Form
Perception, Representation and Symbolization

Heinz Von Foerster

BCL #109, 1962

Ladies and Gentlemen, I would like to deliver my message — if I have any message at all — in two separate parcels. The one parcel will have to do with the interaction of automata with their environment, that is with the universe in which they are embedded; the other parcel will have something to do with the interaction of such automata with each other, when they attempt to communicate some of the features of the universe which I shall assume to be common to all of them.

You will be shocked, perhaps, by the idea that I am going to use automata as the central figures of my discussion about such esoteric matters as perception, representations and eventually aesthetics, instead of making man the actor in this play. However, I hope that first I shall be able to dispel the general notion that an automaton is a pedestrian, predictable, deterministic moron; and second that in specifying the kind of irrationality, ingenuity and the freedom of choice in my fellow-automata I shall be in a much better position to deal with, for instance, the all-elusive problem of esthetic's, which will be seen to belong to the second parcel of my presentation, because the logical modality of this problem requires at least two interacting elements that communicate certain features of their common environment.

My first task in this presentation is a rehabilitation of the term "automaton." Unfortunately, in informal discussion, but also in recent literature, journal articles and in the press, you will find the terms "automaton" and "robot" freely interchanged as if they would refer to one and the same thing. This, however, is not the case. While "robot" is derived from the Czech word *robotnik* — worker, became popular through Capek's delightful play *Rossum's Universal Robots*, and refers to a stupid mechanism carrying out without its own initiative all that it is commanded to do, "automaton" is derived from the Greek *automatizein* — to act according to one's own will, and thus refers to a gadget on a much higher level of sophistication. Indeed, if you care to look up "automaton" in a dictionary[1] you will find that an automaton is "… a contrivance constructed to act as if spontaneously, through concealed motive power." It may be argued that this definition describes still a pedestrian gadget, because with patience and skill we may "reveal" the concealed mechanism. However, the situation changes drastically, if — for some reason or other — we are in principle unable to reveal that hidden mechanism. Under those circumstances we are forced to drop the "as if" in the above definition and we have a truly "free" system before us which acts on its "own will". It may, perhaps, amuse you to note that Aristotle used the term "automaton" in the latter sense[2]. I presume that a bad translation in the early nineteenth century of the famous passage in *De Motu* shifted its meaning to its weaker usage of today.

At this point you may rightly ask how such systems can ever be built. Unfortunately, a tight proof of my assertion of the feasibility of such systems would take up a one-semester seminar, thus, I hope you will believe me that such systems can be built, if they are made up of certain elementary components that can be specified in precise terms[3]. I shall later discuss such elementary components, however not in terms of electronic gadgetry, but in physiological terms. With your permission I shall give you a brief description of structure and function of the physiological elementary component, the nerve-cell, billions of which interact in a most remarkable way to insure proper — and, alas, sometimes improper — action of the most fascinating automaton, namely man. On the other hand, to suggest the feasibility to construct electronic automata of reasonable sophistication, I have in the hall the electronic model of a simple "eye" of such automata of the future. This eye (fig. 1), we call "Numarete", is an artificial "retina" that counts the number of separate entities in its visual field[4]. I shall give you later a more detailed account of such elementary "perceptors". Presently, however, I would like to go a bit deeper into the concepts of "spontaneity" and "freedom", particularly "freedom of will", because these concepts ore generally believed to be exclusively associated with anthropomorphic functions and hence *ipso facto* not associable with artifacts.

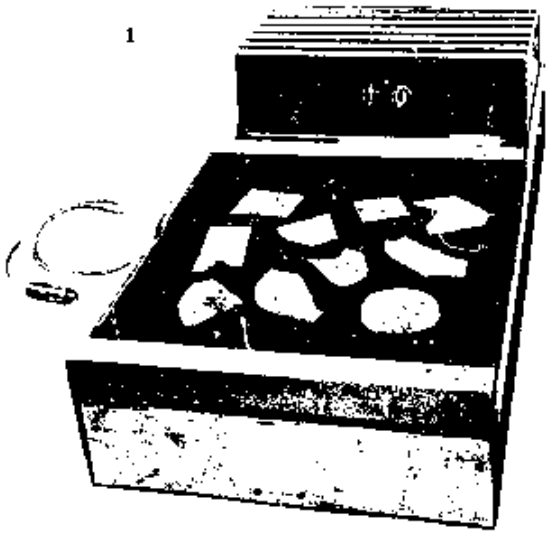


Figure 1: The "Numa-Rete", an artificial "retina" counting the number of objects independent of their size, location and form, and independent of strength of illumination.

As has been noticed relatively early[5] "freedom" is essentially a negative concept in the sense that it is an indication of the absence of constraints, or to put it even more precisely, an indication of our ignorance of the constraints acting on the system. This is best illustrated if we take the degree of predictability of a system as its apparent freedom. This suggests at the same time the relative notion of this concept because predictability may vary from observer to observer. For a people which does not know the laws of celestial mechanics an eclipse may appear as a spontaneous act of a supreme being, while for a student of planetary motion this event may be used to set his clock. To him the constraints that govern the motion of celestial bodies are known and he can predict within fraction of seconds the onset of, say, a solar eclipse. On the other hand, he is lost when invited to predict the outcome of the toss of a die, despite the fact that all laws that govern the motion of the die are known. Here, the sheer complexity of evaluating all determining conditions as, for instance, initial angular momentum, velocity and coordinates of the die, roughness of the table on all points, etc., that make it impossible to predict the outcome. Nevertheless, one can still make certain predictions of the results of tossing a die, however, they will be formulated in a somewhat different way as compared to predictions made in the case of determinable systems. For instance, we are able to say that each outcome will belong to a set of precisely six elements, namely the six different eyes of the die. In addition, we are able to say that in the long run, playing with an "honest" die, each face will come up with approximately the same frequency, that is with approximately 1/6 of the number of tosses. In other words, our predictions will assume statistical character if we are to deal with systems which react under fixed rules, or fixed constraints, but whose behavior is too complex to be analyzed in each particular case. The circumstances I have just described are still very pedestrian indeed, because I can name at once an additional feature which will make my system much more sophisticated, or, let us say, more unpredictable, or more "free", and that is if I assume that the constraints that govern the behavior of the system are themselves influenced by the behavior of the system. For example, I may suggest that my die is filled with a viscous liquid in which a piece of lead is suspended. If in the beginning this "loading" is perfectly centered, this die will first appear as an honest die. However, in the long run this die will have the tendency to favor a particular face, because if a certain face per chance has come up more frequently in the early tosses the load will have sunk to the opposite side shown while at rest and thus enhances the probability to come up with the same side again. This die has, so to say, a "memory" that regulates its future behavior as a consequence of its past "experience". Although this die's behavior becomes easier to predict after every toss, because its internal constraints are growing, from a statistical point of view it is a hard nut to crack, because the laws that describe its behavior are changing all the time. It is this kind of complexity that we have to expect in our automata that elevates them to critters of considerable sophistication with a peculiar kind of unpredictability, and hence with an apparent considerable degree of freedom. On the other hand it is easy to see that if we endow our automata with the capacity to assess their own activity but remaining ignorant about their internal constraints that govern this activity, they, themselves, will boast about their freedom of will, for which we, who have built these constraints into their workings, have only a mild, sympathetic smile.

I hope that I have not overestimated your patience with this somewhat prolonged introduction in which I only stated the postal regulations under which I shall deliver my two parcels.

Let me now briefly outline the contents of my first parcel which deals with the interaction of automata with their environment. In Figure 2 I have attempted to sketch the main points I want to make now. The two entities "environment" and "automaton", that I have mentioned before, are in Figure 2 represented as weakly outlined boxes labeled "universe" and "system" respectively. What I want to show in the following is first, that the constraints (Con) which prevail in this universe give rise to its structuralization (Str.); and second, that our system perceives (dotted arrows) some of the structures

in this universe via its sensory apparatus in form of structural information (St.In.). This information is passed on to its "brain" (B), which, in turn, has to compute, or to reconstruct, or to hypothesize about, the constraints prevailing in the universe that caused the structures, the information of which was passed on to the brain, etc., etc. The circularity of this process is, I believe, quite obvious in this sketch, if one just follows the arrows which indicate the flow of information. Since running around this loop takes time, you may, perhaps, appreciate my suggestion that the self-correcting action in evolution that eliminates systems which permit misrepresentations of environmental structures follows precisely the same circular pattern, if we replace "Brain" by "Gene".

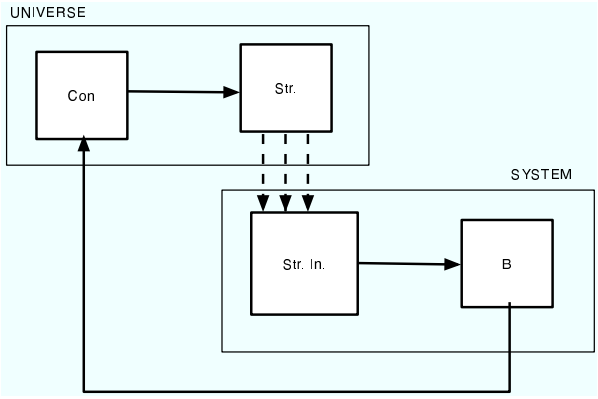


Figure 2: Diagram of the information flow between a "Universe" and a "System."

