may have on the ultimate response of the neuron: under certain circumstances it may produce an "inhibitory effect" (cancellation of another simultaneously arriving pulse); in others a "facilitory effect" (augmenting another pulse to fire the neuron). Consequently, the synaptic gap can be seen as the "micro-environment" of a sensitive tip, the spine, and with this interpretation in mind we may compare the sensitivity of the CNS to changes of the *internal* environment (the sumtotal of all micro-environments) to those of the *external* environment (all sensory receptors). Since there are only a hundred million sensory receptors, and about ten-thousand billion synapses in our nervous system, we are 100,000 times more receptive to changes in our internal than in our external environment.

V. *Cortex.* In order that one may get at least some perspective on the organization of the entire machinery that computes all perceptual, intellectual and emotional experiences, I have attached Fig. 13 which shows magnified a section of about 2 square millimeters of a cat's cortex by a staining method which stains only cell body and dendrites, and of those only 1% of all neurons present.[7] Although you have to imagine the many connections among these neurons provided by the (invisible) axons,
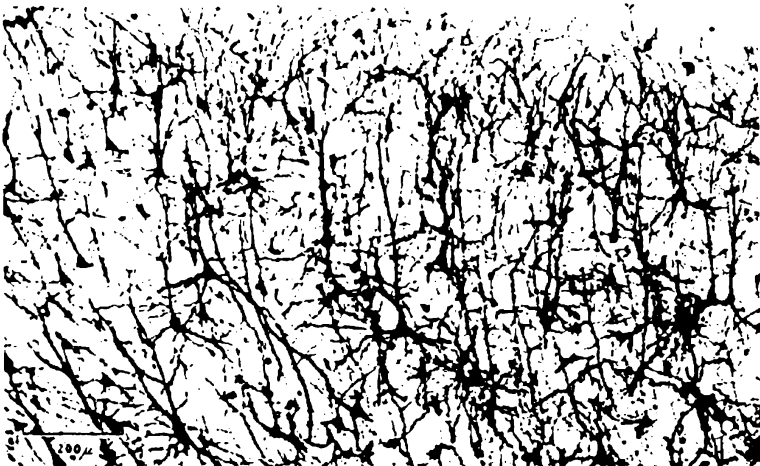
Figure 13

and a density of packing that is a hundred times that shown, the computational power of even this very small part of a brain may be sensed.

VI. *Descartes*. This perspective is a far cry from that being held, say three hundred years ago: "If the fire A is near the foot B (Fig. 14), the particles of this fire, which as you know move with great rapidity, have the power to move the area of the skin of this foot that they touch; and in this way drawing the little thread, c, that you see to be attached at the base of toes and on the nerve, at the same instant they open the entrance of the pore, d, e, at which this little thread terminates, just as by pulling one end of a cord, at the same time one causes the bell to sound that hangs at the other end.[8] Now the entrance of the pore or little conduit, d, e, being thus opened, the animal spirits of the cavity F, enter within and are carried by it, partly into the muscles that serve to withdraw this foot from the fire, partly into those that serve to turn the eyes and the head to look at it, and partly into those that serve to advance the hands and to bend the whole body to protect it."

Note, however, that some behaviorists of today still cling to the same view with one difference only, namely, that in the meantime Descartes' "animal spirit" has gone into oblivion.[9]

Figure 14

VII. *Computation*. The retina of vertebrates with its associated nervous tissue is a typical case of neural computation. Fig. 15 is a schematic representation of a mammalian retina and its post-retinal network. The layer labeled #1 represents the array of rods and cones, and layer #2 the bodies and nuclei of these cells. Layer #3 identifies the general region where the axons of the receptors synapse with the dendritic ramifications of the "bipolar cells" (#4) which, in turn, synapse in layer #5 with the dendrites of the ganglion cells (#6) whose activity is transmitted to deeper regions of the brain via their axons which are bundled together to form the optic nerve (#7). Computation takes place within the two layers labeled #3 and #5, that is, where the synapses are located.

As Maturana has shown, it is there where the sensation of color and some clues as to form are computed.[10]

1
2
3
4
5
6
7
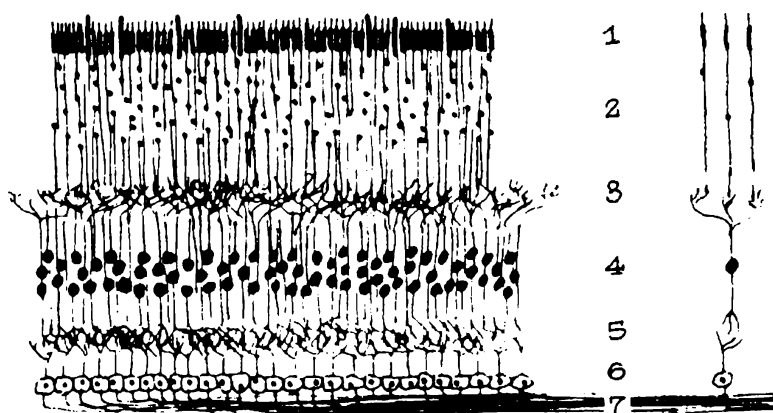
Figure 15

Form computation: take the two-layered periodic network of Fig. 16, the upper layer representing receptor cells sensitive to, say, "light". Each of these receptors is connected to three neurons in the lower (computing) layer, with two excitatory synapses on the neuron directly below (symbolized by buttons attached to the body), and with one inhibitory synapse (symbolized by a loop around the tip) attached to each of the two neurons, one to the left and one to the right. It is clear that the computing layer will not

respond to uniform light projected on the receptive layer, for the two excitatory stimuli on a computer neuron will be exactly compensated by the inhibitory signals coming from the two lateral receptors. This zero-response will prevail under strongest and weakest stimulation as well as to slow or rapid changes of the illumination. The legitimate question may now arise—"Why this complex apparatus that doesn't do a thing?"

