# *Mental Processes -- How the Mind Arises from the Brain*

*Roger Ellman*

Abstract

Cognition is best understood by examining a model of a cognitive system. Such a model is presented in the following paper.

Most discussions of the mind or brain focus on the "hardware", the neural structure and its biological / electrochemical functioning. But, it is the "software", how the neural components logically interact, that produces the results that we experience in our own minds.

The objective is intelligence -- how we see, think, remember, know ourselves, learn, plan create. To describe and explain those sophisticated functions it is necessary to start with simple first steps, building blocks, and gradually erect the total structure. The reader is urged to be patient with the review of fundamentals in the earlier portions of this paper, which review lays the basis for the development.

The development begins with universals and mechanisms for recognizing or identifying them. It then proceeds through perception, learning, and the processing of universals to mental concepts, thoughts, thinking and memory. Then purposive behavior and its related goals, motivation and consciousness are developed. Finally the implications for the issue of free will [versus predestination] and the designing of an artificial intelligence are addressed.

Roger Ellman, The-Origin Foundation, Inc.
320 Gemma Circle, Santa Rosa, CA 95404, USA
RogerEllman@The-Origin.org

# *MENTAL PROCESSES -- HOW THE MIND ARISES FROM THE BRAIN*

*by Roger Ellman*

Most discussions of the mind or brain focus on the "hardware", the neural structure and its biological / electro-chemical functioning. But, it is the "software", how the neural components logically interact, that produces the results that we experience in our own minds.

## *PART 1 -- THE PROBLEM OF INTELLIGENCE*

### *INTRODUCTION*

The problem is to explain the phenomenon of human intelligence. Scientific knowledge has developed to the point where there is generally a sound understanding of most phenomena in the world, and for those phenomena not yet thoroughly understood there is confidence that development of the knowledge is only a matter of a little more time. But, for the phenomenon of human intelligence there is no well developed scientific explanation corresponding to that for evolution, physics, or biology.

The human brain and nervous system is a very complicated and sophisticated system. It not only performs the human functions of thought, intelligence, self-awareness, and so forth, but the lesser functions found in most animals such as purposive behavior and control of voluntary actions of the body. Furthermore, it is also an involuntary control system that monitors and controls all of the bodily functions so as to make the total biological system of the person (or animal) function in its best overall biological interest.

For example, the brain and nervous system control:

- fuel and materials input (food)

- oxidant input (breathing)

- processing and distribution of these (digestion, blood circulation, waste elimination)

- temperature control

- growth and repair

- reproduction, and so forth.

Involved in these processes are systems of nervous and chemical (endocrine) signals and controls and semi- and fully-automatic sub-systems (heart beat, reflexes, etc.)

This system operates on an evolved design. Humans have sub-systems quite like those of lesser animals, These are apparently retained as evolutionary "carry-overs". They can also be viewed as the retaining of well-developed and well-proven systems of evolutionary precursors upon the base of which, as sub-systems, the more sophisticated human systems are built. A partially true biological paradigm is that "ontogeny recapitulates phylogeny" or, in other words, that the development of the fetus from conception to birth recapitulates the evolution of the specie. An analogous, and related partially true paradigm is that the evolved human nervous system recapitulates and has as functioning sub-systems the evolutionary history of the earlier developed stages of nervous system.

Whether, if one were to design a human "from scratch", one would include all of these mechanisms is a hypothetical question to be perhaps answered in the future. Certainly most of the functions would appear to be needed. However, for the present purposes the issue is intelligence, explanation of that high order human function. Digestion is well understood by science and humans have no monopoly on the process. The same is true for reflexes, temperature control, and so forth. Consequently there is no attempt here to go into the detail of the brain's control of all those type activities of the human brain.

The objective is intelligence. How do we see, think, remember, know ourselves, learn, plan, create ?

In setting out to describe and explain these sophisticated functions, probably the most complex and sophisticated in the universe to our knowledge, it is necessary to start with simple first steps, building blocks, and gradually erect the total structure.

That procedure is followed in the next several parts. The reader is urged to be patient with the review of fundamentals in the earlier portion, which lays the basis for the development.

### *OVERVIEW*

Until assigned a name, things are identified by their description. For example the letter "t" in the last word of the prior sentence can be described fairly definitively as: roman letter "t", in the last word of the prior sentence, black, on a white background, CG Times font, size eleven point, lower case, non-italic. Each of the components in that description can apply to a variety of other things, but together they specify the particular instance. A number of other things have some, but not all, of the *characteristics* of that "t" and have other characteristics that the "t" does not.

The specific individual momentary concepts in our heads are likewise describable in terms of a set of characteristics -- ones that collectively are the particular concept of that instant, ones that are partially shared with a variety of different other concepts.

The process that goes on in our minds is a progression of such specific momentary concepts, *thoughts*. Successive thoughts are linked by having most of their characteristics in common but one or more changed. A chain of such successive thoughts is *thinking*. In the following analysis and development the *characteristics* are referred to as *universals*.

Our minds have thoughts by supporting representations of universals and by detecting various universals amid a mass of other data. We think by chains of successive specific momentary sets of universals progressing from set to slightly different set in a systematic (logical, rational) fashion. But, ... how ?

## *PART 2 -- UNIVERSALS AND PERCEPTION - (1)*

By the word *universal* is meant a class or category to which a particular specific example belongs. *Perception* is the process of properly correlating individual specific cases, examples, with universals. Perception particularly includes the proper identifying or recognizing of examples that have not previously been specifically experienced.

Humans perceive, recognize, a very large number of universals, of course. Some examples, in order to clarify the concept, are:

- recognition of the letter *E*, whether capital or lower case, hand written or mechanically produced, large or small, alone or among other symbols, even though the particular *E* being recognized may be different from any ever before seen;

- recognition of all beings that are human as human beings;

- recognition of all shirts.

The universal is the common characteristic of all elements of the group, that is *E*-ness, human-ness, shirt-ness in the above three examples.

Not only humans recognize universals; most animals do also, but the ability in non-humans is apparently more limited. Nevertheless, for example, a dog can recognize another dog as a dog even though the dog recognized was never before seen and is of a significantly different breed or appearance.

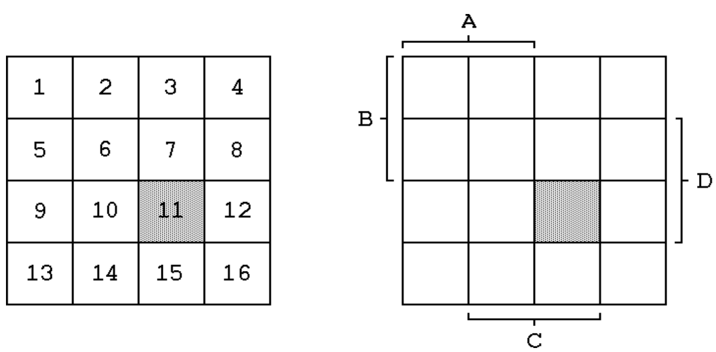
Recognition of universals is not always accurate even though the recognizer is competent. The sample may be a marginal case. For example, everyone is familiar with the problem of reading another person's handwriting, which involves properly recognizing various sample letters as samples of particular letters of the alphabet, which is a set of letter universals.

The process of perception involves an input, a data processor, and an output. For the present case the input is data from a sensory organ: eyes, ears, nose, etc. The processor is some mechanism that operates on the input data so as to correlate examples with universals. The output is data representing that correlation or identification. Of course a given input sample may be a sample of a number of different and perhaps unrelated universals. For example a particular letter *E* might belong to all of the following classes simultaneously: *E*, upper case, small, hand written, in ink, red, moving left to right across the field of vision, upside down, appearing progressively smaller, etc.

The process of recognizing universals is most easily understood by using the case of the sense of sight as the input. The procedures and conclusions apply equally to the other senses or to any coherent or systematic input system. For the purposes here, the sample is projected onto a screen (or the retina of the eye). The screen is not continuous, however. Rather, it is divided into an array of more or less uniformly spaced essentially identical sensors (the "rods and cones" of the retina). Each individual sensor can only register in an on-off manner (for the present); that is, if the part of the image projected onto the screen and falling on a particular sensor is light then the sensor is in the *on* state, if dark the sensor is *off*.

Thus the image projected onto the screen is represented on the screen as an array of black and white dots (*off* and *on* sensors) similar to a photograph in a newspaper as viewed with a magnifying glass.

To initially discuss the process an example using a relatively small array of sixteen sensors arranged in a square of four rows of four sensors each will be used as in Figure 2-1(a), below. It is necessary to be able to refer to each of the individual sensors (elements) of the array. This could be done by sequentially numbering them as in Figure 2-1(a); however, it will be more useful to use the system of Figure 2-1(b), in which the array is divided in half four different ways.




*(a)* *Figure 2-1* *(b)*

(The procedure being used is, of course, the digitizing of the image into binary elements and the description of the sensors and their binary states by means of Boolean algebraic variables and functions. In fact the eye, also, essentially digitizes the image on its retina and supplies signals that are essentially binary to the brain; however, the human processor is not quite Boolean. Boolean discussion will be used for the moment and the conversion to the biological mode of processing will then be presented. For those who are not familiar with these techniques the explanation is continued in simplified terminology.)

The half of the array of Figure 2-1(b) that is labeled *A* will be called *A*. The other half will be called *not A* and be written $\bar{A}$. We can then identify element number #11 of Figure 2-1(a), for example, in Figure 2-1(b) as being in

*(2-1)* $\bar{A}$ and $\bar{B}$ and C and D.

a description that fits no other element of the array.

This procedure makes use of Boolean Algebra, a mathematics of logic originally developed by the Englishman, George Boole, for the purpose of testing and interpreting the logical construct of verbal statements. Although it was developed well before even the notion of digital computers had occurred or could have occurred, Boolean logic is the underlying principle on which digital computers operate.

The letters *A, B, etc.*, are called *variables* meaning that they may vary in value. The allowed values in the present case are *1* ≡ *on* or *"true"* or *"yes"* and *0* ≡ *"not on"* (i.e. *"off"*) or *"not true"* (*"false"*) or *"not yes"* (*"no"*).

Instead of writing "and" over and over as in equation 2-1 the notation

*(2-2)* $\bar{A}\bar{B}CD$ or, when needed for clarity $\bar{A}\cdot\bar{B}\cdot C\cdot D$

will be used and understood to mean the same as equation 2-1. It is read as *"not A and not B and C and D"* and means the state in which *A is not true*, *B is not true*, *C is true* and *D is true*. It also means that portion of the array of Figure 2-1 which is *not in A*, *not in B* and is *in C* and is *in D*.

To refer to more than one element of the array at a time the connective *or* will be used, written as *+*. Thus to refer to the combination of the elements #10 and #11 of the four by four array of sixteen elements (per Figure 2-2, below) the reference is

*(2-3)* $A\bar{B}CD + \bar{A}\bar{B}CD$

which is read as *"A and not B and C and D or not A and not B and C and D"*.