In my introductory remarks I have already suggested what I mean by "constraint", namely that in a universe with constraints not everything happens that may happen from a purely logical point of view. In other words, in a universe with constraints the probability for certain states — or state transitions — is vanishingly small. For instance, that this room suddenly transforms itself into Priam's castle in Troy is, alas, very small indeed. Or — to put it even more bluntly — a universe with constraints acts very much like a complicated, but heavily loaded die.

However, I have not yet shown that prevailing constraints in a universe give rise to its structuralization. Let me show you this interdependence of constraint and structure by using again as an example a die as the generator for some structures. Assume that the movements of a spot in a plane are controlled by the outcome of the tosses of a die in the following way: at each toss our spot makes one step of unit length, however, the direction in which it turns is dictated by the [eyes] shown by the die: if one eye comes up, the spot turns $1 \times 60°$ to the right, if two eyes come up, it turns $2 \times 60°$ to the right; etc., ... if 6 eyes show, it just turns around $6 \times 60° = 360°$. In Figure 3 I have a hexagonal net in which three sets of parallel lines intersect each other at precisely 60°. The intersection points are one unit step apart and hence, our spot will move from intersection point to intersection point, leaving a trace of its activity.

Put yourself for a moment into the shoes of this spot. You have come up to an intersection point where there are precisely six choices of where to go from here. A decision has to be made, and hence these intersection points may be called "decision points ." In absence of any particular philosophy about what to do, you consult a die (the "Eena, meena, mina, mo — method" of children, also called the "Monte Carlo — method" amongst modern computer engineers). Assume the die comes up with face 2. You proceed one unit step at an angle of 120°to the right from where you came on the previous step. With this you have reached a new decision point and again you consult your die as to what to do next, etc. The outcome of such a "random walk" process is given in Figure 3.1 for the first 26 tosses with an "honest" die. The sequence is:

$$311223541131263644152363 2\ldots$$

in the case that you care to check it. Clearly the result is an unpredictable, unstructured zig-zag path, that will eventually cover the whole plane. In strong contrast to this path arc the patterns drawn in Figures 3.2, 3.3, 3.4 and 3.5, the easiest to understand is 3.3. Here the die was heavily loaded and could only show either face 2 or face 4, depending upon whether you run through this hexagon clockwise or counterclockwise. In Figure 3.2, after an initial "transient" of coming up with a "2" the die assumes the stereotyped pattern 34343434... In Figure 3.4 the periodic sequence 44414441... generates the hexagonal star, while the intricate geometrical figure in 3.5 is produced by the period

$$3313132335133.$$

With these examples I wanted to show first that constraints imposed upon a "universe", which is represented here simply by a die and a spot, indeed produce some structure in this universe; second, that with the increase of the intricacies of the constraints, structures of higher complexity do emerge. We may even go one level up in our abstractions and study the class of structures for which the constraints have constraints. For instance, it may amuse you to realize that our moving spot will describe closed figures only if the sum of the digits in the generating period cannot be divided by 3. In our little universe, such theorems play a similar fundamental role as, say, the principles of conservation of energy or Newton's Laws play in the framework of theoretical physics.

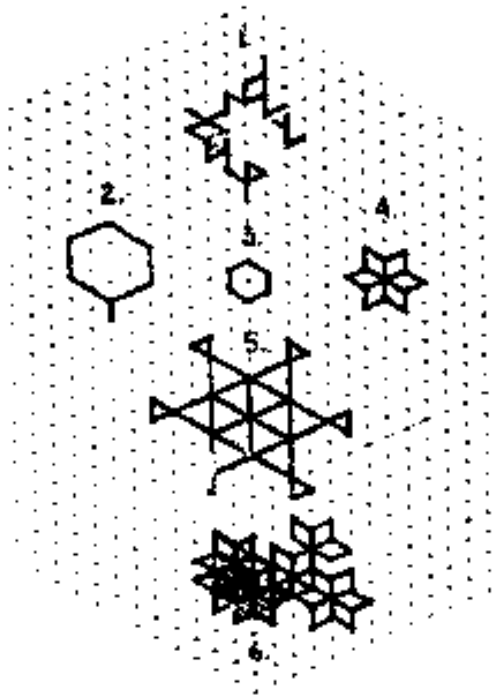


Figure 3. 3. Walking patterns of a completely free (I), a completely deterministic (2) (3) (4) and (5). and a slightly constraint system (6).

I have given two examples of universes, one that is completely unpredictable, indeterministic and free, and another one, where each event is determined by the previous event. In this case the future is absolutely predictable, because the constraints are so strong that there is no freedom left for the system. In my final example concerning this simple universe I shall propose a mixture of these principles in the sense that the probability of a particular sequence is slightly enhanced at the expense of the probability of all other sequences. Take a sequence with a period of four digits. The probability of a particular four digit sequence is $(1/6)^4 = 1/1296$. I propose to construct a die that gives all four digit sequences an extremely small reduction in the probability of their occurrence, namely 1/1297 instead of 1/1296, with the exception of one sequence, which now has an enhanced chance to come up, namely with a probability of 2/1297. I choose the sequence 4441 of our previous example (fig. 3.4) as the predominant sequence and "tossed" an electronic die 10,000 times. The resulting path of our spot is drawn in Figure 3.6. It is — at least for me — always surprising to see the remarkable structuralization that takes place if only slight constraints are in operation. Clearly the old hexagonal star emerges again and again, this time — of course — in a variety of overlappings, since 1295 other four digit sequences have a chance to come up.

I shall leave now this somewhat abstract game and turn to another example of the generation of structure through constraints that have much more concrete appeal. However, the notions developed in the abstract dice game will come in handy in dealing with the regulations of the universe which I propose next. Take the 26 letters of our alphabet plus a "space" as the possible 27 states of our universe. If no constraints prevail in this universe, that is, if all letters may come up with equal probability of 1/27, a sequence of the following form may be generated with an "honest" wheel-of-fortune having 27 positions (6):

(i) `... XFOML RXKHRJFFJVJ ZLPWCFWKCYJ FFJEYVKCQSGHYD QPAAMKBAAACIBZLHJQD ...`

This is called a "zero order approximation" to English because no constraints of letter sequences that exist in English are taken into consideration. We may now introduce the actual frequency distribution of letters in English texts and generate the following sequence with a slightly crooked wheel of fortune:

(ii) `... OCRO HLI RGWR NMIELWIS EV LL NBNESEBYA TH EEI ALHENHTPPA OOBTTVA NAH BRL ...`

Here we have a first order approximation to English and although it does not sound very English yet, one sees at least some structures emerge that may have word character. The situation improves if we go to the second order approximation in which we take notice of the constraints that bind two adjacent letters with preferred probabilities (diagrams):

(iii) `... ON IE ANTSOUTINYS ARE T INCTORE ST BE S DEAMY ACHIND ILONASIVE TVCOOVE AT TEASONARE FUSO TIZIN ANDY TOBE SEACE CTISBE ...`

And finally here an example of a third-order approximation in which the constraints are taken into consideration that give the letter that follows two previous letters its appropriate probability in English:

(iv) `... IN NG IST LAT WHEY CRATICT FROURE BIRS GROCID PONDENOME OF DEMONSTRURES OF THE REPTAGIN IS REGOACTIONA OF CRE ...`

Again, in my opinion, it is not only quite surprising, but also very significant, that by imposing only a slight constraint structures are generated that resemble each other to a certain degree. Consider, for instance, the last example which contains not only several English words, it also suggests at least two delightful specimens which I think should indeed be in our vocabulary as, for instance, "Pondenome" or "Demonstures".

Since the letters of the alphabet are undoubtedly somewhat artificial building blocks of a language —

"phonemes" would have been a more appropriate choice for linguistic particles — I shall give you finally two examples of a slightly loaded random sequence where individual words are the distinguishable states of the universe. The constraints on the randomness of choice will be a second-order approximation to word order in English in the following example (v) and a fifth-order (7) approximation in example (vi):

(v) ... THE HEAD AND IN FRONTAL ATTACK ON AN ENGLISH WRITER THAT THE CHARACTER OF THIS POINT IS THEREFORE ANOTHER METHOD FOR THE LETTERS THAT THE TIME OF WHO EVER TOLD THE PROBLEM FOR AN UNEXPECTED ...

If you consider this to be a fair representation of "Beatnik" poetry, the next example with its somewhat higher constraints may be regarded as an attempt of much higher literary aspirations:

(vi) ... HOUSE TO ASK FOR IS TO EARN OUR LIVING BY WORKING TOWARDS A GOAL FOR HIS TEAM IN OLD NEW YORK WAS A WONDERFUL PLACE WASN'T IT EVEN PLEASANT TO TALK ABOUT AND LAUGH HARD WHEN HE TELLS LIES HE SHOULD NOT TELL ME THE REASON WHY YOU ARE IS EVIDENT ...