Consider now Fig. 17 in which an obstruction is placed in the light path illuminating the layer of receptors. Again all neurons of the lower layer will remain silent, except the one at the edge of the obstruction, for it receives two excitatory signals from the receptor above, but only one inhibitory

Figure 16

0   0   0   0   0

Figure 17

0   1   0   0   0
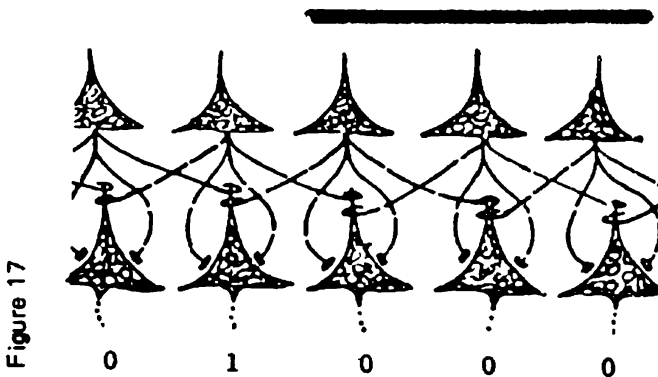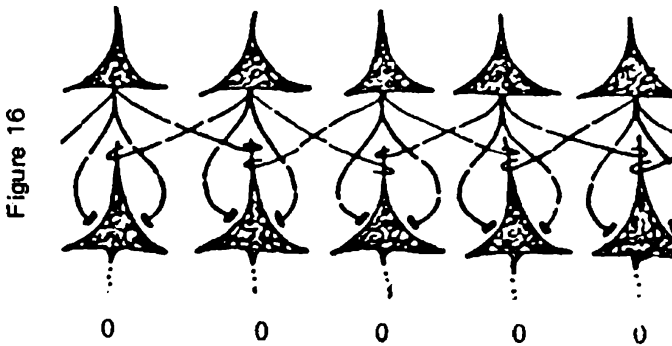
signal from the sensor to the left. We now understand the important function of this net, for it

computes any spatial *variation* in the visual field of this "eye", independent of intensity of the ambient light and its temporal variations, and independent of place and extension of the obstruction.

Although all operations involved in this computation are elementary, the organization of these operations allows us to appreciate a principle of considerable depth, namely, that of the computation of abstracts, here the notion of "edge".

I hope that this simple example is sufficient to suggest to you the possibility of generalizing this principle in the sense that "computation" can be seen on at least two levels, namely, (a) the operations actually performed, and (b) the organization of these operations represented here by the structure of the nerve net. In computer language (a) would again be associated with "operations", but (b) with the "program". As we shall see later, in "biological computers" the programs themselves may be computed on. This leads to the concepts of "meta-programs", "meta-meta-programs", . . . etc. This, of course, is the consequence of the inherent recursive organization of those systems.

VIII. *Closure.* By attending to all the neurophysiological pieces, we may have lost the perspective that sees an organism as a functioning whole. In Fig. 18 I have put these pieces together in their functional context. The black squares labeled N represent bundles of neurons that synapse with neurons of other bundles over the (synaptic) gaps indicated by the spaces between squares. The sensory surface (SS) of the organism is to the left, its motor surface (MS) to the right, and the neuropituitary (NP) the strongly innervated mastergland that regulated the entire endocrinal system, is the stippled lower boundary of the array of squares. Nerve impulses traveling horizontally (from left to right) ultimately act on the motor surface (MS) whose changes (movements) are immediately sensed by the sensory surface (SS), as suggested by the "external" pathway following the arrows. Impulses traveling vertically (from top to bottom) stimulate the neuro-

pituitary (NP) whose activity releases steroids into
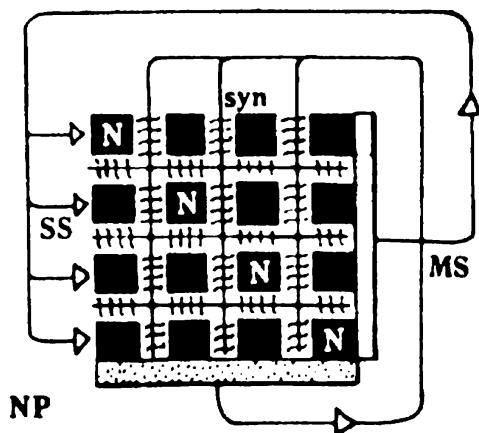
Figure 18