This reference (equation 2-3) can be stated more simply as

*(2-4)* $\bar{B}CD$

because if it is true for *A* or for $\bar{A}$ then it is independent of the value of *A*, whether it is *0* or *1*. That is, for the two elements, #10 and #11, as an area of the array, designation in terms of *A* is to no point. As Figure 2-2 shows, that area is correctly described as the area simultaneously in *not B and C and D* as equation 2-4 presents.

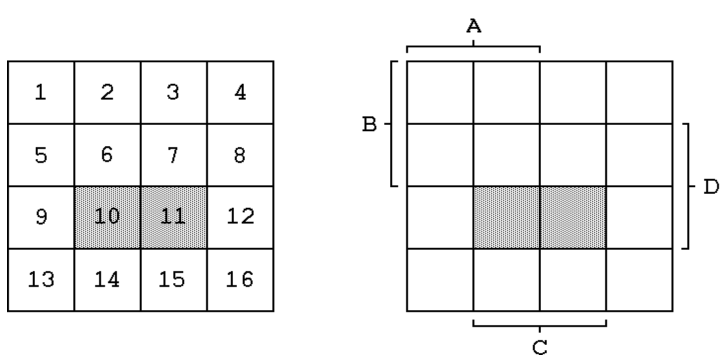
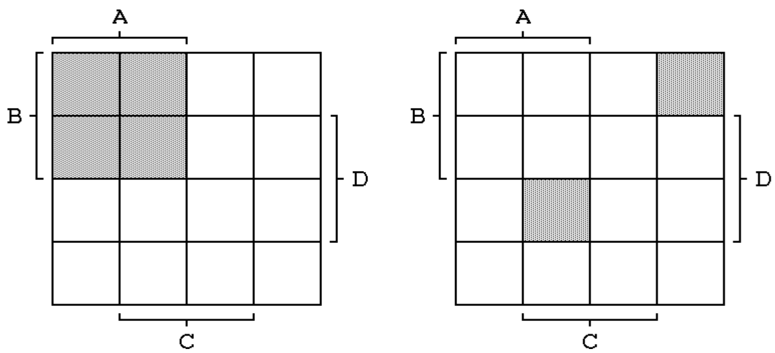



*Figure 2-2*

Other combinations of elements may yield similar such simplifications of the reference and still others may not. For example, the indicated set of four elements in Figure 2-3(a), below, is

$$\text{(2-5)} \qquad AB\overline{CD} + ABC\overline{D} + AB\overline{C}D + ABCD \equiv AB$$

On the other hand, the two elements indicated in Figure 2-3(b), below,




*(a)* *(b)*

*Figure 2-3*

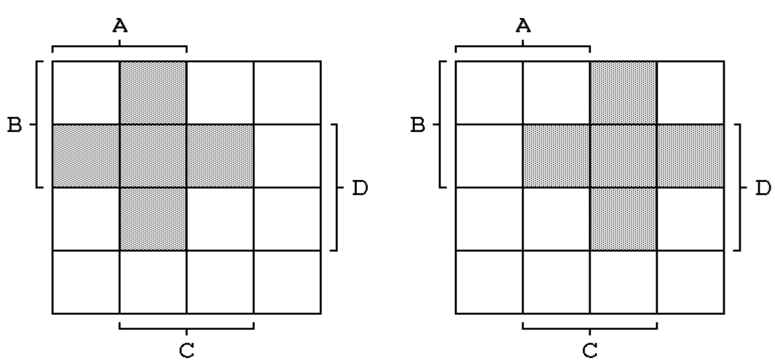
are given by

$$\text{(2-6)} \qquad \overline{A}B\overline{CD} + A\overline{B}CD$$

which cannot be further simplified or reduced.

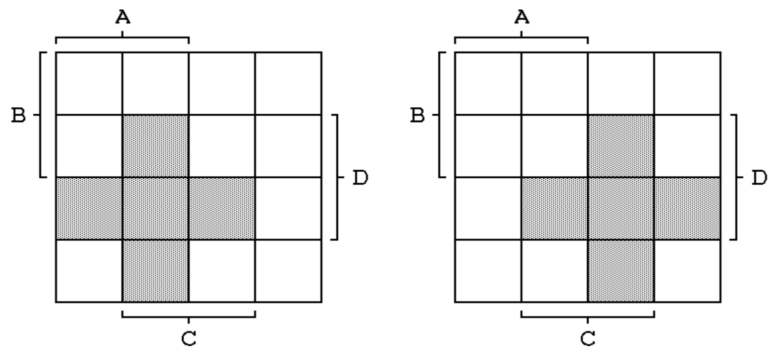
Now let us consider the problem of recognizing (that is identifying the universal of) a simple cross, a horizontal line crossing a vertical line, in this system. More specifically, we wish to obtain a method for recognizing any such cross and only such crosses. Within the special case of the sixteen element array we wish to be able to properly assign any input, as it is projected onto the array, as a member or a non-member of the universal *cross*.

First we consider some examples of the specified input as in Figure 2-4, below and continued on the next page.




*(a)* *(b)*

*Figure 2-4*




*(c)* *(d)*

*Figure 2-4 (continued)*

The elements making up each sample are as follows.

*(2-7)* (a) $ABC\bar{D} + AB\bar{C}D + ABCD + \bar{A}BCD + A\bar{B}CD$

(b) $\bar{A}BC\bar{D} + ABCD + \bar{A}BCD + \bar{A}B\bar{C}D + \overline{AB}CD$

(c) $ABCD + A\overline{BC}D + A\bar{B}CD + \overline{AB}CD + A\bar{B}C\bar{D}$

(d) $\bar{A}BCD + A\bar{B}CD + \overline{AB}CD + \overline{ABC}D + \overline{AB}C\bar{D}$

These can be simplified by expression in terms of two-element areas as follows

*(2-8)* (a) $ABC + ABD + ACD + BCD$

(b) $\bar{A}BC + \bar{A}BD + \bar{A}CD + BCD$

(c) $A\bar{B}C + A\bar{B}D + ACD + \bar{B}CD$

(d) $\overline{AB}C + \overline{AB}D + \bar{A}CD + \bar{B}CD$

(That these expressions include the central element of the cross four times instead of once is mere redundancy and does not affect the accuracy or effect of the expression.)

The commutative principal of mathematics applies to this mathematics; that is, the order of stating variables has no effect on the result. For example

$$\bar{A}BC + A\bar{B}C = A\bar{B}C + \bar{A}BC = B\bar{A}C + CA\bar{B}$$

Likewise, the associative mathematical principal also applies; that is, factoring and the related grouping of variables has no effect on the result. For example

$$A\bar{B}C + AB\bar{C} = A\cdot[\bar{B}C + B\bar{C}]$$ [A is "and-ed" with the bracketed expression.]

Making use of those principals equation 2-8(a) can be expressed as

*(2-9)* AB[C + D] + CD[A + B]

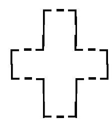
or as

AC[B + D] + BD[A + D]

or as

AD[B + C] + BC[A + D]

Each of equation 2-8(b), (c) and (d) can be similarly expressed. All of the resulting formulations have the same general form:

*(2-10)*  $= V_1 \cdot V_2 \cdot [V_3 + V_4] + V_3 \cdot V_4 \cdot [V_1 + V_2]$

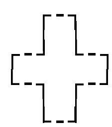
where $V_i$, i = 1, 2, 3, or 4, is a Boolean variable like A, B, etc.

That is, the cross in the four by four element array being treated, is apparently characterized by an identification in terms of the four Boolean variables as the *and-ing* of any two variables with the *or* of the other two, that whole then *or-ed* with the *and-ing* of the other two variables with the *or* of the first two, any or all of the variables being natural or *not-ed* consistently.

Examination of the four by four array being used demonstrates the validity of the following identity

*(2-11)* $A + B = [\overline{\bar{A} \cdot \bar{B}}]$

with the use of which equation 2-10 can be rewritten as

*(2-12)*  $= V_1 \cdot V_2 \cdot \overline{(\bar{V}_3 \cdot \bar{V}_4)} + V_3 \cdot V_4 \cdot \overline{(\bar{V}_1 \cdot \bar{V}_2)}$

Either of the two logically equivalent formulations, equation 2-10 or 2-12, is the extraction of the indicated universal, *cross-ness* from the samples with which the analysis began. Either formulation is, therefore, the means to the perception of that universal in the sample array being studied.

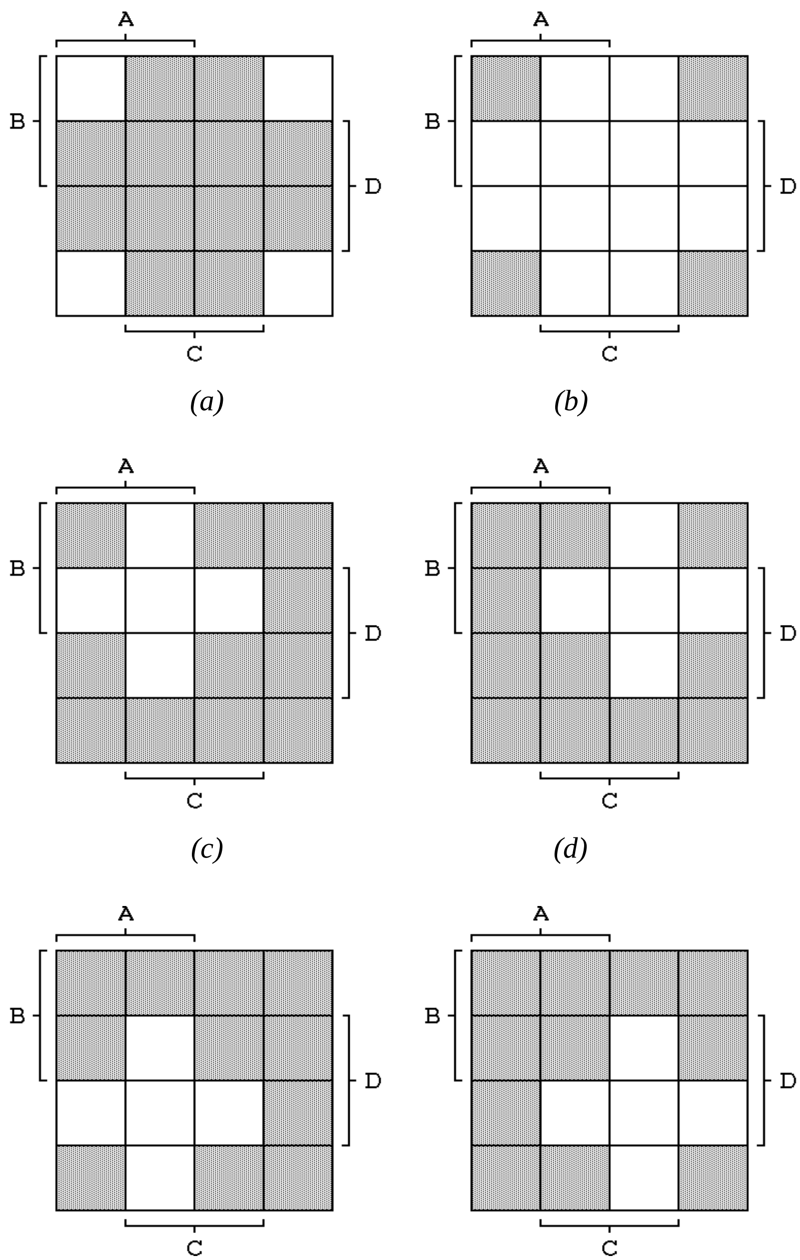
So far in this analysis four sample crosses have been examined. In the simple four by four array being studied there are a total of ten possible symmetrical crosses. The other six are displayed in Figure 2-5 on the following page. All of the ten would be correctly identified by the formulation just derived.