Again one should bear in mind that these word sequences are generated by a random device being constraint only to variously loaded probabilities in the transitions from one state (word) to the next state (word) corresponding to transitions as they appear in natural language. In the second parcel of my presentation I shall return to these examples because they will be useful — I hope — in showing the necessity for structural similarity between repre-

[text missing]

[proba?]-bilities which controlled the *transitions* from one word to the next, and so on. These examples, referring to temporal neighborhood relationships in sequences of structured events, have, of course, their counterparts in spatial neighborhood relationships in geometrical structures as we have seen earlier. Our constructs of Figure 3 were nothing else but geometrical representations of sequentially ordered events. Since, for instance, in a linear design each point has two immediate neighbors, it is the geometrical relationship of at least these two neighbors for all points that determines the shape as whole. With this observation we have an important clue as to some of the constructional principles of our automata, because we may now look for some means to incorporate the appreciation of neighborhood relationships into our systems.

I shall now turn to my second point in my first parcel, and shall discuss the inner workings of my automata, or "systems" as I called them in Figure 2. On the agenda now is a description of how these systems perceive some of their environmental structure and how they manage to make some inferences about the constraints that are responsible for the emergence of these structures.

I shall rid myself of this task by attempting to give you in the shortest feasible time the best account that I am able to give today on the structure and function of man's central nervous system and its remarkable constituent, the single nerve cell, because I would not know of a better system that resembles the kind of automata I have in mind. Aha, you may say, here I was talking big, promising you to describe marvelous little gadgets that can do all sorts of interesting tricks, and now I turn meekly to one of the greatest miracles, to man himself, to help me out of the predicament. No, Ladies and Gentlemen, I have not the slightest intention of avoiding the issue, because I may assure you that within the next half hour it will be rather impossible to account fully for the structure and function of a system as intricate, complex and magnificent as man — besides nobody knows how it works — but that little part which we know is quite sufficient to guide us in constructing automata of considerable sophistication. Since these various functions that are known to us can be put into precise qualitative and quantitative terms, we have a blueprint for the realization of such functions in electronic components and nothing — except the rather staggering expenses that are involved in coping with the sheer size of such systems — can stop us from actually building them. Today, however, these automata will populate my presentation only in form of a "Gedanken-Experiment". But since they are constructed only according to known features of our nervous system, they will help me in avoiding a rather popular pastime, namely to trade in ignorance for nonsense as it is frequently done by introducing pseudo-scientific concepts as, for instance, "Ego". "Id", "Subconscious", etc., which are all examples of pathological grammar of syntactical monstrosities.

A central position in my description of the nervous system will be taken by the nerve cell — or neuron for short — and I have to apologize to those of you who gave me the honor to listen to my presentation of this topic on other occasions[6] [7], if I am going to repeat some of the points that are already familiar to you. As you will see in a moment, this cell is in itself a highly complex automaton, capable of many more operations than the simple electron-tubes or transistors used in modern high speed electronic computers.

After half a century of intensive study of the neuron, its physiology, topology, chemistry, electro-chemistry, molecular structure, etc., we still cannot say today that

we know precisely all the features of this fascinating elementary organ. However, we can say today, that we are in possession of different levels of approximations of its structure as well as of its function, and in the last few years models of these approximations have been realized in electronic "hardware" which opened new vistas for physiologists, psychologists and for inventive and enterprising engineers.

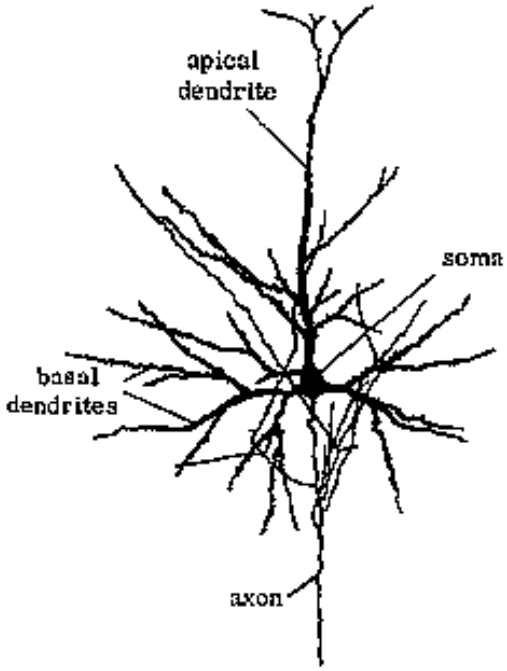


Figure 4. Cell body, dendrites (upwards) and axon (downwards) of a cortical neuron.

Figure 4 shows one of the several billion neurons in the outer folds of a cat's brain[8]. The big blop in the center is the cell-body proper, the "soma", which houses the cell's nucleus and is probably responsible for the neuron's metabolic activity. The same membrane which envelops the soma forms also the tubular sheaths around the many ramifications extending from the soma. There are two kinds. One kind is seen branching off in all directions in a tree-like fashion, the "dendrites". The other one, "the axon", is smoothly surfaced and rather straight. It extends downward, bifurcating further below on many points, exhibiting a more regular, somewhat perpendicular pattern. The diameter of the axons may vary from a few microns to hundreds of microns, its length from a few millimeters to a meter or more. Most of these axons terminate on other neurons and establish two different kinds of connections as sketched in Figure 5. One is a direct attachment to the soma of the other cell by formation of an "end bulb", the other one is a somewhat haphazard ascending intertwining with the dendritic ramifications of the target neuron. If one penetrates with an electric micro-probe the enclosing membrane at any point of the neuron, one finds a change of the electric potential of somewhat less than a tenth of a volt, which indicates that the whole structure in its rest state is a charged, distributed electric battery.

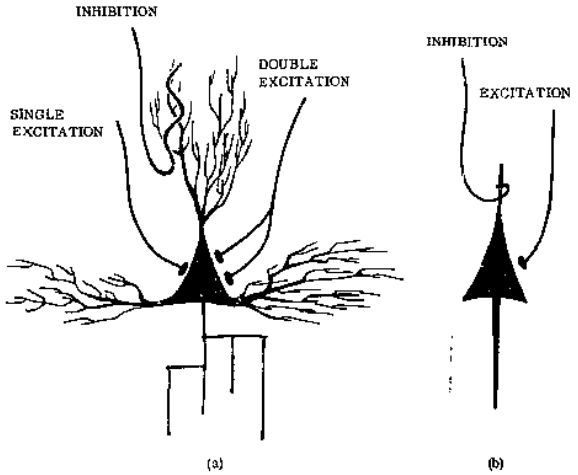


Figure 5: Schematic (a), and symbolic (b), representations of a single neuron.

If at the soma this electric potential is momentarily perturbed beyond a certain threshold value, the neuron will "fire", i.e., the perturbation will travel along the axon in form of an electric pulse and will pass on this perturbation to its connectees which, in turn, will respond to this perturbation by producing again an electric pulse. You may note the following interesting details in connection with this interaction process. First, the magnitude of the electric pulse is independent of the size of the perturbation. However, a prolonged perturbation in sensory cells will produce a chain of pulses[9] with a repetition frequency approximately proportional to the logarithm of the intensity of the stimulus (fig. 6a). Thus, frequency modulation has long been employed by nature, before it was discovered as a noise-evading method for transmitting signals, and the neat little trick of coding intensity into its logarithm reduces not only the problem of multiplication to simple addition, but also compresses a wide range of intensities into a narrow band of measurable quantities. However, after the frequency modulated signal that comes from our sensory receptors has been "processed" by a variety of layers of neurons that received these incoming signals, the pattern of electric activity is not any more as simple as we observe it immediately after the sensors. If one probes deeper into the

brain and picks up activity of single nerve-cells complicated, but still periodic, pulse activity is observed (fig. 6b). I would like to stress at this point that we have to make it utterly clear to ourselves that every thought that crosses our mind, every motion that grips our heart and every impression we obtain from the outside world is solely represented by the pulse activity of our neurons which is the only language within our system that is spoken and understood by all nerve cells.

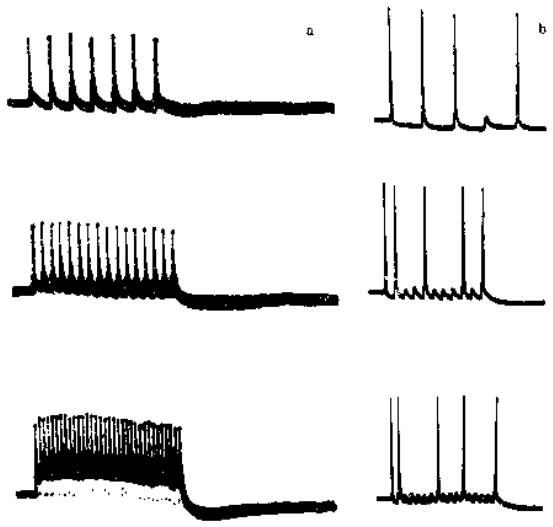


Figure 6: Electrical activity measured with a micro-probe on the axon of a sensory (a) and a cortical (b) neuron. (a) Frequency modulation, (b) pulse interval modulation.