the synaptic gaps, as suggested by the wiggly ter-
minations of the lines following the arrow, and thus
modify the *modus operandi* of all synaptic junctures,
hence the *modus operandi* of the system as a whole.
Note the double closure of the system which now
recursively operates not only on what it "sees" but
on its operators as well. In order to make this two-
fold closure even more apparent I propose to wrap
the diagram of Fig. 18 around its two axes of circular
symmetry until the artificial boundaries disappear
and the torus (doughnut) as in Fig. 19 is obtained.
Here the "synaptic gap" between the motor and
sensory surfaces is the striated meridian in the front
center, the neuropituitary the stippled equator.
This, I submit, is the functional organization of a
living organism in a (dough)nut shell. (Fig. 19)
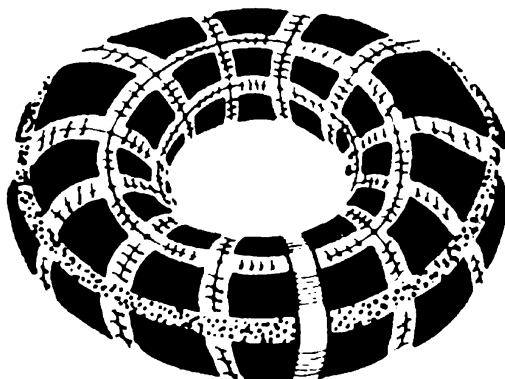
Figure 19

The computations within this torus are subject to a non-trivial constraint, and this is expressed in the Postulate of Cognitive Homeostasis:

> The nervous system is organized (or organizes itself) so that it computes a stable reality.

This postulate stipulates "autonomy", i.e., "self-regulation", for every living organism. Since the semantic structure of nouns with prefix "self-" becomes more transparent when this prefix is replaced by the noun, "autonomy" becomes synonymous with "regulation of regulation". This is precisely what the doubly closed, recursively computing torus does: it regulates its own regulation.

Significance. It may be strange in times like these to stipulate autonomy, for autonomy implies responsibility: If I am the only one who decides how I act then I am responsible for my action. Since the rule of the most popular game played today is to make someone else responsible for *my* acts—the name of the game is "heteronomy"—my arguments make, I understand, a most unpopular claim. One way of sweeping it under the rug is to dismiss it as just another attempt to rescue "solipsism", the view that this world is only in my imagination and the only reality is the imagining "I". Indeed, that was precisely what I was saying before, but I was talking only about a single organism. The situation is quite different when there are two, as I shall demonstrate with the aid of the gentlemen with the bowler hat (Fig. 20).

He insists that he is the sole reality, while everything else appears only in his imagination. However, he cannot deny that his imaginary universe is populated with apparitions that are not unlike himself. Hence, he has to concede that they themselves may insist that they are the sole reality and everything else is only a concoction of their imagination. In that case their imaginary universe will be populated with apparitions, one of which may be *he*, the gentleman with the bowler hat.
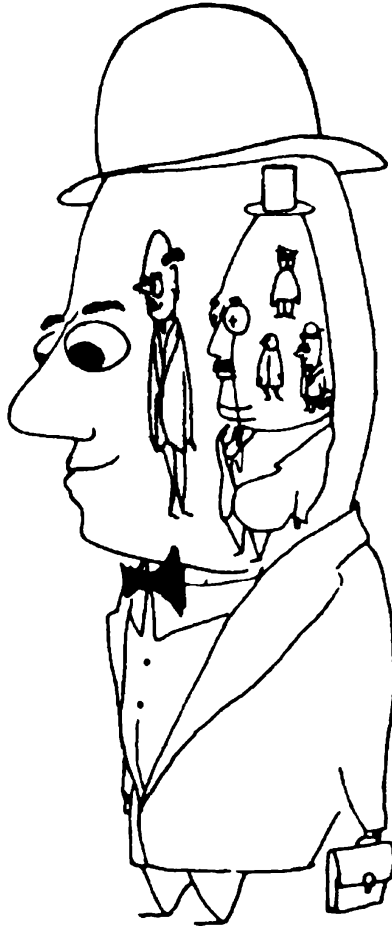
Figure 20

According to the Principle of Relativity which rejects a hypothesis when it does not hold for two instances together, although it holds for each instance separately (Earthlings and Venusians may be consistent in claiming to be in the center of the universe, but their claims fall to pieces if they should ever get together), the solipsistic claim falls to pieces when besides me I invent another autonomous organism. However, it should be noted that since the Principle of Relativity is not a logical necessity, nor is it a proposition that can be proven to be either true or false, the crucial point to be recognized here is that I am free to choose either to adopt this principle or to reject it. If I reject it, I am

the center of the universe, my reality are my dreams and my nightmares, my language is monologue, and my logic mono-logic. If I adopt it, neither me nor the other can be the center of the universe. As in the heliocentric system, there must be a third that is the central reference. It is the relation between Thou and I, and this relation is IDENTITY:

Reality = Community.

What are the consequences of all this in ethics and aesthetics?

The Ethical Imperative: Act always so as to increase the number of choices.

The Aesthetical Imperative: If you desire to see, learn how to act.