Some asymmetrical crosses would also be identified by the formulation, for example as in Figure 2-6. For each symmetrical cross consisting of five elements there are three ways that it can be asymmetrical: horizontally, vertically or both. Thus each such cross can appear in four forms. The total number of possible crosses, symmetrical and asymmetrical is that four times the eight possible symmetrical five-element crosses, equals *32.* That plus the two larger crosses equals a total of *34.*

The total number of different patterns that can be displayed on this sample sixteen element array is

$$2^{16} = 65{,}536$$

The logical formulation just developed detects the *34* cases having the universal *cross-ness* in common out of the *65,536* total cases. (If one wished the universal to be so defined as to reject the asymmetrical crosses, it can be done with only a little more complexity.)

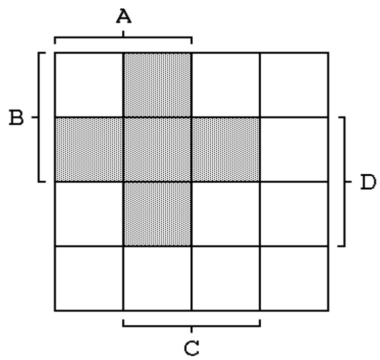



(a) (b)

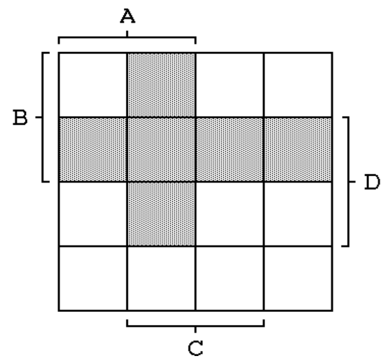
(c) (d)

(e) (f)

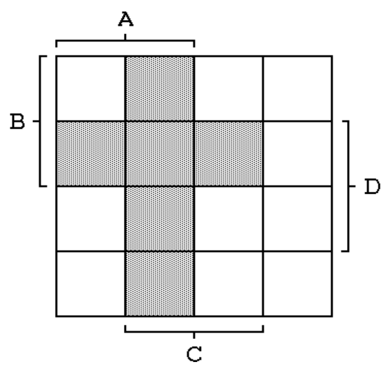
*Figure 2-5*



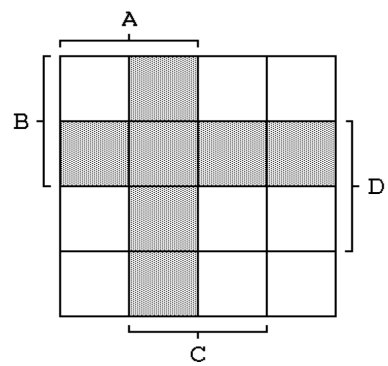

*(a)*




*(b)*




*(c)*




*(d)*

*Figure 2-6*

At this point the example of the four by four array of sixteen elements can be abandoned in favor of the general case of a practical, human perception system. The purpose of the example was to illustrate in a general sense that:

· a sensory input can be analyzed,

· a limited number of input samples can be sufficient to reasonably well establish a formulation for a universal,

· that can be done by digitizing the data into binary representation,

· and the formulation of the universal can be of a kind corresponding to well known digital logic arrangements as used in digital computers and some automatic control systems.

(However, intelligence functions quite differently from the functioning of a digital computer.)

### *Complex Perception Systems*

Instead of sixteen sensory elements as in the preceding example, the human eye has about *7,000,000* such sensory elements, the *rods and cones* of the retina. The number of different patterns that can be represented on a binary digital array is

(2-13) Number of possible patterns

$$= 2^{[\text{number of elements in the array}]}$$

Thus the human eye can deal with about $2^{7,000,000}$ different patterns. This is an extremely large number.

Since $2^{10} = 1,024$, then, taking $1,000$ as an approximation to $1,024$,

(2-14) $$2^{7,000,000} = (2^{10})^{700,000}$$

$$\approx (1000)^{700,000} = (10^3)^{700,000}$$

$$= 10^{2,100,000}$$

$$\approx \text{1 followed by 2,100,000 zeros}$$

If the eye saw a different pattern every $1/10$ of a second it would take $30$ years to see $10,000,000,000$ patterns ($1$ followed by $10$ zeros, not $2,100,000$ zeros) and an essentially inconceivable number of years to see all of the different patterns possible to the eye.

When it is considered further that relationships among different patterns are significant in that they provide information on time sequence, changes, motion, etc., so that different groups of patterns and different orders of occurrence within groups are further input data beyond that of the input patterns taken individually, it is clear that the amount of information available from the human eye, the vision input sensor, is immense.

When an image, an input pattern, is projected onto the retina of the eye, a family of signals from the individual sensory elements of the retina is transmitted to the nervous system for processing. The first level of processing (which actually occurs in the eye, in cell layers of the retina) is to identify all of the *first order* universals in the input image. By *first order* is simply meant any universals identifiable at this first level of processing. These are universals that detect or identify: corners, edges, shape types, motion and so forth, universals similar to the cross of the recent example.

The possible number of such first order universals is quite large, large enough in fact to constitute a complete description of the input image, of any possible input image. Such a description for a particular input image consists of all of the universals identified as present in the input image and their location or orientation in the input image, where they occur. The input is converted from being an array of points in a one-to-one correspondence with the original of the image (each point being light or dark, *on* or *off* as its corresponding point in the original) to an array of characteristics of the input image, the set of *first order* universals that have been identified as present or absent, located in that array according to location in the input image.

This new array, the output of the first level of input processing is the input for all further processing. If we could look at that array as an image on a flat screen it would make little sense to us and would not appear to much resemble the original input. That is because the original input has been re-expressed, encoded, mapped into a new terminology different from the one-to-one correspondence with which we are familiar. But, while meaningless to our conscious selves, that information is quite meaningful to our nervous system. It is the kind of information needed by our nervous system (needed by any rational mechanism) in order to effectively process, to understand and use input information.

However, further processing of the input, the using and understanding of it, must be set aside for the moment in favor of concentrating attention on how the *first order* perception of universals actually takes place.

If we refer to each of the *7,000,000* sensors in the retina individually as *#A*, *#B*, ... for all *7,000,000* of them, then any single image projected on that retina can be represented as the *and* of the signals from all of the *on* sensors *and-ed* with the *and* of the *not* of the signals from each of the *off* sensors. For example

*(2-15)* Some image = $A\bar{B}\bar{C}DE\bar{F}G...$ [7,000,000 letters].

A group of input images, each individual one represented in the form of equation 2-15, could be described as a group by the *or-ing* together of the equation 2-15 type expression for each of the images of the group. The expression for any single image, image *#1* for example, identifies it as the image having (for example) *Sensor A on and Sensor B off and Sensor C off and* .... The expression for the group of images describes the group as (for example) *Image #1 or Image #2 or Image #3 or* .... It would appear (for example) as

*(2-16)* Some group of images =

$$= A\bar{B}\bar{C}DE\bar{F}G... + AB\bar{C}\bar{D}EF\bar{G}... + \bar{A}BCD\bar{E}FG... + ...$$

[Total number of letters = 7,000,000 letters per image times the number of images.]

Such an expression would be the universal of that group of images. That is, any image belonging to the group matches or fits a part of the expression and any image not a member of the group fails to so satisfy the expression. If an image is tested against the expression then a Boolean output result of *1* or *yes* or *on* or *expression satisfied* means that the image being tested exhibits the universal of the group. If an image is tested and produces a *0* or *no* or *off* or *expression not satisfied* Boolean output result that failure is a signal that the image being tested does not exhibit the universal of the group.

These kinds of Boolean logical expressions are readily implemented electronically with simple devices called *logic gates* that produce the *and-ing* and *or-ing* and devices called *flip-flops* that represent the Boolean variables (*A*, *B*, etc.) and remember their current value. They also yield the *not* operation where called for.

However, there are several problems with this approach to constructing a mechanism to recognize and implement universals. The first is that the large number of variables makes the Boolean expressions much to large and cumbersome. Implementing those expressions electronically requires far to many logic gates and flip-flops. As a practical procedure it is unworkable.

In addition, however, and far more serious as a problem, is that this procedure can only correctly test input images that were used in the original setting up of the expression. It is unable to generalize, "to get the idea" of what the universal is, and apply that learning to correctly treating new images never before experienced. In the above approach the universal detecting mechanism must be constructed from the beginning using all possible examples of the intended universal plus all possible examples that are not of the universal. Not only would such a device be far too large and expensive; most likely it is impossible to even identify all of the possible input cases called for.

In other words, such a system has no ability to learn, to modify and improve its behavior on the basis of experience. That defect makes the system far too cumbersome to be practical and also leaves the system not corresponding to that which we know about rational systems -- rational systems do learn. Not only do intelligent humans learn; all animals having some form of nervous system exhibit some learning, learning that varies from the sophistication of chimpanzees to the much simpler, yet still quite complex, worm.

Referring to equation 2-16 again, suppose that every input image that exhibits the universal of interest has sensor *#B = on* regardless of the state of any of the other sensors. Likewise suppose that every input image that does not exhibit the universal of interest has sensor *#B = off* regardless of the state of any of the other sensors. Then sensor *#B* alone would represent the universal. The logical expression to represent the universal and test for its presence or absence in input images would be very simple -- a case of examining sensor *#B* and ignoring the rest of the image for this purpose.

In general it is the nature of universals that they exhibit such simplified expressions although not necessarily nor usually as radically simple as the example just used. A universal is a kind of generalization, an omission of non-relevant specifics in favor of a focus on the broad commonality. Its expression tends to be simpler than the expression for the collection of all images exhibiting the universal and all that do not. This simplified representation of commonality among input images is precisely what a universal is.

The problem at this point is, then, how does a rational system operate in a fashion that overcomes the above problems ? How does it extract a simplified universal from a group of sample inputs ? How does it develop the ability to recognize an input never before experienced ? How does a rational system learn ? For, the process of extracting simplified universals from a partial set of input examples is what learning is.