Another important point I would like you to note is that the magnitude of the traveling pulse does not diminish while it travels along the axon, even if a bifurcation is reached and from then on two pulses travel along their tracks. This is, of course, due to the cleverly distributed battery which supplies the necessary energy at any point along the line. Hence, a set of neurons in series may act as an impressive signal amplifier.

The last point which has to be made with respect to the interaction of neurons is that the two types of connections as sketched in Figure 5a fulfill two different kinds of functions. Connections via end bulbs provide units of excitations, while axons terminating in the dendritic ramifications will inhibit the firing tendency of the target neuron. In other words, if two pulses arrive almost simultaneously at a nerve cell, the one over an end bulb, the other one over the dendrites, their actions will cancel out and nothing will happen. If, however, at a particular instant the only active input is a single end bulb and the threshold of the cell is zero, this will suffice to trip the neuron and it will transmit a pulse via its output axon to other neurons in its connection field.

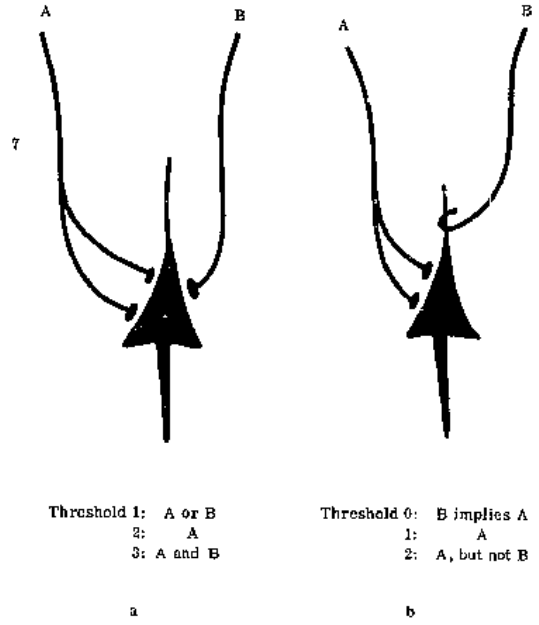


Figure 7: Single neurons computing a variety of logical functions.

It was precisely this functional property which suggests the possi- bility of an idealization of the function of a neuron in form of a logical operation, the affirmative represented by excitation, negation represented by inhibition. Figure 5b is a symbolic representation of this situation, the triangle standing for the soma, its single upward extension indicating dendritic ramifications with the loop around it symbolizing the inhibitory axon. Excitatory inputs have unmistakably the characteristic end bulbs. Depending upon the symbolic neuron's threshold (which is assumed to change in unit steps) and the arrangement of the input fibers, one or more of these idealized elements are capable of computing all possible "logical functions". In Figure 7a, for instance, for a threshold of one unit the neuron will fire when fiber A is active, or when fiber B is active or when both A and B are active, hence the logical function "A *or* B" is computed. Raising the threshold one unit, it will fire always when A is active irrespective of the activity of B, since the logical function "A" is computed; finally if the threshold is three units, only the simultaneous arrival of pulses over A and B will trip the neuron and the function "A *and* B" is computed. Similarly in Figure 7b the functions "B *implies* A", "A" and "A, but not B" are computed if the threshold moves through the values 0, 1 and 2. Networks of these idealized neurons can be made capable of computing more and more complicated logical relationships, and, as McCulloch and Pitts[10] showed in their celebrated paper, "A Logical Calculus of the Ideas Imminent in Nervous Activity", any functional behavior

which can be defined logically, strictly unambiguously in a finite number of words can also be realized by such a formal neural network.

In my opinion this result is perhaps one of the most significant contributions to the theory of knowledge in the last half century. What are the implications of their findings? I believe that they are best summarized in the words of the late John von Neumann[11], the mathematician. I quote: "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."

I am sure that you have recognized in this citation by von Neumann a paraphrase to my earlier assertion, namely, that we are today in a position to build — in principle — automata that will do everything that is appropriately specified by us. We only have to equip this automaton with the kind of formal neural network which corresponds to the tasks the automaton is required to perform. This puts an end also to the discussions of whether or not we can build a machine that can, for instance, compose music à la Mozart. The answer is that such a machine — in principle — can be built if somebody says what "à la Mozart" means. This is not an evading of the answer. It is, in my opinion, a challenge to us to come up with more precise specifications. If we mean "in the style of Mozart" then we are compelled to specify this particular style. If we mean "music as beautiful as Mozart's music" we have to define "beauty" in the special sense of Mozart's beautiful music, etc. Hence, having these automata in the back of our minds, we are compelled to formulate our ideas in precise and rigorous terms, a task that is in its difficulty comparable to building these machines. However, I believe it is worthwhile to undergo this purification process, because ambiguity in symbolic representation is as dangerous as misrepresentation of environmental features in a single system. As we shall see later, two communicating systems that cannot converge to the same interpretation of a symbol will be eliminated by the harsh rules of evolution in the same way as a single system is doomed to extinction that makes false interpretations of certain features of its surrounding universe.

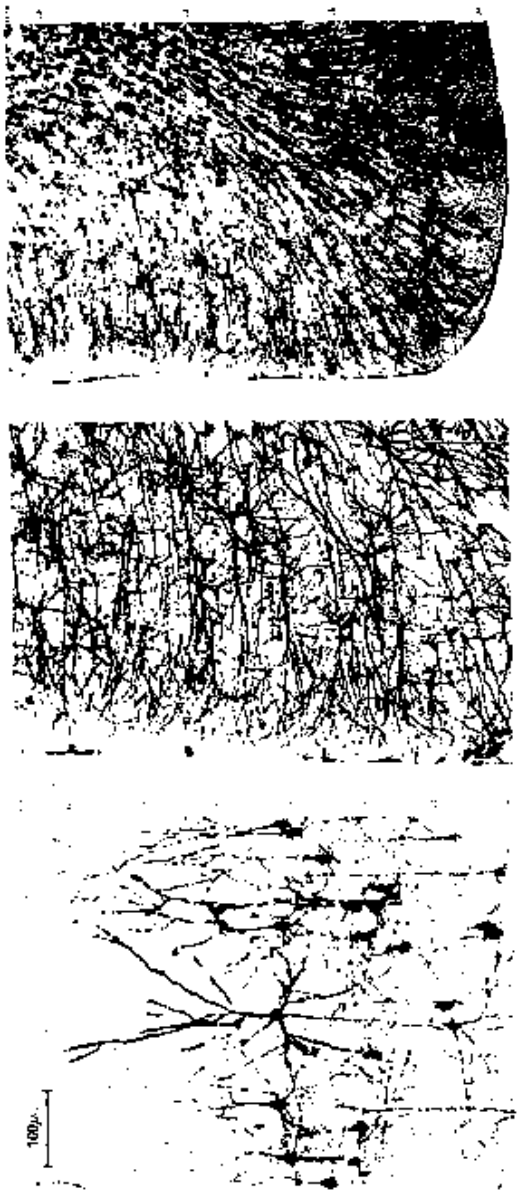


Figure 8: Figures (a) (b) (c) show with decreasing magnification the structure of neural nets in the cortex of a cat.

Let me show you now a few examples of physiological nerve nets as they can be studied on microscopic slides. Figure 8a is an enlargement of a small region in the cortex of a cat[12]. The structure that stands out most strongly is a single neuron with its cell-body and a few dendritic ramifications extending upward. It is surrounded by many other neurons with which it may or may not have communicative connections. I may point out that these neurons are made visible by a refined staining method which, however, stains only about 1% of the

neurons in this region. You have, therefore, to imagine a dense jungle of such neurons, approximately a hundred times denser as this picture indicates. This accounts for the extraordinary large number of neurons which can be squeezed into a relatively small container such as our skull. We carry in our brain approximately ten billion neurons, each of which is a most sophisticated computer element. In order to give an idea of the expenses that would be involved in just manufacturing these complex components, I am giving you the latest estimates of a single electronic unit that would do all the tricks we know of that the physiological neuron does. Considering present state of the art, comparison shopping indicates that a single neuron would run to about $125, that is as much as one has to pay for a good FM-Transmitter-Receiver. The sheer componentry of a "single brain" would cost us today about 1250 billion dollars — not to speak of the labor costs that are involved in connecting this Mount Everest of gadgetry. Figures 8b and 8c are magnifications of decreasing power of the same cortical region of the cat, giving you more and more of the whole structural setup. Although you may recognize in 8c a certain directionality and order in the way in which the fibers are aligned, two of the most basic principles of the operation of these nets had to be established by measurements with electric microprobes.

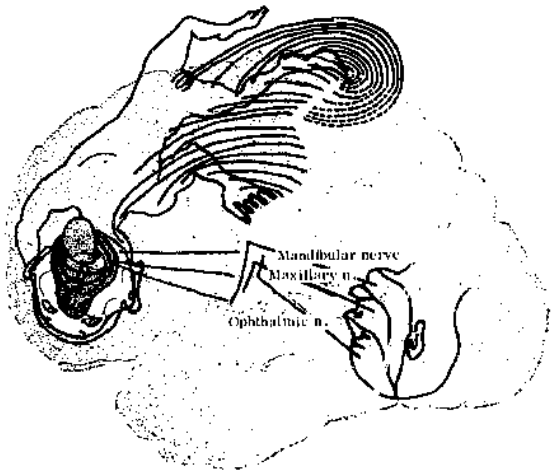


Figure 9: "Homunculus", Sensory neighborhoods are mapped into cortical neighborhoods. Topological mapping of the sensation of touch into the outer layer of the brain.