Acknowledgement

References

1. Brown, G. S., *Laws of Form*. New York, Julian Press, page 3, 1972.

2. Teuber, H. L., "Neuere Betrachtungen über Sehstrahlung und Sehrinde" in Jung, R., Kornhuber, H. (Eds.) *Das Visuelle System*, Berlin, Springer, pages 256–274, 1961.

3. Naeser, M. A., and Lilly, J. C., "The Repeating Word Effect: Phonetic Analysis of Reported Alternates", *Journal of Speech and Hearing Research*, 1971.

4. Worden, F. G., "EEG Studies and Conditional Reflexes in Man", in Brazier, Mary A. B., *The Central Nervous System and Behavior*, New York, Josiah Macy Jr. Foundation, pages 270–291, 1959.

5. Maturana, H. R., "Neurophysiology of Cognition" in Garvin, P., *Cognition: A Multiple View*, New York, Spartan Press, pages 3–23, 1970.

6. Maturana, H. R., *Biology of Cognition*, University of Illinois, 1970.

7. Sholl, D. A., *The Organization of the Cerebral Cortex*, London, Methuen, 1956.

8. Descartes, R., *L'Homme*, Paris, Angot, 1664. Reprinted in *Ouevres de Descartes*, XI, Paris, Adam and Tannery, pages 119–209, 1957.

9. Skinner, B. F., *Beyond Freedom and Dignity*, New York, Knopf, 1971.

10. Maturana, H. R., "A Biological Theory of Relativistic Colour Coding in the Primate Retina", *Arch. Biologia y Medicina Exper., Suppl.* No. 1, Soc. Biologia de Chile, Santiago, Universidad de Chile, 1968.

PART III
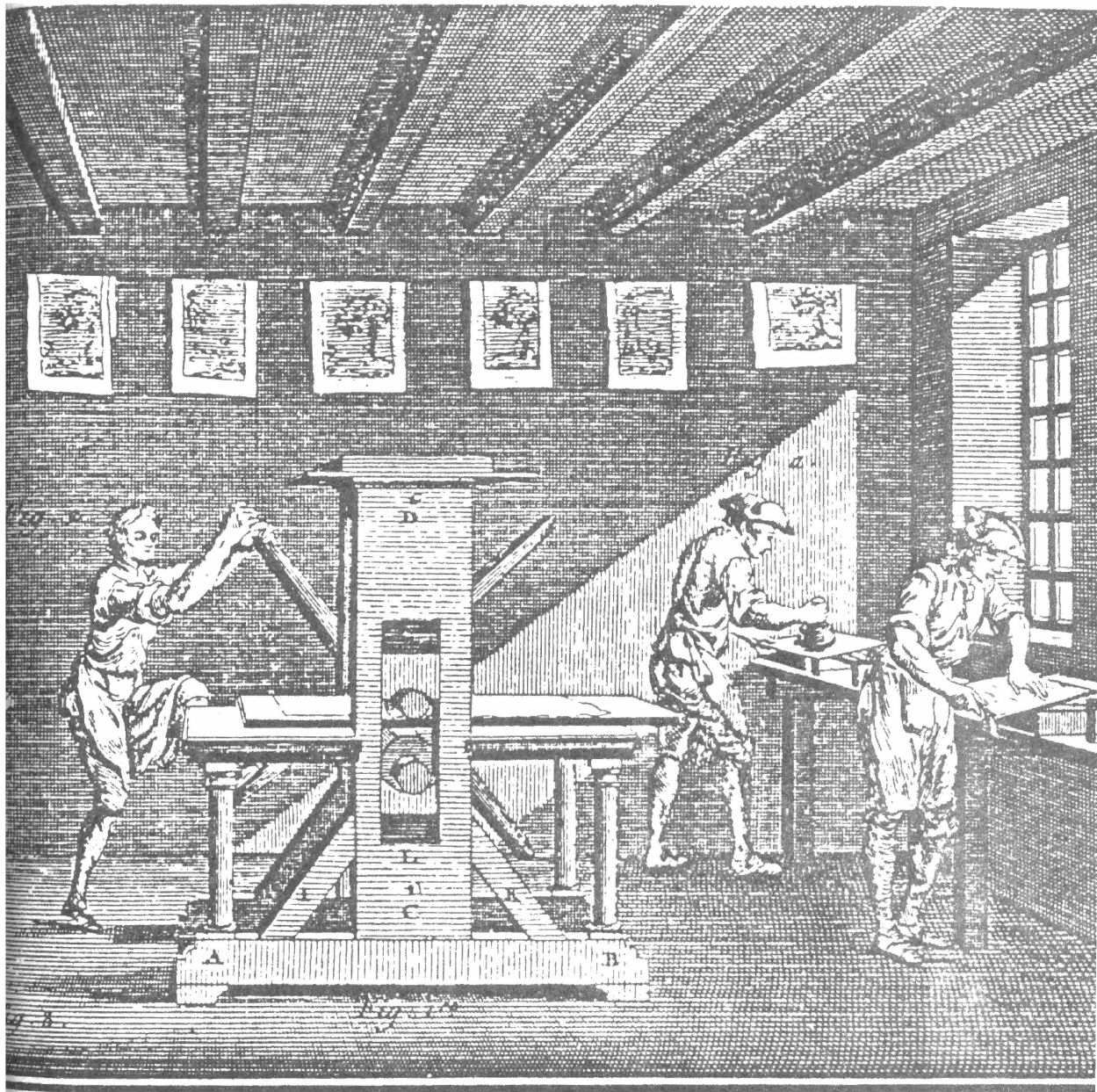
Publications

by

HEINZ VON FOERSTER

PUBLICATIONS

Heinz Von Foerster

1943

1. *Uber das Leistungsproblem beim Klystron*, Ber. Lilienthal Ges. Luftfahrtforschung, 155, pp. 1-5 (1943).