## *NEURAL-TYPE LOGIC DEVICES*

The *neuron* is a special type of biological cell which is the operating component in the nervous system of all life on Earth that has a nervous system, whether human, animal, insect or whatever. By *neural-type logic* is meant systems in which the principal operating component is the *neuron* or systems in which the principal operating component is a device, a man-made device, that operates *logically in the same way as a neuron*.

The logic technique used in such neural-type rational systems, including the human brain, is slightly different from the *and / or* logic examined so far. The basic logic function (procedure) used in biological systems is *majority logic*. Using the notation *M(...)*, where the *M* stands for *majority of* and the variables involved (e.g. array or retina element signals) are listed in the parentheses, then a translation between *majority* and *and / or* logic is, for example

(2-17)    $M(A,B,C) = AB + AC + BC$

That is, a majority logic operator has an output of *on* if a majority of its inputs are *on* and otherwise an output of *off*. In the example of equation 2-17 any two of the three variables is a majority of them.

For convenience of notation, and because Boolean algebra employs binary logic (a logic based on the binary number system having base *2* instead of *10* and digits *0* through *1* instead of *0* through *9*), the binary digit *1* will be used to represent *on* or *yes* or *satisfied* hereafter with regard to Boolean algebra expressions and the digit *0* to represent the opposite. In those terms equation 2-17 states that the output is *1* if any two or all three of the inputs are *1*. Otherwise the output is *0*.

In addition to variables such as the *A*, *B*, etc. already used, majority logic can also use *logic constants*. Here a constant is like a variable in all respects except that it always has the same, fixed value. Since the system is binary there are only two values that a constant can have, *1* or *0*.

In *and / or* logic, constants are essentially meaningless as the following examples illustrate.

(2-18)    $A + B + 1 = 1$    [In spite of the variables the result is always "1". The variables are meaningless because of the constant.)

$A + B + 0 = A + B$    [The constant has no effect.]

$A \cdot B \cdot 1 = A \cdot B$    [The constant has no effect.]

$A \cdot B \cdot 0 = 0$    [In spite of the variables the result is always "0". The variables are meaningless because of the constant.)

However, in *majority logic*, *constants* play a useful and important role; they enable *majority logic* to represent *Boolean logic*. For example:

(2-19)    $M(A,B,1) = A \cdot B + A \cdot 1 + B \cdot 1 = A + B$

$M(A,B,C,1,1) = \ . \ . \ . \ . \ = A + B + C$

$M(A,B,0) = A \cdot B + A \cdot 0 + B \cdot 0 = A \cdot B$

$M(A,B,C,0,0) = \ . \ . \ . \ . \ = A \cdot B \cdot C$

The *not* operation still applies in *majority logic*; that is, the majority operation may operate on *natural* or *not-ed* variables. For example

(2-20)    $M(A,\overline{B},\overline{C},1,1) = A + \overline{B} + \overline{C}$

$M(\overline{A},B,0) = \overline{A} \cdot B$

Thus majority logic with both constants and variables can produce all of the fundamental type logical constructs that Boolean logic uses.

Likewise, a majority operation's output can be an input variable in another majority operation just as in Boolean logic. For example

```
            ┌-------┐
(2-21)  M(A,M(B,C,0),1)
            └-------┘
```

where the bracket indicates the "*The Majority of B, C* and *0*" as one of the variables in the overall expression, which reads as "*The Majority of A, The Majority of B, C* and *0*, and *1*. Such complex majority operations, which can have many more levels than the two-level case illustrated in equation 2-21, enable majority logic to implement any Boolean logic whatsoever.

In fact majority logic can do more than that. The very same physical structure, that is the same connection of inputs to a given majority processor, can yield controllably different logical constructs, logical results, depending on the value of the constants applied to that majority processor. Majority logic makes possible fixed "pre-wired" interconnections in a configuration where the logical effect of the physically fixed structure can be controlled and varied by varying the values of the constants involved.

That is precisely the process that goes on in a rational system based on neurons, whether that system is in a human, a cow, an ant or whatever. The inputs to a neuron are the outputs of other neurons or of sensors (e.g. the retina of the eye). Those inputs are such that some act on the neuron in an *excitatory* fashion and some act on it in an *inhibitory* fashion. That is, *excitatory* inputs are analogous to *natural* variables (as opposed to *not-ed* ones) and have the logical effect of an input of *1* if activated and *0* if not. *Inhibitory* inputs are analogous to *not-ed* variables and have the logical effect of an input of 0 *if activated and* 1 if not.

In a neuron the presence or absence of a majority is not determined by counting the total possible inputs and comparing the number of them that are *1* to that count. Rather the effect is as if the *1* inputs are each *+1* (excitatory) and the *0* inputs are each *-1* (inhibitory). If the algebraic sum, the excitatory plus the inhibitory (the number of excitatories less the number of inhibitories), is greater than zero then a majority is present.

There is still another component of a neuron's operation, however. That *algebraic sum* of the excitatory *+1* and the inhibitory *-1* inputs is not compared to *zero* as such. Rather it is compared to a *threshold* level present in that neuron. If the *threshold* happens to be *zero* then the logical construct of the neuron is simply the majority of its inputs.

But, if the threshold is greater than *zero*, meaning that for the neuron to have an output of *1* the number of excitatory inputs must be that much (the *threshold* amount) greater in number than the number of inhibitory inputs, then the effect is the same as if there were as many constants equal to *0* present and acting as the level of the *threshold*. Likewise, a *threshold* less than *zero* corresponds to there being that many constants equal to *1* present and acting. Thus the value of the *threshold* represents the net value of constants in the input and variation of the *threshold* produces variation of the net value of the constants which produces variation in the Boolean logic that the majority operator is equivalent to.

For example, if the inputs to the neuron are *A*, *B*, *C*, *...* and all of them are excitatory (simply for this example), then:

| (2-22) | With Threshold | The Neuron Performs |
|---|---|---|
| | 0 | M(A,B,C,...) |
| | +1 | M(A,B,C,...,0) |
| | +2 | M(A,B,C,...,0,0) |
| | -1 | M(A,B,C,...,1) |
| | -2 | M(A,B,C,...,1,1) |

The threshold is equivalent to the net number of constants involved, constants of *-1* for positive threshold and of *+1* for negative threshold. The output is *1* if the majority of the input variables and those constants is greater than *zero*.

But, the special power of the neuron is that its threshold can be changed. That means that its constants can be changed and that means that the logical effect, the Boolean logic that the neuron is implementing, can be changed. The neuron "remembers" the value of the threshold so that the threshold is, in that sense, some set number of majority logic constants operating as such in the logical construct that the neuron effects. However, that set value or level of the threshold can be changed, adjusted so that the logical construct that the neuron effects is slightly, gradually changed. It is that process that enables learning. Learning is, in effect, the directed adjustment of neural thresholds to achieve the desired result.

The input to the neuron from other neurons or from sensors is received by the neuron as various excitatory and inhibitory, *+1* and *-1*, inputs. The neuron emits an output that is *1* or *0* depending on the internal operation of the neuron. That output acting as an excitatory input to another neuron is a *+1* input to it if the output was *1*. That output acting as an inhibitory input to another neuron is a *-1* input to it if the output was *1*. The internal operation of the neuron simply determines whether the majority of the inputs plus the threshold is greater than zero (neuron output is *1*) or not (neuron output is *0*). (How the threshold changes occur will be treated shortly, in the next section of this work.)

Actual biological neurons operate in this manner. A single biological neuron consists of a central cell body, a number of input lines (filaments or fibers of cell material) called dendrites, and an output line (also a filament or fiber of cell material) called an axon. Output signals of neurons travel to the end of the axon where they then communicate, as inputs, with the dendrites of other neurons. The junction where the signal transmission from neuron to neuron takes place is called a synapse. Within a neuron some of the dendrites (inputs) are excitatory and some are inhibitory. The threshold, at the main cell body, determines whether the net effective input signal causes or fails to cause an output signal on the axon. The processes within the neurons and at the synapses are electrochemical in nature.

When neurons, whether biological or man made neural-type electronic devices, are interconnected so that the outputs of some neurons are inputs to other neurons then a multilevel neural network exists. Such a network makes possible neuron-implemented complex majority logic structures that can effect logic such as illustrated in equation 2-21. Multilevel networks of neurons use the neuron's majority logic, modified by the individual neuron's thresholds, to represent the equivalent of complex Boolean logical descriptions. Such descriptions are the logical representation of universals. Complex neural networks can thus represent specific universals if the individual neural thresholds are correctly set to make them do so.

Let us now operate a simple such neural network using as its input the sample four-by-four, `16` element, array used in the first part of this section. That array was there used to illustrate the universal *cross-ness* among the various possible images that could appear as input on the array.

An individual neuron or neural-type device will be symbolized as in Figure 2-7, below.

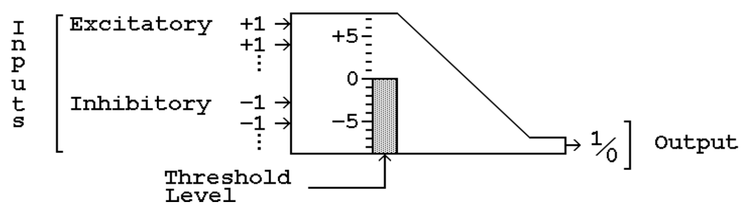



*Figure 2-7*

The outputs of the four-by-four array will be interconnected to the inputs of a number of such neurons and then the outputs of those *first level* neurons will be interconnected to the inputs of one more neuron. The output of that final, single, neuron will be deemed the representation of the action of the entire neural network. (See Figure 2-8 on the following page.)

But, how should the interconnections be made; that is, which sensors should be connected to which inputs of which neurons ? This question is quite fundamental to neural networks as is the matter of how threshold changes occur. As with the control of threshold changes, the subject will be treated fully in the following section. For the moment let us assume that those aspects of the problem have been correctly implemented in the sample neural network being used.

Let us now teach the neural network to recognize the universal *cross-ness*; that is, let us cause it to learn how to discriminate between input images exhibiting *cross-ness* and those lacking it. Our objective is that the neural network should give an output of *`1`* if the input image has *cross-ness* and *`0`* otherwise.

We use the following procedure.

(1) Show the input array an input image (project an image onto the four-by-four, `16` element array). That is, cause various of the `16` elements in the array to be *`on`* and others *`off`* so that the desired pattern is represented on the array

(2) (Being the teacher in this case, the authority, we) note whether the image exhibits the universal *cross-ness* or not. (The problem of where, in general, the teacher comes from is also addressed in the next section.)