One of these operational principles is a mapping of neighborhoods into neighborhoods, or the principle of "topological mapping". We speak of topological mappings whenever we can set up a continuous one-to-one correspondence between, say, an "object" and its "image", although, by no means, the image has to resemble the object in an obvious way. All geographical maps are, of course, topological mappings of some terrestrial features, all projections and all deformations we may obtain from painting figures on a rubber balloon and squeezing afterwords this balloon into the craziest shapes. Obviously, in all these transformations neighborhoods will map into neighborhoods and continuity is not destroyed. From my previous remarks on the manifestations of constraints in form of certain relationships amongst temporal or spatial neighbors it is clear that a transformation from one representational modality into another representational modality that does not destroy neighborhoods will still permit the analysis of the original neighborhood relationship and thus will allow the eventual reconstruction of the operating constraints. In many instances our nervous system indeed performs such topological mappings, particularly when sensory information is projected into the deeper regions of the brain where this information is further processed and reduced. A typical example is the topological preservation of our body with respect to the sensation of touch in the appropriate regions of the brain as you may see in Figure 9. This "Homunculus" is obtained[13] by registering with microprobes those regions in the brain which become active when certain regions on the body are stimulated. Thus, we carry in our brain a "signal-representation" of ourselves, with more or less emphasis upon those regions to which evolution allows us to pay more or less attention. Of course, this signal representation is only a preliminary step in further abstractions to which I would like to turn now your attention.

I am referring to the second basic principle of operation of nerve-nets. It is only a couple of years ago that Lettvin[14] and a team of neuro-physiologists established this principle beyond doubt in a series of brilliant experiments. Let me briefly describe the experimental set-up. A frog is tied to a small stand, with his head fixed in a certain position. He is surrounded by a white, hemispherical horizon covering his entire visual field. The illumination of this horizon can be varied, strong shades of objects of various configurations can be projected on this horizon and these shades can be moved about. The frog's response to these optical stimuli is measured in two ways. His muscular action is recorded from the stand to which he is attached. His neuro-optical response is measured by microprobes which are inserted into single fibers of the optic stalk which is made up of the bundle of nerve-fillers leading from the frog's retina to its "brain". Exposing the frog's visual system to a variety of stimuli, Lettvin and his colleagues found most interesting results. Maintaining the micro-probe in a particular fiber, response was only elicited when a certain kind of stimulus was pre-

sented, for instance, when the light was suddenly turned out. This particular fiber remained silent under strongest exposure of light, fast illumination, movement of objects, etc. However, when they moved into another fiber, no response was elicited when the light was suddenly turned out, but strong responses were obtained when, for instance, the shadow of a straight-edge appeared in the visual field. Again this fiber remains inactive for all other kinds of stimuli. In a delightful article entitled "What the Frog's Eye Tells the Frog's Brain", Lettvin and his colleagues reported their findings. Moving from fiber to fiber they found that "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 movement or rapid general darkening."

At first glance these "operations" appear to be quite mysterious. Are there special fibers which are sensitive only to edges, curves, or other geometric properties? No, this is impossible. Because, first, these properties are distributed properties, revealing themselves only when neighborhoods are inspected; second, all neurons are more or less alike and no specificity for certain stimulation can be explained.

It was perhaps an accident that about the same time when these observations were made, we in the Biological Computer Laboratory of the University of Illinois[15] were in the midst of a comprehensive investigation of the computing capabilities of large networks particularly so-called "periodic nets", which are characterized by a periodic repetition of one and the same connection pattern, hence by a repetition of one and the same computational operation. The study of these periodic structures proved to be most rewarding and led to important clues in our understanding of the problem of "cognition" not only in the frog, but also in higher animals, including man.

Figure 10a gives an example of such a periodic one-dimensional net. A series of photosensitive elements are connected to a series of corresponding idealized neurons such that the left and right hand neighbor neuron is singly inhibited, while the associated neuron is doubly excited. This connection scheme repeats itself periodically over the entire strip. With all thresholds equal and slightly above zero, clearly, if all photo cells are uniformly illuminated, the neural net will not respond, irrespective of whether the illumination is strong or faint, because the double inhibition converging on each neuron from photo cells to the left and right will cancel the double excitation coming from its corresponding photo cell.

An obvious effect of this particular inter-connectivity is its insensitivity to variations of light, despite the fact that the "sensory layer" — the photocells — are all highly sensitive light receptors. You may ask now, why all this effort to make a light sensitive organ insensitive to light? However, you will see in a moment the interesting feature of this net. If now an obstruction is placed into the light path (fig. 10b), the edge of this obstruction will be detected at once, because the only neuron which will now respond is the one on the edge of the obstruction, receiving insufficient inhibition from only one of its neighbor photocells in the light, while the other one is in the shade and silent. In other words, this net "computes" the environmental property "edge", independent of its location and independent of the strength of illumination. The efferent fibers of this network will be active only if edges are present in the visual field of this one-dimensional "retina".

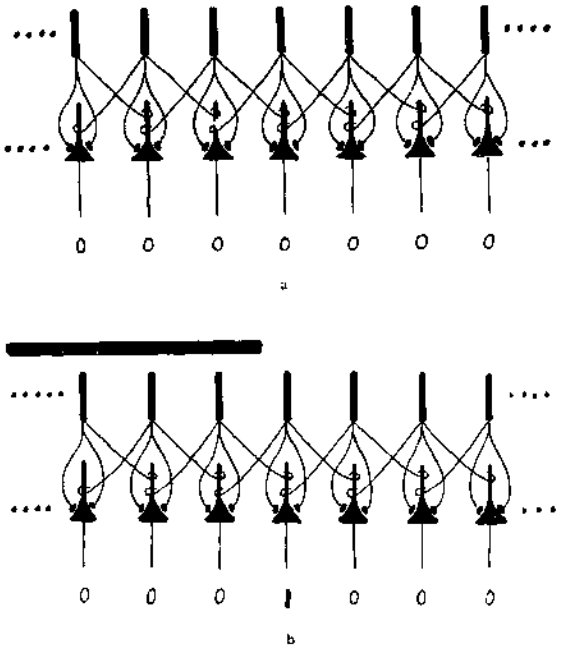


Figure 10. Periodic neural net (a), which computes the environmental property "edge" (b), independent of location and size of object, and independent of strength of illumination.

Although this network is admittedly simple, it can detect an environmental property which cannot be detected by the nervous system built into us. Consider the simple fact that any finite, one-dimensional obstruction must have two edges (fig. 10b). If N objects obstruct the light path to our edge detector, 2N neurons will be active and their total output divided by two gives exactly the number of objects in the visual field of that retina. In other words, this strip sees each different number of

objects as a different entity, say, "seven-ness", "twenty-ness", etc., as we see different electromagnetic frequencies as different colors "red-ness", "green-ness", etc. An electronic model of a more sophisticated version of such a counting network I have exhibited in the hall. It is the "Numa-Rete" which I had mentioned already in my introductory remarks (fig. 1). Its small receptor field consists of a quadratic array of 20×20, that is 400, light sensitive cells each of which produces an electric potential when illuminated. Compared to the human eye, this is a very small sensory organ indeed, because we count in the human retina over five million such light sensitive cells: the cones and the rods. While in the human eye a multitude of neighborhood relationships of shapes that are projected on the retina are computed by the immediately following post-retinal neural networks, in our Numa-Rete the "neighborhood-logic", i.e., the mode of interaction between neighboring elements in the computer net that follows the sensory layer, computes only one geometrical property presented to the stimulus field of the sensors, namely, "having at least one neighbor that is also not illuminated". If cut-outs of various geometrical shapes are placed on this "retina", all cells covered by these cut-outs form — so to say — a coalition which responds in unison if only one cell of this coalition is interrogated by a scanning pulse. Since each coalition responds only with one "here" when asked, the number of replies is identical with the number of shapes placed on this retina, independent of shape, location and size of these shapes and independent of the level of illumination. Such a process is called an "abstraction" because it abstracts from the stimulus field which contains plenty of information regarding size, color, shape, roughness, illumination, etc, etc., only *one* property, "connectedness", which is invariant to all variations regarding the previously mentioned environmental properties. Hence, we say that an "abstractor" computes an "invariant", or an "abstract". We are today in possession of a general theory of the required interaction between neighboring elements in networks so that these networks compute an almost infinite variety of abstracts as, e.g., straightness, curvature, topological connectedness, motion of shapes, flicker, etc., in the visual field; chord and timbre independent of pitch, voicing and variation of frequencies typical for definition of spoken phonemes in auditory perception, etc., etc.[16][17]

With this brief detour into the gadgetry of cognitive systems I have concluded my narrative of the contents of my first parcel. In summary we have seen that cognitive systems that exist in an environment with constraints compute these constraints with the aid of appropriately constructed networks. The result of these computations are "abstracts" that are linked to the constraints by the structure that is generated in the universe and also perceived by the system. The information about presence or absence of some environmental features is transmitted in these systems in form of a universal language whose symbols are periodic or aperiodic trains of electric pulses. As we understand it today, the sensory apparatus is not "like a camera" that just projects a picture of the world into the brain, but is a highly sophisticated computer system that transmits to the brain selective, structural information. Since at each operational level in this system the "input" can be taken as the "environment" for the higher operational levels, information of the "self-state" of the system can be processed equally well (Homunculus). In the normal person information of "self-states" are usually coded such that they are distinguishable from true environmental information. In some pathological cases, however, we know that this distinction collapses. In hallucinations the patient, for instance, "hears" that invisible chickens are talking to him, etc.