1948

2. Das Gedachtnis: Eine quantenmechanische Untersuchung, F. Deuticke, Vienna, 40 pp. (1948).

1949

3. Cybernetics: Transactions of the Sixth Conference (ed.), Josiah Macy Jr. Foundation, New York, 202 pp. (1949).

4. *Quantum Mechanical Theory of Memory* in Cybernetics: Transactions of the Sixth Conference (ed.), Josiah Macy Jr. Foundation, New York, pp. 112-145 (1949).

1950

5. With Margaret Mead and H. L. Teuber, Cybernetics: Transactions of the Seventh Conference (eds.), Josiah Macy Jr. Foundation, New York, 251 pp. (1950).

1951

6. With Margaret Mead and H. L. Teuber, Cybernetics: Transactions of the Eighth Conference (eds.), Josiah Macy Jr. Foundation, New York, 240 pp. (1951).

1953

7. With M. L. Babcock and D. F. Holshouser, *Diode Characteristic of a Hollow Cathode*, Phys. Rev., 91, 755 (1953).

8. With Margaret Mead and H. L. Teuber, Cybernetics: Transactions of the Ninth Conference (eds.), Josiah Mac Jr. Foundation, New York, 184 pp. (1953).

1954

9. With E. W. Ernst, *Electron Bunches of Short Time Duration*, J. of Appl. Phys., 25, 674 (1954).

10. With L. R. Bloom, *Ultra-High Frequency Beam Analyzer*, Rev. of Sci. Instr., 25, pp. 640-653 (1954).

11. *Experiment in Popularization*, Nature, 174, 4424, London (1954).

1955

12. With Margaret Mead and H. L. Teuber, Cybernetics: Transactions of the Tenth Conference (eds.), Josiah Macy Jr. Foundation, New York, 100 pp. (1955).

13. With O. T. Purl, *Velocity Spectrography of Electron Dynamics in the Traveling Field*, Journ. of Appl. Phys., 26, pp. 351-353 (1955).

14. With E. W. Ernst, *Time Dispersion of Secondary Electron Emission*, Journ. of Appl. Phys., 26, pp. 781-782 (1955).

1956

15. With M. Weinstein, *Space Charge Effects in Dense, Velocity Modulated Electron Beams*, Journ. of Appl. Phys., 27, pp. 344-346 (1956).

1957

16. With E. W. Ernst, O. T. Purl, M. Weinstein, *Oscillographie analyse d'un faisceau hyperfrequences*, LE VIDE, 70, pp. 341-351 (1957).

1958

17. *Basic Concepts of Homeostasis* in Homeostatic Mechanisms, Upton, New York, pp. 216-242 (1958).

1959

18. With G. Brecher and E. P. Cronkite, *Production, Ausreifung and Lebensdauer der Leukozyten* in Physiologie und Physiopathologie der weissen Blutzellen, H. Braunsteiner (ed.), George Thieme Verlag, Stuttgart, pp. 188-214 (1959).

315

19. *Some Remarks on Changing Populations* in The Kinetics of Cellular Proliferation, F. Stohlman Jr. (ed.), Grune and Stratton, New York, pp. 382-407 (1959).

1960
20. *On Self-Organizing Systems and Their Environments* in Self-Organizing Systems, M.C. Yovits and S. Cameron (eds.), Pergamon Press, London, pp. 31-50 (1960).
20.1 *O Samoorganizuyushchiesja Sistemach i ich Okrooshenii* in Samoorganizuyushchiesju Sistemi,113-139, M. I. R. Moscow (1964).
20.2 *Sobre Sistemas Autoorganizados y sus Contornos* in Epistemologia de la Communicacion, Juan Antonio Bofil (ed.), Fernando Torres, Valencia, pp. 187-214 (1976).
21. With P. M. Mora and L. W. Amiot, *Doomsday: Friday, November 13, AD 2026*, Science, 132, pp. 1291-1295 (1960).
22. *Bionics* in Bionics Symposium, Wright Air Development Division, Technical Report 60-600, J. Steele (ed.), pp. 1-4 (1960).
23. *Some Aspects in the Design of Biological Computers* in Sec. Inter. Congress on Cybernetics, Namur, pp. 241-255 (1960).