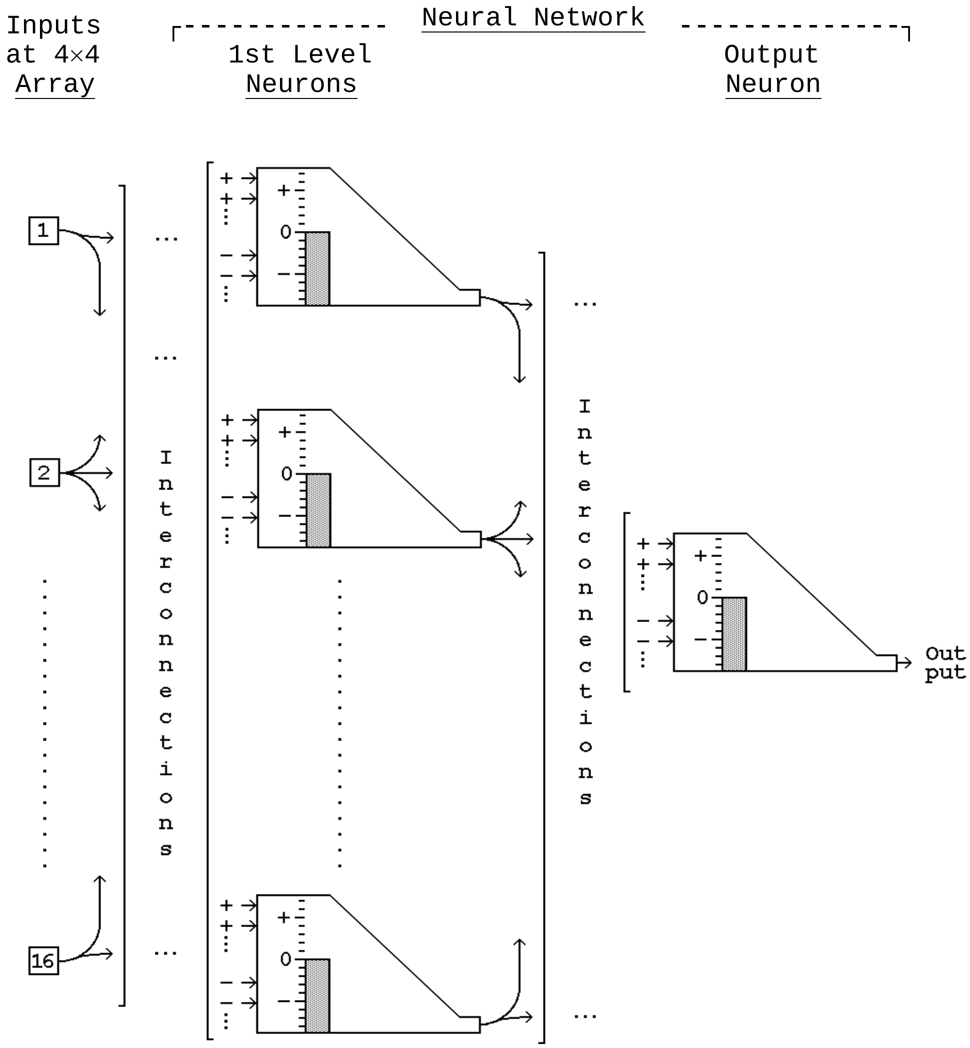



*Figure 2-8*

(procedure continued)

(3) Observe the output of the neural network (whether it is *1* or *0*).

(4) Evaluate the performance of the network, which could be any of the following four possible cases.

| Input Image | Output | Result |
|---|---|---|
| cross | 1 | correct |
| cross | 0 | wrong |
| not cross | 1 | wrong |
| not cross | 0 | correct |

(5) Change the threshold of each neuron of the neural network as follows:

- If the neural network output was correct reinforce that behavior by adjusting each neuron's threshold in the direction that makes that result more likely.

  · If its output was `1` lower its threshold by `1` unit (making even more likely a `1` output for another input like this one).

  · If its output was `0` raise its threshold by `1` unit (making even more likely a `0` output for another input like this one).

- If the neural network output was wrong discourage that behavior by adjusting each neuron's threshold in the direction that makes that result less likely.

  · If the output was `1` raise its threshold by `1` unit (making less likely a `1` output for another input like this one).

  · If the output was `0` lower its threshold by `1` unit (making less likely a `0` output for another input like this one).

(6) Repeat the above five steps using a different input image each time until the neural network's performance is sufficiently consistently correct.

This has the appearance of a reward-and-punishment type procedure but that is not the case here. The neurons do not understand anything, certainly not reward and punishment. The procedure simply changes the thresholds in a direction tending to increase the chances that for input images similar to the one just processed the neural network's operation on the input variables, with its now changed thresholds, will yield the desired correct output.

But, whether the neurons "understand" this or not is irrelevant. The end result of the process is that the neural network actually becomes able to discriminate *cross-ness* even though at the start of the process it could not do so. The neural network has learned, been taught by the teacher, to discriminate. It effectively *perceives* the *universal* taught, *cross-ness* in this example, having *learned* to do so.

That learning was accomplished by directed, logical adjustments to each neuron's threshold level. Such adjustments have already been shown to change the Boolean logical construct that is effected by each neuron's majority operation in conjunction with the constants represented by its threshold.

In other words, the above described learning process causes the Boolean logical construct or operation that the neural network performs on the input variables to gradually change until it is identical to, or it sufficiently resembles, the Boolean logical construct that corresponds to the universal being taught.

The accomplishment of that <u>is</u> the learning to perceive that universal. The subsequent using of that to make correct outputs in response to input images <u>is</u> the perceiving of that universal.

[This concept and laboratory research with regard to it were first developed and pursued at the Cornell Aeronautical Laboratory in the latter 1950's. The research was reported in the Proceedings of the Electronic Computers Group of the (then) Institute of Radio Engineers, IRE, (now the Institute of Electrical and Electronic Engineers, IEEE) circa 1960. The neuron simulator device, operating as herein described, was called the "perceptron". Laboratory development demonstrated that the type device does learn and operate as here described.

[The first generation of commercially produced machines using these principles were in the mid 1990's appearing on the market and being used. The machines employ neural networks similar to those described above. The machines are used to perceive patterns in data in situations where humans may be too slow or unable to perceive the pattern.]

In general summary so far:

· Perception is the correlating of an experienced example with a universal, a class to which it belongs.

· Learning is the developing of the ability to so perceive.

· The perception is accomplished by having -- the learning is the process of constructing -- a logical mechanism that operates on the experienced example in a fashion that detects the presence or absence of the universal.

· That "logical mechanism" is a physical implementation that is, in effect, a Boolean logical expression that conforms to the universal.

· The "logical mechanism" is "constructed", exists and operates, by means of majority logic with constants as implemented by neurons or neural-type devices having majority logic and adjustable thresholds.

While this process has been discussed in terms of our sense of vision the same process operates with regard to all of the senses: hearing, smell, touch, etc. Hearing involves the universals in sounds and hearing and understanding language involves universals just as numerous and complex in their effect as in the case of vision. The blind read by their sense of touch and process a similarly numerous and complex set of universals through their fingertips. And some of the animals, unlike we humans, derive quite extensive information from their sense of smell.

## *PART 3 -- UNIVERSALS AND PERCEPTION - (2)*

The perception mechanism as developed so far in the preceding section is a simple, one-output, system that would appear to be a suitable building block or prototype design for more complex rational systems. The problems of adapting it to multi-output systems, let alone the problems of managing to obtain true intelligence from such systems, are yet to be addressed. However, even as the simple, one-output model that it is, it requires a *teacher*, external to the system, to direct the *perceptron* in its learning. It is the *teacher* who decides whether and how much to increase or decrease the thresholds after each input image is processed. Who or what is that *teacher*. Where did, does, that *teacher* come from ?

### *THE "TEACHER" -- THRESHOLD CHANGES*

This is not a difficult question as it turns out. First it must be recognized that this system is not a complete control, let alone rational, system for any life form. It is only a basic building block of such nervous systems. There are many nervous systems, those of humans, apes, snakes, worms, ants and so forth. This building block need not achieve any spectacular or sophisticated performance such as composing a symphony or originating Newton's Laws. It need merely be a component, one of at least hundreds if not millions or more, of similar components, in a system that for example enables an ant to walk, seek food, defend itself, etc.

Second, it may well be that an aspect of evolution, of variation among individuals of a specie, variation that gradually or suddenly leads to a different type specie, is that of some of the new individual's neurons having different threshold settings as the individual is born (hatched, or whatever) than the parent had at birth. The individual would therefore have different pre-learned-because-born-with-them perception mechanisms. (On a larger scale what can *instinct* be except *pre-learned behavior* somehow embedded in the nervous system. The nervous system consists essentially only of neurons. If *instinct* is to be embedded in it, the only available way is in the thresholds of various neurons and in the interconnections among them.)

But, third and finally, how do we humans, and the other animals, learn something ? By repeating it, repeating it over and over until it "sinks in". There is no other way that we or any animals learn. Explanations, demonstrations, experiences are only the means to learning. The learning only happens when the lesson is repeated sufficiently enough that it "sticks".

There is only one explanation of that effect that is plausible and reasonable. Since:

- In the operation of real world nervous systems there is no external *teacher* to adjust the thresholds, and

- learning does occur in such neural systems, and

- that can only take place by means of threshold changes, and

- things are learned by repetition of the thing over and over until it is learned,

then it must be that

- <u>the threshold changes are automatic</u>, that they occur by a simple, inherent process within the neuron.

Whenever a neuron "fires", that is delivers a *1* output because a majority of its inputs and its threshold so correspond, then its threshold naturally and automatically must decrease slightly so that a similar "firing" will be more likely under subsequent similar input conditions. Whenever a neuron "does not fire", that is it effectively delivers a *0* output because a majority of its inputs and its threshold is not present, then its threshold naturally and automatically must increase slightly so that a similar "non-firing" will be more likely to occur under subsequent similar input conditions.

That would be a behavior of learning by repetition. The repeating of successes is the repeating of the same, or very similar, inputs thus obtaining the same output. That would cause the repeating of the same, or very similar, actions in each of the individual neurons involved. That would tend to change the thresholds of those neurons in the direction that makes the same outcome even more likely.

The repeating of failures would tend to, at least, break up the above pattern of developed successful thresholds. It might, at most, correspond to the learning of the "failure" by its repetition.

In fact, what happens if a human or an animal lapses in regular practice or rehearsal of something learned ? We begin to gradually forget it, to lose the skill, to find it somewhat harder to remember the point. That would exactly correspond to the gradual decay of learned thresholds if they are not regularly reinforced by repetition, by practicing the learned behavior or fact.

This is not an unreasonable situation. The neuron is an electrochemical device. Its operation is the propagation of electrical potentials that are generated and transmitted within the neuron by chemical actions. The "firing" of a neuron, the delivering of an output amount of electrochemical energy, would logically be expected to temporarily deplete the neuron in some sense, perhaps also depleting its threshold, which itself is an electrochemical element in the neuron's overall functioning. Likewise, the absence of "firings" could be expected to give the on-going restorative actions of the neuron, its metabolism, the opportunity to accumulate more threshold.