I am going to embark now into the description of the contents of my second parcel that deals with the interaction of automata in an environment with constraints. By interaction I am, of course, referring essentially to interactions for the purpose of communication and not, as it may be interpreted, in terms of these fellows just bouncing into each other, or kicking each other in a fight for food or for the charms of a third automaton. Although these latter forms of interaction are doubtless special cases of communicative endeavors.

Although the problem we are facing is one of much higher order than the previous one, because now we have to deal with "representations of representations" of structures, we are well prepared to cope with the new difficulties, because most of the basic concepts I am going to employ, I have explicated already in my earlier remarks.

Let me again open the discussion with a diagram that may help to illuminate the situation (fig. 11). In contrast to Figure 2, in which I had only one system that was in close contact with Universe U, I have now two systems S_1 and S_2 which receive information (arrows) from the environment which is common to both. Of course, the interaction mode of these systems with their environment, which is symbolized in this diagram by a single arrow only, can be replaced for each system by the information flow diagram of Figure 2. Again, in contrast to the simple situation of Figure 2 in which our system was faced only with a universe with given constraints, now each of our systems is confronted with the additional complication that he sees his environment populated with at least one other automaton that generates events in the Universe U. Hence S_2 sees in addition to the events generated by Uni-

verse U those generated by S_1, and since these events take place in U, I shall label them U_1; aid, conversely system S_1 sees in addition to events generated by U those generated by S_2 which I shall label U_2. Hence, the surprising upshot of this situation is that in spite of the fact that both, S_1 and S_2, are immersed in the same environment U, each of these systems sees a different environment, namely S_1 has to cope with (U,U_2) and S_2 with (U,U_1). In other words, the situation for both systems is asymmetrical, with U being the only symmetrical part.

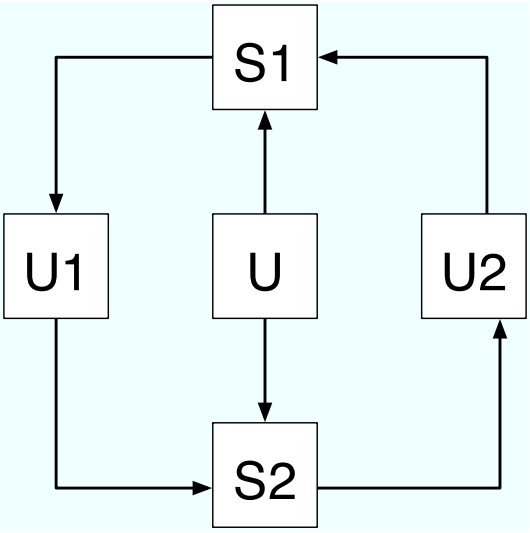
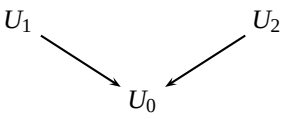


Figure 11. Diagram of information flow between two "systems" coupled to the same "Universe".

In anticipation of what I shall try to discuss in more detail in a moment, let me become a bit more specific about this asymmetrical situation. Assume for the moment that U_1, U_2, "the events generated" by S_1 and S_2 respectively, are in some way or another representations of some structural features of the Universe U. Or, to put it more dramatically, assume S_1 "says" something or "gesticulates" something that is intended to convey some environmental features known to S_1. This activity I have called U_1. Now assume S_2 does likewise and represents the same features by U_2. It is intuitively clear that these two chaps will have a heck of a time to figure out that they are talking about the same thing unless — and here comes the crucial point — they succeed in adapting their representations to a common denominator, or — in other words — if they succeed in eventually converging to like representations for like universal features. This process may be symbolically expressed by



where the arrows indicate the convergence process, and U_0 stands for the final, universal "language" spoken by both systems. You may note that when this paradisical situation obtains, the initial asymmetry for our two systems ceases to exist, because both, S_1 and S_2, see the same universe (U,U_0). Evolutionary processes that acted selectively for the last two billion years on the development of our brain have succeeded to a certain extent to let our neurons speak a universal language. With only a few interruptions man has attempted for the last 500,000 years to proceed along similar lines to improve his powers of communication. One interruption for instance, is reported in Genesis 11:1–9, when God punished the would-be builders of Babel's tower by suddenly letting them speak in different tongues. Again today, He seems to have similar ideas in mind by sending us the "look-say" boys, the "New Webster Dictionary" boys and the "Madison Avenue" boys. Only this time He is so kind as to punish only those who accept their gibberish.

The rest of my time I shall devote to the process of convergence into a universal language. Since I believe this process to be of considerable importance in our understanding of the basis of communication and its non-trivial consequences, let me recapitulate the idea with the aid of Figure 12. Let us assume that certain environmental features (three people) in our Universe U are observed by two systems S_1 and S_2. Both systems carry now an internal representation of these features. I indicated these representations by sequences of pulses in order to remind you of the electric pulse activity of the neurons. Our two systems, in turn, proceed now to represent these observations in form of external activity, which in both cases maintains some of the environmental structure, but the representation is different for S_1 and S_2, namely, U_1 and U_2, respectively. At this point commences the convergence process which will bring about three major effects:

1. A *universal* representation of environmental structures.
2. A *unique* representation of environmental structures.
3. A transition from representation to *symbolization*.

Since I am going to catch these three impressive flies with only a single stroke — the "convergence process" — let me give you, at least, a few plausibility arguments that this process has indeed a considerable chance to succeed. The first argument involves the particular way in which the information about the environment is chased around and around between the two systems S_1 and S_2. This is best seen in Figure 11 if you follow the arrows from S_1 to U_1 to S_2 to U_2 to S_1 and so on and so on. However, at each instance of representation, the systems

have a chance to compare their representations with the universe, and in the case that, say, S_1 registers deviations between its internal representations of U and U_2, he may give a correcting signal to S_2. Likewise, S_2 may deliver such a signal to S_1, which causes S_2 in turn, to make further corrections, and so on. In the engineering sciences, such a circular causal loop is called a "feed-back" system, because the information about certain deviations are "fed back" to the system which is now in a position to react so as to reduce the deviation. I may remind you of the "volume control" in your radio that is a gadget which measures the power that is delivered to your loudspeaker and compares it with the value that you desired when you set the knob into a particular position. If, for some reasons, the deviation between actual power and desired power is positive, this deviation turns down the amplifiers until the deviation vanishes; and, conversely, if the actual power fades away, the negative deviation turns up the amplifiers until again the difference ceases to exist.

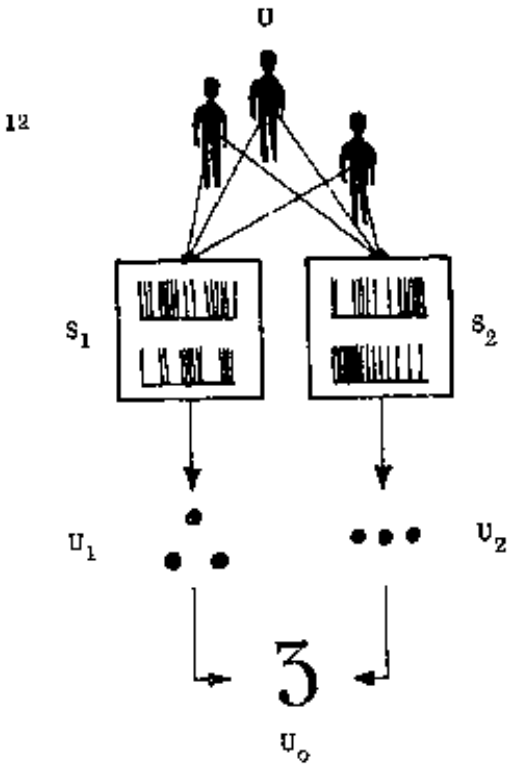


Figure 12: Diagram suggesting the transformation of environmental structures into internal representations, external representations and, finally, their convergence into symbols.

"Cybernetics" a whole new science has such control mechanism at the core of its study, and I can only report to you that we know that these systems, after a few oscillations (which you barely hear in a good radio), converge indeed to a fixed value.

I shall now present another plausibility argument about the convergence of representations made by our two systems which is based essentially on the observation that our systems, immersed in the same universe, are also faced with the same constraints in expressing their internal representations. Since this is supposed to be a plausibility argument only, let me forgo all mathematical and generalized rigor, and let me explicate this idea on a concrete story[18].