1961
24. With G. Pask, *A Predictive Model for Self-Organizing Systems* , Part I: Cybernetica, 3, pp. 258-300; Part II: Cybernetica, 4, pp. 20-55 (1961).
25. With P. M. Mora and L. W. Amiot, *Doomsday*, Science, 133, pp. 936-946 (1961).
26. With D. F. Holshouser and G. L. Clark, *Microwave Modulation of Light Using the Kerr Effect*, Journ. Opt. Soc. Amer., 51, pp. 1360-1365 (1961).
27. With P. M. Mora and L. W. Amiot, *Population Density and Growth*, Science, 133, pp. 1931-1937 (1961).

1962
28. With G. Brecher and E. P. Cronkite, *Production, Differentiation and Lifespan of Leukocytes* in The Physiology and Pathology of Leukocytes, H. Braunsteiner (ed.), Grune & Stratton, New York, pp. 170-195 (1962).
29. With G. W. Zopf, Jr., Principles of Self-Organization: The Illinois Symposium on Theory and Technology of Self-Organizing Systems (eds.), Pergamon Press, London, 526 pp. (1962).
30. *Communication Amongst Automata*, Amer. Journ. Psychiatry, 118, pp. 865-871 (1962).
31. With P. M. Mora and L. W. Amiot, *"Projections" versus "Forecasts" in Human Population Studies*, Science, 136, pp. 173-174 (1962).
32. *Biological Ideas for the Engineer*, The New Scientist, 15, 173-174 (1962).
33. *Bio-Logic* in Biological Prototypes and Synthetic Systems, E. E. Bernard and M. A. Kare (eds.), Plenum Press, New York, pp. 1-12 (1962).
33.1 *Bio-Logika* in Problemi Bioniki, pp. 9-23, M. I. R., Moscow (1965).
34. *Circuitry of Clues of Platonic Ideation* in Aspects of the Theory of Artificial Intelligence, C. A. Muses (ed.), Plenum Press, New York, pp. 43-82 (1962).
35. *Perception of Form in Biological and Man-Made Systems* in Trans. I.D.E.A. Symp., E. J. Zagorski (ed.), University of Illinois, Urbana, pp. 10-37 (1962).
36. With W. R. Ashby and C. C. Walker, *Instability of Pulse Activity in a Net with Threshold*, Nature, 196, pp. 561-562 (1962).

1963
37. *Bionics* in McGraw-Hill Yearbook Science and Technology, McGraw-Hill, New York, pp. 148-151 (1963).
38. *Logical Structure of Environment and Its Internal Representation* in Trans. Internat'l Design Conf., Aspen, R. E. Eckerstrom (ed.), H. Miller, Inc., Zeeland, Mich., pp. 27-38 (1963).
39. With W. R. Ashby and C. C. Walker, *The Essential Instability of Systems with Threshold, and Some Possible Applications to Psychiatry* in Nerve, Brain and Memory Models, N. Wiener and J. P. Schade (eds.), Elsevier, Amsterdam, pp. 236-243 (1963).

1964

40. *Molecular Bionics* in Information Processing by Living Organisms and Machines, H. L. Oestreicher (ed.), Aerospace Medical Division, Dayton, pp. 161-190 (1964).

41. With W. R. Ashby, *Biological Computers* in Bioastronautics, K. E. Schaefer (ed.), The Macmillan Co., New York, pp. 333-360 (1964).

42. *Form: Perception, Representation and Symbolization* in Form and Meaning, N. Perman (ed.), Soc. Typographic Arts, Chicago, pp. 21-54 (1964).

43. *Structural Models of Functional Interactions* in Information Processing in the Nervous System, R. W. Gerard and J. W. Duyff (eds.), Excerpta Medica Foundation, Amsterdam, The Netherlands, pp. 370-383 (1964).

44. *Physics and Anthropology*, Current Anthropology, 5, pp. 330-331 (1964).

1965

45. *Memory without Record* in The Anatomy of Memory, D. P. Kimble (ed.), Science and Behavior Books, Palo Alto, pp. 388-433 (1965).

46. *Bionics Principles* in Bionics, R. A. Willaume (ed.), AGARD, Paris, pp. 1-12 (1965).

1966

47. *From Stimulus to Symbol* in Sign, Image, Symbol, G. Kepes (ed.), George Braziller, New York, pp. 42-61 (1966).

1967

48. *Computation in Neural Nets*, Currents Mod. Biol., 1, pp. 47-93 (1967).

49. *Time and Memory* in Interdisciplinary Perspectives of Time, R. Fischer (ed.), New York Academy of Sciences, New York, pp. 866-873 (1967).

50. With G. Gunther, *The Logical Structure of Evolution and Emanation* in Interdisciplinary Perspectives of Time, R. Fischer (ed.), New York Academy of Sciences, New York, pp. 874-891 (1967).

51. *Biological Principles of Information Storage and Retrieval* in Electronic Handling of Information: Testing and Evaluation, Allen Kent et al. (eds.), Academic Press, London, pp. 123-147 (1967).

1968

52. With A. Inselberg and P. Weston, *Memory and Inductive Inference* in Cybernetic Problems in Bionics, Proceedings of Bionics 1966, H. Oestreicher and D. Moore (eds.), Gordon & Breach, New York, pp. 31-68 (1968).