The point here is not to specifically analyze this process, a process the analysis of which is a lifetime activity for a microbiologist. Rather, the point is that:

- the learning-type threshold changes really do occur in the neurons of any and all nervous systems, whether spider or man, and

- there is no other source for the directing of those changes than that of the effect of simple repetition, and

- simple repetition corresponds in any case to the way in which things are learned in the real world.

### *SYNCHRONIZATION*

Another question with regard to this perception system concerns the synchronization of its operation. The analysis and discussion has implicitly contained the idea that all of the data from the sensors (the retina) are simultaneously available at the input to each first level neuron and are evaluated there simultaneously. Likewise, it has been implicitly assumed that all of the first level neurons simultaneously deliver their outputs as inputs to the second level, final, neuron for its evaluation of them.

All of that simultaneity seems quite unlikely in a real biological neural system where the travel paths over various different dendrites and axons will be of different lengths so that the time of travel of their electrochemical signals in the various neurons must be different. In addition it would not seem reasonable that nature rely on such exact same processing or reaction time relative to the threshold within each of the neurons.

This even raises the question as to what does the "non-firing" of a neuron mean, as it is used in the discussion of thresholds and their changes. The implication is that at a time or under a set of conditions where a "firing" or a "non-firing" could or should occur it is the "non-firing" that occurs and is observed. How does this happen ?

This same problem exists in human-made logic systems as employed in digital computers. Those machines always employ a clock, an overall synchronizing mechanism. The clock is an oscillator, a generator of a train of pulses (*1's*) at a preset constant rate or frequency. Essentially, the input to every flip-flop, every memory element, is *and-ed* with the clock pulses. Consequently, regardless of what goes on in between the clock pulses, it is only the conditions at the time of the clock pulses that cause the next logical step in the operation of the digital computer's logic.

For that system to exist in a biological rational system it would be necessary to have the clock generator, some kind of oscillator, as a part of that system somewhere within it, and to *and* its output at the input to every neuron. It is very clear that biological neural systems do not have some same input that appears at every neuron, whether from a "clock" or from anything else. The brain simply is not constructed that way. Neurons connect densely to other neurons and sensors that are physically near to them, less densely to those that are somewhat distant, and rarely or not at all to those that are very distant (excepting sensor neurons that carry signals from distant sensors to the brain and motor neurons that carry signals from the brain to muscles).

Then, how is a biological logic system synchronized ? It isn't. It simply is not (evolutionarily was not) practical to employ such a system in a biological rational mechanism. So, the systems evolved with the ability to operate without synchronization. In that sense almost all of the time all neurons are putting out a *0* signal. Then from time to time (so to speak) occur the exceptions, the pulses of *1* signals here and there as the neural logic dictates. Those signals from neurons produce excitatory or inhibitory neural inputs to other neurons. Depending on how they interconnect to the particular neuron that neuron experiences *+1* and *-1* inputs in consequence

Furthermore, that neuron experiences *0* inputs most of the time, that is inputs of "nothing happening". Such an input from some other neuron or sensor means "just now the source sensor or neuron is not participating -- the Boolean variable or variable combination that it represents is not part of the logic being effected at this moment".

The neurons cannot emit output pulses continuously. After an output a period of time must elapse during which the neuron metabolism produces sufficient electrochemical recovery from the expenditure involved in an output firing. During that time inputs received may enter into the determination of the neuron's next firing -- the "when" of that firing because of their affecting the amount of recovery the neuron must achieve and the "what" of that firing by their affecting the net electrochemical changes of state within the neuron.

The non-synchronized mode of operation of such neural systems facilitates another characteristic of living neural systems. Even though the systems are essentially binary in that they transmit pulses that are treated as present or not present, *1's* and *0*'s , those pulses also convey valuable non-binary information: that of *how much*. Whether the sensor is one that detects touch or temperature at a point on the body or one that detects sound in the ear, or light in the eye, the information conveyed to the neural system by sensor outputs is that of both *what* and *how much*.

The *what* depends on where the sensor is located and how it relates in physical position and neural network logic to the rest of the system. The *how much* is communicated by the rate of such sensor neuron firings, by the rate of pulses output by the neuron. More frequent pulses are caused by, and therefore signify, brighter light or louder sound or more harsh touch sensation. Less frequent pulses imply the opposite. It is simply that the greater the rate at which the sensor receives excitation energy the greater is the rate at which it is able to deliver output energy in repeated firings.

During the period that a neuron is recovering from its most recent firing and changing electrochemically until it reaches a state that is able to produce another firing, that neuron may receive a number of input pulses on one or more of its input dendrites. That conveys information as to *how much* as well as *what* to the neuron's operation.

This unsynchronized mode with magnitude conveyed by pulse repetition would appear to not be completely compatible with the simple logic system operation that was presented in the previous section. However, the net operational and logical effect is still retained.

The neuron implements a piece of Boolean logic which is determined by its fixed input connections and its variable threshold. That piece of logic in conjunction with those of similar other neurons implements an overall complex

Boolean logic corresponding to the defining of some universal. But, the neuron does not do that by understanding and operating Boolean majority logic in an overt sense. The neural network simply automatically adjusts its thresholds, on the basis of repetition-learning, until the appropriate resulting output ends up occurring.

The combination of the neural majority logic and the learning-directed variable thresholds naturally leads toward the objective of the learning: identification of the related universal. The simple system described in the previous section is a *static* system. But, with synchronization removed the system becomes *dynamic*. It responds to *how much* data. It deals with input patterns in *time* as well as in *array space,* that is patterns which includes elements of both kinds in their input.

### *MULTIPLE UNIVERSALS*

Each neuron receives input from a number of sensors and / or other neurons (tens, hundreds, and in many cases thousands of inputs). Each neuron's output is input to a number of other neurons. The neurons are extensively interconnected. This is schematically illustrated in Figure 3-1, below. The figure also illustrates a structure of the neurons in layers. The layered structure is more pronounced in neurons near to sensor arrays, but occurs to some extent throughout a large scale neural system. Deep within such systems there are more interconnections within layers as well as between them, a diffusing of the sharp layer boundaries depicted below.

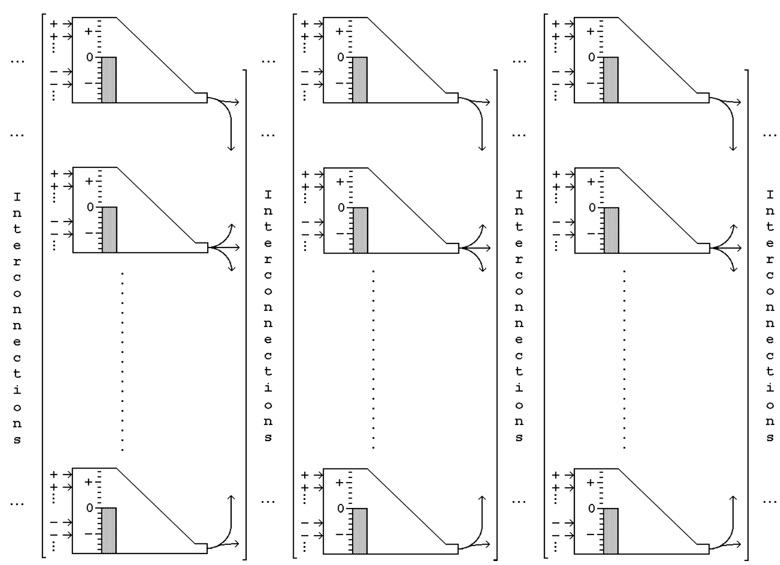



*Figure 3-1*

Considering, for example, the case of vision, the number of universals to be perceived is quite large. And, mechanisms for perceiving each of those universals must be replicated a large number of times over the field of vision or whatever the input sensor system is. That is because the significance of such universals is not only that of which universal is involved but that of where in the input array it appears.

Thus it would be quite impractical for individual neurons to be dedicated to participating in only a single universal. Far too many neurons and far too large a network or brain would be needed. The actual situation must be more like that of the above Figure 3-1 where each of the second layer neurons has an output that is a different result derived from the same overall set of first layer neurons. Each first layer neuron's output participates in a number of universals, a number of next layer logic processes and their outputs, a number of particular Boolean logic expressions, all simultaneously.

For example, the top neuron in the middle layer (column) of the above Figure 3-1 is, relative to the left layer, in the same role as the rightmost neuron (output neuron) of the earlier Figure 2-8, reproduced below. Each of the middle layer neurons of Figure 3-1 is, relative to the left layer, in an output neuron role of Figure 2-8, but each is implementing a different logic, a different universal because each has a different set of inputs..

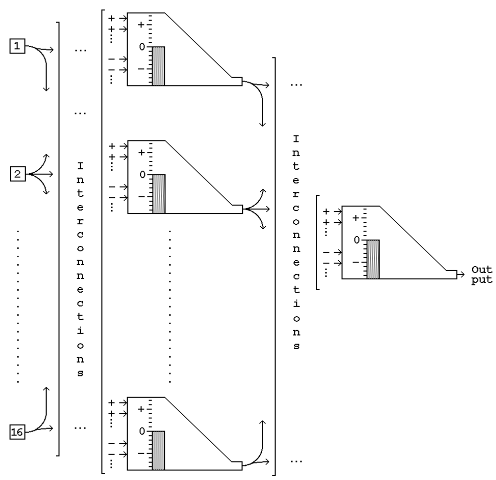



*Figure 2-8 (repeated)*

Likewise, the top neuron in the middle layer (column) of the above Figure 3-1 is, relative to the right layer, in the same role as the left column neurons (first level neurons) of Figure 2-8. Each of the middle layer neurons of Figure 3-1 is, relative to the right layer, in an input neuron role like the role of the left column neurons of Figure 2-8. Thus, in input role a neuron supplies input to many other neurons that are in output role relative to it. As a result the input role neuron participates in implementing a number of different universals.

The involvement of individual neurons in a number of universals simultaneously is necessary not only because the otherwise inefficient use of neurons would require too many neurons in the system. It is also unavoidable given the complex system of interconnection. While it has some drawbacks in its effect on the logic system it also offers a quite tremendous advantage.

The significant drawback is that the threshold of an individual neuron is changed by actions involving any of the universals in which it participates. So to speak, having been thoroughly trained on *cross-ness* and having its threshold well adjusted for that purpose, it then must learn *circle-ness* and in the process of being so trained its threshold is further changed. That most likely would degrade its ability to identify crosses. Over a period of experiencing inputs randomly varying between circles and crosses the threshold would become the best compromise achievable for perceiving either of the inputs.

The effect of this kind of behavior is experienced by us on the large scale. Having learned some thing fairly well and then progressing to further related learning we find that our learning of the former thing has degraded somewhat.

The quite tremendous advantage, however, of individual neurons participating in a large number of different universals is that that multiple-participation creates the capability for *thinking* to take place. However, treatment of that process must be deferred until the next section while the remaining details of neural perception of universals are resolved.