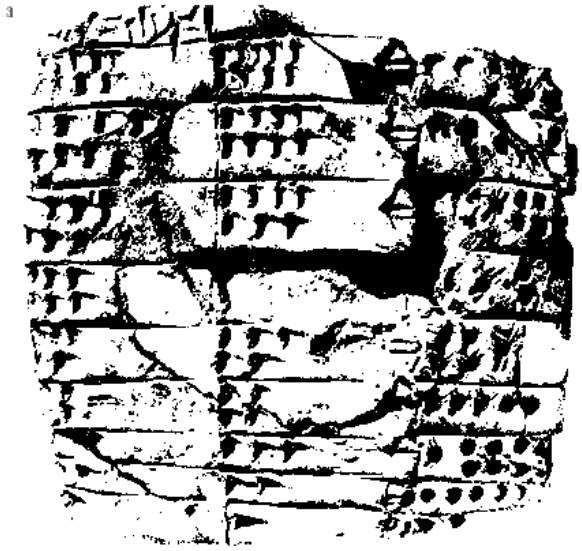


Figure 13: Clay tablet of a Sumerian mathematical table of the fourth millennium B.C.

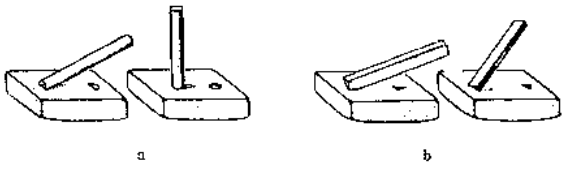


Figure 14: Development of styli for writing Sumerian numerals. Ancient (a) and "modern" (b) forms.

My story begins approximately 5500 years ago when in the lower Euphrates River Valley a Sumerian scribe prepared the mathematical text reproduced in Figure 13. The numerals representing the appropriate numbers are "written" by pressing a cylindrical stylus into soft clay. Presenting the numbers N between 1 and 9, the stylus is held in a tilted position and N impressions are made. The middle column in Figure 13 reads these numbers from bottom to top. This method of writing was the standard procedure for Sumerians at that time. Clay was abundant, dried quickly in the hot sun, and the inscribed tablet could be nicely preserved. Here the physical properties

of clay, being soft and malleable when wet, and hard and persistent when dry, the properties of wood which can be cut into convenient shapes, etc., all these properties define the constraints that shape the modes of representation for our "systems", that is the Sumerians, who are immersed in the same environment, the lower Euphrates River Valley.

If these tools are very frequently used, as we have ample reason to believe, it becomes quite obvious that a cylindrical stylus can be used differently, for instance, by pressing it perpendicularly into the clay, and also, that the form of the stylus itself may be changed in order to give impressions of different character. In Figure 14 I give you the two types of Styli as they probably evolved in Sumer between the fourth and the third millennium B.C., the stylus with triangular cross-section being the later development and giving rise to the type of writing we call "cuneiform". A calligraphical masterpiece, and by the way an important mathematical text, is reproduced in Figure 15 which is estimated to have been prepared in the second millennium. I have chosen these examples to show how strongly the constraints inherent in the tools of expression define the structure of the whole "language", if you take language in a general sense as the totality of the vehicles of communication. However, there is another point which can be made with these examples, namely, that by changing the modality of representation, "that which is referred to" may also be changed. In the early mathematical tablet (fig. 13), use has been made of this possibility. There are two kinds of impression, the one kind is made with the tilted stylus, the other kind with a stylus held in a perpendicular position. We know today that N impressions of the latter form are numerals standing for numbers 10 N, for example, four round holes is the numeral for 40, etc. To put it even more concisely we may say that N "flat holes" stand for $N \times 1$, while N "round holes" stand for $N \times 10$. Clearly, the number of holes *represents* in both cases the number of entities which it is supposed to represent, but by changing the modality of representation "flat hole" $\longrightarrow$ "round hole" one changes the represented entities: "ones" $\longrightarrow$ "tens". However, I venture to say that not even the wildest imagination will be able to dream up a representational correspondence between

$$flat\ hole \iff one$$

$$round\ hole \iff ten$$

In this case representation has been abandoned and non-representational symbolization has begun. Since in the structure of these holes there is absolutely no hint as to their "meaning", it is purely by convention that a correspondence between a symbol and what it stands for can be achieved. But convention is working only if it is universally accepted. A symbol is a riddle if it is not explained. And every explanation is a further extension of its universality. I shall return to this point in a moment. Presently I would [like] to show first the gradual transition of representational "pictograms" to highly stylized representational forms (fig. 16.) This transition is believed to have taken place in the two millenia of Sumerian [cultural?] activity between 4000 and 2000 B.C. As we go down the various rows it is clearly seen how the constraints imposed by the writing tools strongly modified the early pictograms given in the top row. Simultaneously with the departure of structural representation goes an increase in the possibility to represent various entities in connection with appropriate modifiers. While the pictogram at the top of the right hand column indeed says "foot", after two thousand years of stylization (bottom row) it may say "walking", "running", "delivering a message" and, probably "playing footsie" if associated with other signs which give it the appropriate modification. Nevertheless, with sufficient imagination, the original representational forms of early times may be traced back from later representations.

However, around the turn of the second millennium B.C. this situation changed drastically. A new people with an entirely different language and culture moved into the Euphrates River Valley, absorbing and utilizing whatever they found useful, eventually supplanting and eradicating the Sumerian ethnic element. As it is mostly the case with great conquerors, they do not suffer from too much educational ballast. In this particular case they could neither read nor write. In order to show that this fortunate circumstance provided a turning point in the history of mankind, because in one stroke, it transformed the whole body of Sumerian writing from a representational formalism into a symbolic algorithm, I have to explain their linguistic differences.

The Sumerians, a non-semitic people, spoke a language similar to Chinese, with monosyllabic nouns and modifiers "RAS", "TA", "PUM", etc., standing for entities as, perhaps e.g., "geometry", "palace", "cocktail party", etc. The invaders, a semitic people, who established themselves later in Babel, hence referred to as Babylonians, spoke a multi-syllabic semitic language similar to Hebrew, where inflections of nouns and verbs are generated by insertion of vowels between three consonants which define the root of the word. For instance PRS is the root for aPaRas "I cut", iPaRas "he cuts", iPRuS "he has cut" etc.

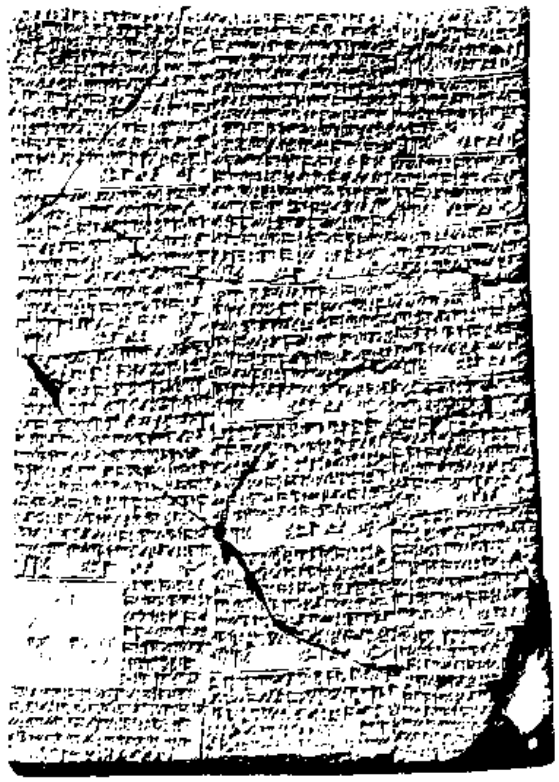


Figure 15: Babylonian mathematical text of the 16th century B.C.

From the Sumerians they learned the remarkable feat that sounds can be written. To take the above artificial example, they learned that geometry which is pronoun ed RAS, is representable by some cuneiform combination. Why not use this cuneiform combination to write the last syllable of iparas? Indeed, this is what they did. With plenty of mono-syllabic sounds available in Sumerian, almost all Babylonian words could be transcribed by this phonetic method. It is to be noted that no attention was paid to the Sumerian meaning of these sounds. As a consequence, a certain cuneiform combination that sounded when read i-pa-ras, meant "he cuts" in Babylonian, but may mean in Sumerian "horse – father – geometry".

It is clear that with this forceful transition of the meaning of signs, all connections with their original structural representation were lost, and these signs assumed the detached character of symbols. With this transition, I believe, the power of abstract symbolic operations was discovered, and the rapid development of mathematics, astronomy and law during the Baby- lonian reign may support my conjecture.