53. With J. White, L. Peterson and J. Russell, Purposive Systems, Proceedings of the 1st Annual Symposium of the American Society for Cybernetics (eds.), Spartan Books, New York, 179 pp. (1968).

1969

54. With J. W. Beauchamp, Music by Computers (eds.), John Wiley & Sons, New York, 139 pp. (1969).

55. *Sounds and Music* in Music by Computers, H. Von Foerster and J. W. Beauchamp (eds.), John Wiley & Sons, New York, pp. 3-10 (1969).

56. *What Is Memory that It May Have Hindsight and Foresight as well?* in The Future of the Brain Sciences, Proceedings of a Conference held at the New York Academy of Medicine, S. Bogoch (ed.), Plenum Press, New York, pp. 19-64 (1969).

57. *Laws of Form*, (Book Review of Laws of Form, G. Spencer Brown), Whole Earth Catalog, Portola Institute; Palo Alto, California, p. 14, (Spring 1969).

1970

58. *Molecular Ethology, An Immodest Proposal for Semantic Clarification* in Molecular Mechanisms in Memory and Learning, G. Ungar (ed.), Plenum Press, New York, pp. 213-248 (1970).

59. With A. Inselberg, *A Mathematical Model of the Basilar Membrane*, Mathematical Biosciences, 7, pp. 341-363, (1970).

60. *Thoughts and Notes on Cognition* in Cognition: A Multiple View, P. Garvin (ed.), Spartan Books, New York, pp. 25-48 (1970).

317

61. *Bionics, Critique and Outlook* in Principles and Practice of Bionics, H. E. von Gierke, W. D. Keidel and H. L. Oestreicher (eds.), Technivision Service, Slough, pp. 467-473 (1970).

62. *Embodiments of Mind*, (Book Review of Embodiments of Mind, Warren S. McCulloch), Computer Studies in the Humanities and Verbal Behavior, III (2), pp. 111-112 (1970).

63. With L. Peterson, *Cybernetics of Taxation: The Optimization of Economic Participation*, Journal of Cybernetics, 1 (2), pp. 5-22 (1970).

64. *Obituary for Warren S. McCulloch* in ASC Newsletter, 3 (1), (1970).

1971

65. *Preface* in Shape of Community by S. Chermayeff and A. Tzonis, Penguin Books, Baltimore, pp. xvii-xxi (1971).

66. Interpersonal Relational Networks (ed.); CIDOC Cuaderno No. 1014, Centro Intercultural de Documentacion, Cuernavaca, Mexico, 139 pp. (1971).

67. *Technology: What Will It Mean to Librarians?*, Illinois Libraries, 53 (9) 785-803 (1971).

68. *Computing in the Semantic Domain*, in Annals of the New York Academy of Science, 184, 239-241 (1971).

1972

69. *Responsibilities of Competence*, Journal of Cybernetics, 2 (2) pp. 1-6 (1972).

70. *Perception of the Future and the Future of Perception* in Instructional Science, 1 (1), pp. 31-43 (1972).

70.1 *La Percepcion de Futuro y el Futuro de Percepcion* in Comunicacion No. 24 Madrid (1975).

1973

71. With P. E. Weston, *Artificial Intelligence and Machines that Understand*, Annual Review of Physical Chemistry, H. Eyring, C. J. Christensen, H. S. Johnston (eds.), Annual Reviews, Inc.; Palo Alto, pp. 353-378 (1973).

72. *On Constructing a Reality*, in Environmental Design Research, Vol. 2, F. E. Preiser (ed.), Dowden, Hutchinson & Ross; Stroudberg, pp. 35-46 (1973).

72.1 *Construir la Realidad* in Infancia y Aprendizaje, 1 (1), pp. 79-92, Madrid (1978).

72.2 *Das Konstruieren einer Wirklichkeit* in Die erfundene Wirklichkeit, Paul Watzlawick (ed.), Piper & Co. , Muenchen, pp. 39-60 (1981)

72.3 *On Constructing a Reality*, in The Invented Reality, Paul Watzlawick (ed.), W.W. Norton & Co., New York, 41-62 (1984).

1974

73. Giving with a Purpose: The Cybernetics of Philanthropy, Occasional Paper No. 5, Center for a Voluntary Society, Washington, D.C., 19 pp. (1974).

74. *Kybernetik einer Erkenntnistheorie*, in Kybernetic und Bionik, W.D. Keidel, W. Handler & M. Spring (eds.), Oldenburg; Munich, 27-46, (1974).

75. *Epilogue to Afterwords*, in After Brockman: A Symposium, ABYSS, 4, 68-69, (1974).

76. With R. Howe, *Cybernetics at Illinois* (Part One), Forum, 6 (3), 15-17 (1974); (Part Two), Forum, 6 (4), 22-28 (1974).

77. Notes on an Epistemology for Living Things, BCL Report No. 9.3 (BCL Fiche No. 104/1), Biological Computer Laboratory, Department of Electrical Engineering, University of Illinois, Urbana, 24 pp., (1972).

77.1 *Notes pour un epistemologie des objets vivants*, in L'unite de L'homme, Edgar Morin and Massimo Piatelli-Palmarini (eds.), Edition du Seuil; Paris, 401-417. (1974).