### *THE NEURAL INTERCONNECTIONS*

The operation of these neural systems depends on two variable quantities: the interconnections between neurons and the threshold adjustments. The interconnections are only a variable during the design phase, while the physical device is being built. Any neural system, whether biological or man made is ultimately "hard wired", a fixed system of interconnections. Of course in biological neural systems the system is not "hard wired" during the initial formation and immature growth phase, which may include the first part of the period after birth, hatching or whatever. But, in any case, the issue is that of how those interconnections should be (are in biological systems) for optimum performance. This problem is best approached by the process of imagining the designing of such a system

The fact of the matter is that, initially, we have no idea how to interconnect the sensors and neurons and the neurons with other neurons. That, is the key to the solution. Nature had no idea, originally, either.

Under that circumstance the best choice is randomness.

The interconnections determine (in conjunction with the thresholds) the specific Boolean logic that will be implemented by that part of the system. However, we have no idea what the specific logic is, nor what it should be. If we knew the logical interconnections for the desired universal we could "wire them in". But, we do not know the logic even if we did know the intended universal. And, in any case, we need a system that can deal with any universals, with any input.

That is the point. A biological neural system has to be able to deal with a great variety of inputs. It is not possible to design in advance for all of the possibilities. Given that, then the only way for a neural logic system to maximize its ability to deal with the unknown is to use random interconnections.

Of course there is one alternative, that of including every possible interconnection alternative in the neural system. That would certainly equip it to deal with all situations. However, that is impossible to do. With the immense numbers of neurons the interconnection possibilities are just too inconceivably large. And every added neuron adds even more interconnection requirements than it contributed to satisfying.

Our biological system is not like a merchandise bar code reader in the check-out of a market, which device is designed to deal with only a very specific input. Rather, we humans and the other animals must deal with all the variety of experiences that are encountered in life. Our neural system must be flexible. Any systematic method of interconnection inevitably must favor some logic and disfavor other logic. Only a random interconnection system yields a system able to deal with any (or at least most of the) logic required of it.

Random interconnections is also the easiest and most natural system for nature to implement. It requires no plan and no control. It calls for merely allowing what happens to happen.

Then Darwin's variation and natural selection step in. Some "random" interconnection systems turn out to perform better than some others. ("Perform" here means promote the success of the life form, its ability to reproduce and perpetuate its specie.) The process tends over time to select optimal interconnection systems.

But "optimal" depends on the specific situation being dealt with. Vision systems have existed in nature for hundreds of millions of years. There has been sufficient time and experience for the optimal set or family of retinal first order universal processing interconnections to develop to optimum. Whether the being is a fish, a dinosaur, a mammal of prehistoric times or man, the fundamentals of vision are well defined by experience and largely the same: detection of size, motion, corners, solid areas, and so forth.

But, what is optimal brain operation for astronauts, steelworkers, gourmet chefs ? Man experiencing so many different geographies, weathers, food supplies, dangers, and so on confronts a thinking need that cannot be predetermined. While his vision system may be well defined, the needs of his thinking system are very broad. The most likely success is one that can adapt to any circumstances.

Thus, at the higher levels of neural systems randomness is most likely still the optimum design even though specific sub-systems, vision, digestion, breathing, can be optimized in special ways.

The point of this is that, if one were designing an artificial intelligence random interconnections would be called for (although employing our brain's pattern of greater density of interconnection to near by neurons and less to distant ones). And, the point is that that is apparently the case in the cerebral cortex, the "thinking part" of our brains as compared to the retina, the seeing (but not the understanding of what is seen) part of our brains.

### *PROCESSING OF UNIVERSALS*

One simple level of first order universals processing is not enough to operate the vision mechanism. For example, it contains no provision for dealing with changes from input image to input image, changes which carry information about motion, growth, death (lack of change) and so forth. The same is true of any other sensory input system, hearing and so forth.

The process already described must take place again a number of times. At the first level the sole input was data from the sensor array. At the next level the input is that data and the output of first level neurons. Processing can then yield a more highly processed second level, and third, and so forth. This performs a few levels of re-mapping, re-encoding, and further re-mapping and encoding of the first-level or prior-level, itself mapped and encoded, description in terms of somewhat more sophisticated universals. The system progresses from simple, fundamental micro-universals to more and more sophisticated and abstract universals.

While it is convenient to think of the operation as taking place in discrete layers of neurons and to ignore inputs to earlier level neurons that come from later level neurons, that does not conform to the real situation. In the evolutionarily developed "wired-in" systems like the early levels of vision processing that is somewhat the case; however, even there there is interaction such that levels are not completely discrete.

But for the more sophisticated higher levels of neural activity, those closer to or actually part of the intelligent processes, the concept of levels and arrays must yield to a complex broad body of interaction. Yet that body still operates on the underlying principles of universals, learning by threshold adjustments and Boolean logic implementations.

Which leads to the issue of what are thoughts, memory, the next level of sophistication in neural mechanisms ?

## *PART 4 -- CONCEPTS, THOUGHTS, THINKING AND MEMORY*

As this discussion has developed so far, the neural system is dealing with inputs from the external world. Such an input is a specific example and has characteristics that are peculiar in combination to it, alone. Each of those characteristics is an example of some universal in which the example participates. The distinction between an input example and a universal is important.

The universal does not exist external to the neural system in the sense that it has no representation there. (It is, of course, the commonality among all possible input examples with regard to the characteristic that the universal represents.) It exists in the neural system as a configuration of neurons and their thresholds that is able to discriminate between the presence or absence of that universal in an input example that is presented to it.

The input example exists in the world external to the neural system. It does not exist within the neural system except as a brief representation in terms of the universals that it participates in. That representation consists of a momentary state of some set of neural *1's*, neural *firings* by the appropriate neurons.

The universal is relatively permanent in the neural system being the "wiring" configuration of interconnections among the neurons plus their threshold settings. The latter do not change rapidly in amount except in the early learning phase of the system's operation.

The input example is temporary in the neural system and is merely the pattern of which neurons are then firing.

A *concept* is a *mental universal*; that is, not only the universal but also any specific examples of circumstances that have characteristics fitting it, can only exist and act in a neural system, a rational mechanism such as a brain. This is as distinguished from *material universals*, characteristics of descriptions of material things having specific examples external to the neural system. See Figure 4-1, below.

Universals

| Material ≡ Descriptions | | Mental ≡ Concepts | |
|---|---|---|---|
| blue | soft | good | trouble |
| round | heavy | busy | anxious |
| visually depicted letter "e" -- a spatial form | | mentally conceived letter "e" -- its role in language | |

*Figure 4-1*

A universal, whether material or mental, a description or a concept, potentially exists whenever a set of examples have something in common. The universal is overtly expressed only by and in a functioning mechanism that is capable of abstracting the universal from a group of real samples. The universal, itself, has no tangible existence other than that. But, the specific examples of a material universal, a description, are material.

Concepts are purely mental. They arise and exist only in a functioning neural mechanism, a brain. Both the universal, that is the concept, and specific examples of circumstances fitting the concept are mental actions, only. A *thought* is such a specific example fitting a concept. Just as is the case with a specific example of a material universal, a thought, which is a specific example of a mental universal, is represented in the neural system only momentarily by the firing of the appropriate neurons, the ones whose output firing means that the universals of which the thought is a specific example are indeed present in the thought -- in effect *are* the thought.

A thought can also be of, about, a specific example fitting a material universal rather than a mental one (thinking about blue or soft objects, for example). A thought can also be of the universal itself rather than of a specific case that fits the universal (even as in reading this our thoughts are about universals). Such a thought is still only a thought, not the functioning universal, even though such a thought is our only way of consciously, overtly, being aware of the universal.

We are aware of and can control our thoughts (but how that and how thinking in general occur remains to be developed, below). Our overt awareness of our universals, of the "wiring" and set of thresholds in our neural system, and our ability to control them is much less.

For example, if something seems to us to be honest or dishonest the distinction is seemingly automatically made by our mind at an unconscious level. We instinctively and automatically have the opinion without any thinking having gone on. That is the operation of the universal concept "honest" or any other universal.

If we, in fact, think about the issue trying to decide if it is honest or dishonest that process takes place because the specific example is sufficiently borderline or so complex that our universal extraction and identification system yields a "no decision" output.

Of course, just as with the words here written on this page, the purely mental concepts and thoughts can be recorded in writing or other media of communications. But, the recorded form, as this page, is merely a code that causes the concepts and thoughts to arise in a rational system able to read the writing, able to decode the record. The concepts and thoughts themselves exist only in that rational system, a mind. They exist there only in the combination of that mind's "hard wiring" and its developed, learned, thresholds.

Yet, a mind starts with nothing and what the mind develops, learns, comes from interpreting sensory input from the material world, that is from the extracting of material universals from sensory data. Then, how do concepts arise at all ?

*Thinking* is associations and sequences of thoughts, that is associations and sequences of specific examples fitting certain universals. Each (momentary) thought is a particular set of (momentary) neuron firings. Thinking is sequences of such firings of particular sets of neurons, the content of the set changing somewhat from firing to firing. Such sequential firings, such thoughts, associate, that is form a succession, become a sequence, develop the trend of the thinking, by having in common parts of the logic for their universals.

For example, and greatly simplified, suppose that thought *#1* consists of universals [a, b, d, f, g] out of the *26 [a ... z]* total universals available in this simple example. The next following thought consists of the prior *[a, b, d, f, g]* plus *[k]*. The third thought consists of the set comprising the second thought less *[d]*. The three such thoughts in that sequence and because of those changes in the included universals are "a line of thought", *thinking*.

In the preceding section it was pointed out that a specific neuron is usually involved in a number of universals rather than being dedicated to just a single one. It was pointed out that the result was a drawback in that the thresholds must be attempted optimum compromises among the family of universals in which the neuron participates.

But, it was also presented that this was a tremendous advantage in that it made thinking possible. Thinking is associations and sequences of thoughts. A thought is the firing of a particular set of neurons. Those neurons as a set, collectively represent that thought. But individually, each neuron also represents a part, a component, of a number of other possible thoughts. At the moment of the current thought those other thoughts are not active because their exact complete set of component neuron firings is not active; only some parts or pieces of them are active.

However, the activation of the current thought could, with only a little additional help, result in the activation of one or more of those other thoughts that share a significant proportion of their neurons with the current thought. That "little help" would be something that has the effect of activating some other related neurons and / or deactivating some of the currently active neurons. And, because of the sharing of neurons, of universals, in common between the two thoughts, the successive thoughts will be related; they will tend to follow logically in terms of rational thinking.

While neurons participate in more than one universal they clearly cannot participate in contradictory universals because it would be impossible to achieve a compromise threshold that yet worked for both of the contradictory universals. In general the set of universals in which a neuron participates must be a somewhat related family not totally unrelated. That is part of the nature of universals, their condensation of the characteristic's various appearances into a commonality.

The neural network in which this process of thinking takes place is not the type that is relatively simple and to a fair degree a layered type structure as encountered where sensory input occurs. Rather it is the most complex and sophisticated form of neural network, that which has a very large amount of interconnection with relatively little layering and involving a very large number of neurons overall and in each thought.