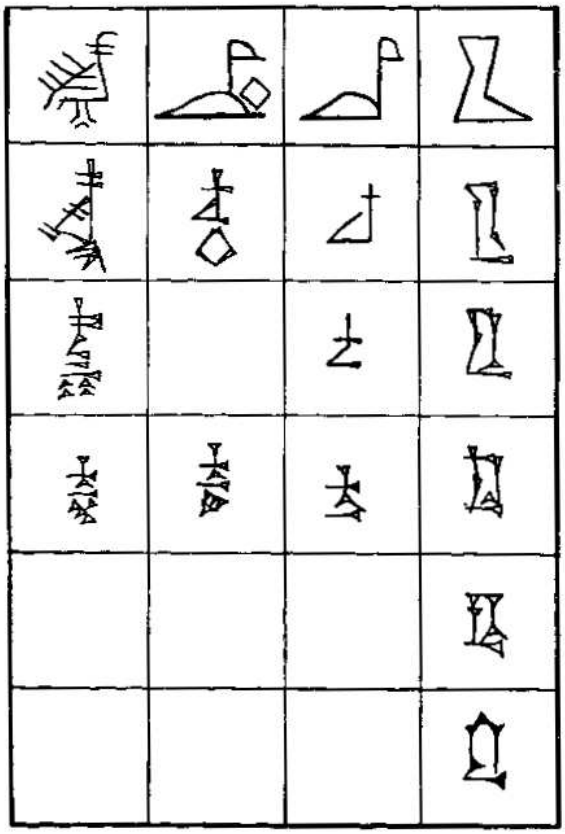


Figure 16: Formalization of pictograms through constraints imposed by writing tools. This development is estimated to have taken place in a period from the 5th to the 3rd millenium B.C.

At this point it seems that I deny symbols all reference to structure. Certainly, I do not hold this position. I maintain, however, that the symbol-structure relationship manifests itself in a rather abstract and not so obvious way, because symbols carry rules of connectivity and not so much rules of entity. A symbol may be compared to an atom which can form compounds only with a small number of other atoms but remains inert to the rest of them. Take, for instance, these "molecular" sentences:

(i) Socrates is identical.

(ii) 4 + 4 = green.

The disturbing thing about these sentences is that they are neither true nor false; they are nonsensical. The connection rules of the symbols has been violated in these examples. "Identical" sets up a relation between two entities. "Socrates is identical with Plato" is a sentence that gives sense, although it happens to be a false proposition. The compound "4 + 4 = " requires a number to follow. Putting "6" at the end is fine! It's a good guess. But "green" is an operator with an entirely different struc-

ture. This indicates that somehow structure is still preserved in symbolical discourse, but in a syntactical and not in a representational sense. The language of symbols has, so to say, its own logical grammar[19]. Uniqueness in symbolic expressions is established in the same sense as pieces of a jigsaw puzzle can be put together in one, and only one, way. It is the neighborhood relationship amongst the various pieces — the symbols — that puts them into place.

With these remarks I have come to the end of my paper "Form: Perception, Representation and Symbolization". I have briefly touched upon the mechanisms of perception that filter out some structural entities in the universe. I have sketched the modes of internal representation of these structures and I have alluded to the constraints in the universe that again shape representations when projected externally for the purpose of communication. Finally I have attempted to make it plausible that an error-correcting circular information flow mechanism between two systems eventually transforms representations into unique and universal symbols whose logical connectivity preserves environmental order.

Nevertheless, permit me to give you, so to say, as an Appendix to my story an application of these ideas to a problem that has puzzled many generations of philosophers. I refer to the class of problems which usually go under the title "Theory of Values", that is the theory that copes with questions of the form "Why does X prefer A over B?"; "Is A better than B?"; "What are the conditions for A being better than B?", etc. Aesthetics, or the theory of the conditions that make A "more beautiful" than B, belongs into this class of problems. If my universe, populated with two automata is worth your time and patience in listening to me, this simple model should be able to throw some light on this perennial question.

First, let me eliminate one pitfall in discussing these questions, namely, the belief that values can be presented in a hierarchical order only. By this it is usually meant that all entities in competition can be finally lined up like horses on a race track, one being the best, another one the next to the best, etc. This idea has lead to the postulate of a *summum bonum*, an "ultimate best", which is supposed to be second to none. Alas, in actual choice situations this charming theory does not hold up to the facts. A man — and *a fortiori* one of our automata — given the choice between apples and prunes may choose prunes; given the choice between prunes and grapes may choose grapes; and given the choice between grapes and apples may choose apples. This "heterarchical" arrangement of values can be derived from the finitude of our systems. It has also been shown that this pattern of circular choices is a consequence of the organization of our nervous systems. Nevertheless, the "hierarchists" go on to advertise a *summum bonum*, upon which there can't be an agreement as we have just seen. This topic is therefore a truly inexhaustible source for lecturing, preaching and writing.

Next I have to remove another fallacy that creeps into the discussion of aesthetic values, namely, the belief that these values arc some objective properties of the universe. From all that has been said so far the nonsensicality of this proposition is obvious. If there is no observer in this universe, who is going to make the decision of whether or not A is better than B? Clearly, we need a system which possesses an internal representation of the structures A, B, C ... etc. Comparison is therefore made amongst *representations* of A, B, C ... and if our system, per chance, comes up with a choice, this choice must fulfill precisely the following requirement in order to be regarded as a choice at all: It must be communicable. This does not mean that this choice has to be communicated, it only means that, in order to finalize any choice it must, in principle, be capable to be witnessed. Hence, we need at least a second automaton S_2 that witnesses the choices of our aesthetic automaton S_1. What does our witness S_2 lean about the universe while listening to S_1's value judgments? Very little indeed. However, S_2 leans a lot about S_1.

Hence, you can see why I am so impatient to have my little automata finally built. Because I am going to ask them what they think about Pop-Art. Whatever they will say, you can be sure I shall learn a lot about my little automata.

Transactions of the Allerton Conference 1962 sponsored by the Society of Typographic Arts, Norman Perman, Editor. October 19–21, 1962.

Notes

[1]American Collegiate Dictionary. Harper & Brothers, New York (1948). (It may be noted that Webster's New World Dictionary, 1950, does not use the "as if.")

[2]Aristotle: *De Motu Animalium*. In *The Works of Aristotle*, Smith, J.A., and Ross, W.D., (Trans.): Vol. V, 701b–703a. Oxford, (1958).

[3]Von Foerster, H.: "Communication Amongst Automata", Am. J. Psychiatry, **118**, 805–871, (1962).

[4]Weston P.: "Photocell Field Counts Random Objects", Electronics, pp. 44–46, (September 22. 1961).

[5]Schopenhauer, A.: "Über die Freiheit des Willens" *Sämliche Werke*, Bd. **6**, p. 159 ff. A Weichert, Berlin.

[6]Von Foerster, H.: "Perception of Form in Biological and Man-Made Systems"; in *Industrial Design Education Assn.* E.J. Zagorski (ed), Superior Printing Co., Urbana, pp. 10–37 (1962).

[7]Von Foerster, H.: "The Logical Structure of Environment and its Internal Representation" in *Int. Design Conf.*, Aspen 1962, R. Eckerstrom (ed), Miller Corporation, Zeeland, Mich. (1963).

[8]Scholl, D.A.: *The Organization of the Cerebral Cortex*, Methuen ard Cic., London, p. 3, (1956).

[9]Eccles, J.C.: *The Neurophysiological Basis of Mind*, The Clarendon Press, Oxford, p. 122, (1952).

[10]McCulloch, W.S. and Pitts, W.: "A Logical Calculus of the Ideas Imminent in Nervous Activity", Bull. Math. Biophys. **5**, 115–133. (1943).

[11]von Neumann, J.: "The General and Logical Theory of Automata", in *Cerebral Mechanisms of Behavior*, Jeffress, L.A. (ed), John Wiley & Sons, New York, p. 21 ff., (1951).

[12]14 Ref 5, plate 6, 12 and 3.

[13]Mettler, F.A.: Neuroanatomy, The C.V. Mosby Co., St. Louis, p. 432, (1958).

[14]Lettvin, J.Y., Maturana, H.R., McCulloch, W.S., Pitts, W.: "What the Frog's Eye Tells the Frog's Brain". Proc. IRE **47**, 1940–1951 (1959).

[15]Babcock, M.L., Inselberg, A., Lofgren, L., Weston, Pl, and Zopf, G.W., Jr.: *Some Principles of Pre-Organization in Self-Organizing Systems*, Contract Nonr 1834(21) Tech. Rep. #2, Electrical Engineering Research Laboratory, Engineering Experiment Station, University of Illinois, Urbana, Illinois, (1960).

[16]Babcock, M.L., Erickson, R.J., and Neill, D.M.: *A Dynamic Signal Analyzer*, Contract AF 6428, Tech. Rep. No. 3–1, Electrical Engineering Research Laboratory, Engineering Experiment Station, University of Illinois, Urbana, Illinois, (1962).

[17]Inselberg, A. and Von Foerster, H.: *Property Extraction in Linear Networks*, NSF Grant 17414, Tech. Rep. No. 2, Electrical Engineering Research Laboratory, Engineering Experiment Station, University of Illinois, Urbana, Illinois, (1962)

[18]Neugebauer, O.: *Vorgriechische Mathemathik*, Springer, Berlin, pp. 40–78, (1934).

[19]Wittgenstein, L.: *Tractatus Logico Philosophicus*, Humanities Publ., New York, N.Y., (1956).