78. With P. Arnold, B. Aton, D. Rosenfeld, K. Saxena, *Diversity H: A Measure Complementing Uncertainty H*, SYSTEMA, No. 2, January 1974.

79. *Culture and Biological Man* (Book Review of Culture and Biological Man, by Elliot D. Chapple), Current Anthropology, 15 (1), 61 (1974).

80. *Comunicacion, Autonomia y Libertad* (Entrovista con H.V.F. Comunicacion No. 14, 33-37, Madrid (1974).

1975

81. With R. Howe, *Introductory Comments to Francisco Varela's Calculus for Self-Reference*, Int. Jour. for General Systems, 2, 1-3 (1975).

82. *Two Cybernetics Frontiers*, (Book Review of Two Cybernetics Frontiers by Stewart Brand), The Co-Evolutionary Quarterly, 2, (Summer), 143 (1975).

83. *Oops: Gaia's Cybernetics Badly Expressed*, The Co-Evolutionary Quarterly, 2, (Fall), 51 (1975).

1976

84. *Objects: Tokens for (Eigen-)Behaviors*, ASC Cybernetics Forum, 8, (3 & 4) 91-96 (1976).(English version of 84.1).

1977

84.1 *Formalisation de Certains Aspects de l'Équilibration de Structures Cognitives*, in Epistémologie Génétique et Équilibration, B. Inhelder, R. Garcias and J. Voneche (eds.), Delachaux et Niestle, Neuchatel, 76-89 (1977).

95. *The Curious Behavior of Complex Systems: Lessons from Biology*, in Futures Research, H.A. Linstone and W.H.C. Simmonds (eds.), Addison-Wesley, Reading, 104-113 (1977).

1979

96. *Where Do We Go from Here?* in The History and Philosophy of Technology, George Bugliarello and Dean B. Doner (eds.), University of Illinois Press, Urbana, 358-370 (1979).

1980

85. *Minicomputer — verbindende Elemente*, Chip, Jan. 1980, p. 8.

86. *Epistemology of Communication*, in The Myths of Information: Technology and Postindustrial Culture, Kathleen Woodward (ed.), Coda Press, Madison, 18-27 (1980).

1981

87. *Morality Play*, The Sciences, 21 (8) 24-25 (1981).

88. *Gregory Bateson*, The Esalen Catalogue, 20 (1), 10 (1981).

89. *On Cybernetics of Cybernetics and Social Theory*, in Self-Organizing Systems, G. Roth and H. Schwegler (eds.), Campus Verlag, Frankfurt, 102-105 (1981).

90. *Foreword*, in Rigor and Imagination, C. Wilder-Mott and John H. Weakland (eds.), Praeger, New York, vii-xi (1981).

91. *Understanding Understanding: An Epistemology of Second Order Concepts* , in Aprendizagem/Desenvolvimento, 1 (3), 83-85 (1981).

1982

92. Observing Systems, with an Introduction by Francisco J. Varela, Intersystems Publications, Seaside,

93. *A Constructivist Epistemology*, Cahiers de la Fondation Archives Jean Piaget, No. 3, Geneve, 191-213 (1982).

94. *To Know and to Let Know: An Applied Theory of Knowledge*, Canadian Library Journal, 39, no. 5, 277-284 (October, 1982).

BCL

THE COMPLETE PUBLICATIONS
OF THE BIOLOGICAL COMPUTER LABORATORY
THE UNIVERSITY OF ILLINOIS

The complete publications of the Biological Computer Laboratory from 1957 to 1976 as edited by Kenneth Wilson with over 300 papers by more than 100 authors is now available in microfiche through the Illinois Blueprint Corporation, Micrographics Dept.*(821 Bond Street, Peoria, Illinois 61603). This 14,000 page collection is contained on 148 standard (24x) microfiche cards with a complete index and an introduction.

Among many authors the collection contains works by:

W. R. ASHBY: 55 papers on
Information Theory, Self-Organizing Systems, Regulation and Control, System Stability, Many Dimensional Relations in Large Systems

GOTTHARD GÜNTHER: 34 papers on
Multivalued Logic, Non-Classical Logical Structures, Philosophical Foundations of Cybernetics

LARS LÖFGREN: 12 papers on
Systems Theory, Self-Reproducing Systems, Automata Theory, The Theory of Descriptions

HUMBERTO MATURANA: 2 books and 5 papers on
The Biology of Cognition, The Organization of Living Systems, Epistemology, Language, Memory, Consciousness, and other aspects of Cognition

GORDON PASK: 6 papers on
System Modeling, Goal Oriented Systems, Computer Aided Instruction

HEINZ VON FOERSTER: 91 papers on
Self-Organizing Systems, Information Theory, Biological Computers, Memory, Perception and Cognition, and the Philosophy of Cybernetics

TOPICAL AREAS INCLUDE: Cybernetics, Cognition, Perception, Memory, Learning, Systems Theory (General, Control, Biological and Social), Multi-Valued Logic, Computer Science (Semantic Computation, Relational Data Structures, Information Processing), Automata Theory, Philosophy, Linguistics, Movement Notation, and the Cybernetics of Cybernetics.

---

*$59.00

INDEX