From equations 4-1 and 4-2, on below, the number of different possible thoughts that can appear in only one percent of the total human brain's number of neurons is the immense-beyond-comprehension number: `1,000,000 ... (30,000,000 zeros or about 10,000 pages of zeros) ... 000`.

With the extensive interconnection of neurons, including the feed-back or recirculation of output firings as inputs elsewhere in the network, and with the essentially continuous sensory input constantly delivering new data, the "little additional help" necessary to progress to a next, related, thought is constantly present. Inevitably, then, the existence of a current thought results in an immediately following next thought and that next thought is inevitably related to, but not identical to, the former thought.

The associations and consequent transitions from thought to thought are then progressive changes of one, usually some, and perhaps rarely all, of the specific individual universals that comprise the current thought. In the complex neural thinking structure with thoughts involving inconceivably large numbers of universals the opportunities for a variety of associations are quite large.

This overall process is what we call *thinking* (but not, yet, purposive thinking). It is a process that can take place in neural systems over a wide range of size and complexity. Certainly man thinks. But thinking is also performed by dogs, birds, snakes and beetles. In each of the cases as the size and sophistication of the neural system is smaller and simpler then the complexity of the thoughts is reduced. But, the thinking takes place.

The sequence of thoughts, which thought comes next, which specific mental example fitting what universals is the next to appear, is determined by the interaction and relative significance of the universals of the current and the prior thought(s) plus the "little additional help", the effect of new sensory input and of the feed-back of current firings to recirculate in the network.

At the same time each thought can modify the then existing universals. Since the universal is an abstraction of a common element from a family of samples, then if the thought is a new sample added to the family, the thought must produce at least a small change in the pertinent universals. Each neuron's firing reinforces its threshold and each failure to fire de-inforces its threshold.

The result is an iterative process of evolving universals and sequences of specific examples where the examples modify the universals and the universals determine what the various available directions that the sequence of examples may take is.

· Early thinking, learning, operates with material universals, only, and produces some relatively simple concepts.

· Most thinking operates with material universals and existing concepts and produces changed and new concepts.

· Abstract thinking operates purely with concepts and produces more and changed concepts.

Thoughts, thinking, inherently involve the developing of new concepts from the interaction of existing material universals and concepts. That is accomplished by the changes in the thresholds which results in the formation of the ability to perceive new associations, relationships, among the existing material universals and concepts. Those new associations and relationships become new concepts.

A *memory*, that is a thing remembered, is a *thinking pattern*. It is a short or long sequence of thoughts, simple or complex. To the extent that the memory is mentally repeated (the thinking through the sequence of thoughts again) to that extent its thresholds become more firmly set; the memory becomes more permanent. To the extent that the memory is not repeated, to that extent other thoughts that use some of the same logic as is used in that particular memory, produce threshold changes that degrade that particular memorization.

Access to the memory, that is the remembering of it, is via the same kind of associations of thought universals as in any thinking. To access the memory we must think of something associated with it, something that will trigger the sequence of thoughts that are the memory.

Thus, memories reside in a diffused, distributed manner over a large number of neurons. They are not in some separate "library" or "memory file cabinet" of the brain. They are "right out there" intermixed with and inter-operating with the brain's overall activity. The only difference between a *remembering* and a *thinking* is whether the pattern of thoughts is new, original, or is the retracing of an earlier pattern.

What with the vast amount of information that we remember and the complexity of our thinking, one wonders how our neural system can contain it all. Of even more concern could be that, with thresholds being constantly affected by current mental activity how can things learned and things remembered last a long time ?

The four by four array examined earlier contained only sixteen discrete elements -- in effect neurons. Yet that array is capable of representing $2^{16} = 65,536$ different patterns. The human brain contains on the order of *one hundred billion* neurons, about $10^{11}$. Let us arbitrarily assign *10%* of those to sensory, motor, automatic (for example heart beat) and intercommunication activities within the body and brain. (That is quite generous. A *Tyrannosaurus Rex* had a brain of fewer than *10%* a human's number of neurons for all purposes yet it did a pretty good job of functioning.)

Let us then recognize that the complex human brain has a number of regions of specialization. One local region interprets vision; another deals with language, another handles emotion, and so forth. Let us provide for one hundred such sub-systems. Then any one such sub-system would have

$$(4\text{-}1) \quad 10^{(11 - 1 - 2)} = 10^8 \text{ neurons}$$

and could represent

$$(4\text{-}2) \quad 2^{10^8} = (2^{10})^{10^7} \approx (1000)^{10^7} = (10^3)^{10^7}$$

$$\approx 10^{3\cdot 10^7}$$

$$\approx 1{,}000{,}000, \ldots \text{[30 million zeros]} \ldots ,000.$$

different patterns per each such sub-system.

Even our neural system, having that great capacity, is not able to really appreciate what an immense number that is. At the rate of a page being able to contain about 3,000 zeros it would take 10,000 pages of zeros just to write out the number -- to write it down not to express the value of the number. (It takes four digits to write down "1000" but it has the numerical value *1,000*.) That vast capability certainly suffices for our neural system, our brain, to readily learn and retain everything that we give it over a lifetime.

Yes, a certain amount of memory loss occurs because of disuse of some memories or learnings and the consequent blurring of their thresholds. And yes, a brain cell dies here and there regularly and takes its participation in the logic with it when it goes. But those degradations are negligible in the overall system. The number of neurons involved in any single thought or memory is so large that a problem with, or a failure of, a single neuron here or there, now and then, is of no importance.

On the other hand, a popular saying that is valid in its context is that "we are what we eat". It is likewise true that our mind (which, after all, is our conscious selves) is what we think. We tend to become, to think as, to behave as, that which we feed into our neural logic networks and threshold settings. That is something to think about.

The word *conscious* has just turned up and that leads to the next aspect of neural systems: how does *thinking* as just presented become *purposive thinking*, what is *consciousness* and how does it happen ?

## *PART 5 -- PURPOSIVE BEHAVIOR: GOALS, MOTIVATION, CONSCIOUSNESS*

*Purposive* behavior is behavior that involves goals and making choices from among alternative options. If there is only one option then the behavior is not purposive. But, if any one of two or more alternative actions can be taken then there is a choice and the selecting of one of the options is *purposive*.

For example:

| Purposive Behavior | Non-Purposive Behavior |
|---|---|
| Deciding when or what to eat | Digesting |
| Holding the breath temporarily | Routine Breathing |
| Reading a book | Dreaming. |

*Figure 5-1*

A *goal* is a type of purely abstract thought. That is, it is one of a number of types, forms, occurrences of thoughts that are activated by neural associations, alone, with no sensory inputs. The *goal* itself may be material or abstract. A material goal might be, for example, turning on a water valve. An abstract goal might be to do the sum of 21 plus 32 "in one's head".

Therefore, a material goal is an abstract thought describing / corresponding to a material state that does not presently exist but a state that is *intended* or *desired* to be obtained. (What *intended* and *desired* are and how they come about is developed further below.) Therefore, it is a firing of a set of neurons that signal a specific set of universals (the set of universals that describe the material state intended or desired). The signaling of the same set of universals where the signaling is set off by sensory input signals means that the goal has been accomplished / realized (or at least would appear to so be).

Thinking about the goal is repeated goal thoughts. The realization of the goal is repeated goal-satisfying sensory inputs. The goal is described by the set of universals that either of those actions signal by their neuron firings. The neuron firings have a set of outputs that are produced whenever the goal's set of universals is present in the cause of the inputs to those neurons. If those inputs are neural the process is thinking of the goal. If those inputs are sensory the process is sensing that the conditions that make up the goal do materially exist (the goal is satisfied).

An abstract, or mental, goal is, then, a set of universals the triggering of which as a thought by some associated preceding thought represents the goal and the triggering of which by a set of some other related abstract thoughts may represent the accomplishment of the goal. (Or it may represent work in process on accomplishing the goal, or thinking about how to, or about whether it is possible, etc.). A material goal is the same except that sensory input must be involved in triggering the representation of its being accomplished.

*Purposive behavior* involves:

· setting a goal,

· making choices among options to achieve the realization of the goal,

· comparing current progress made against the ultimate goal,

· adjusting behavior by modifying choices,

and so forth, all iterated until the goal is achieved or the process is interrupted.

When such behavior is present then the neural system is conscious. When such behavior is absent the system is not conscious; it is unconscious. (We sometimes refer to an aspect or event in our behavior as being unconscious even though it occurs when we are, overall, conscious. That is because a neural system can have unconscious, that is non-purposive, aspects of its behavior even while it is overall conscious, that is behaving purposively overall.)

We humans develop patterns of purposive behavior with which we become so familiar that we can initiate them and then cease to pay attention to them for a while. A common such experience is to be driving a car and suddenly realize that you have been thinking about work, or dinner, or whatever, and that you don't seem to know, for a moment or two, where you are or how you got there. You then realize that, obviously, you drove the car to where you now are, apparently you did it properly and safely, but you did it without attending to it. Your attention was on some other purposive behavior running through your mind.

Then:

· how do we pay attention,

· how do our goals come to be, and

· what causes our neural system to seek to satisfy our goals rather than ignore them ?

## *"Pain", Motivation*

As previously presented, the rate of repeated firings by a sensory neuron delivers information as to *how much*: how loud a sound is, how bright a light is, the magnitude of a touch, and so forth. That kind of information is essential to an organism's functioning. It greatly enhances the sensory information's description of the material world, and for motor purposes (muscle operation, physical action) *how much* supplies progress reports so that, for example, an object can be grasped without over-reaching or under-reaching, without squeezing so hard as to crush it or so lightly as to let it slip out of the grasp.

The *how much* data is natural to biological electrochemical sensors. Brighter light or louder sound or larger touch delivers more energy which more easily triggers the sensory neurons' firings. But if the sensory input is too large it can be destructive:

- too loud a sound damages the ear (and may also represent an external threat of some kind)

- too bright a light destroys the eye (and also may signal an external danger)

- too large a touch (cutting, breaking, burning) injures the body.

The most (evolutionarily) early, simple neural networks in early, simple organisms received *how much* data from their early, simple sensors. If it was a signal of *too much* the sensor might be destroyed and the organism most likely would fail to survive. But, some organisms responded to the *too much* sensory inputs by action to avoid the cause of the excessive input, by action to get away.

That must have been fairly common because in simple neural systems the sensors would be closely linked to the motor action neurons. The most simple such early neural system would have consisted of a sensor neuron that was connected directly to a motor neuron. Such a mechanism could, for example, produce opening in response to light (like day lilies), closing in response to a touch within the food receiving-digesting cavity (not unlike the action of today's sea anemones), or flagella waving in response to excessive temperature (effectively causing swimming away). The *too much* sensory signals producing a rapid neuron firing rate deliver a rapid rate of input pulses to motor neurons tending to produce some kind of action, some kind of change. When confronted with destruction any change is preferable to no change.

Whether that kind of response was initially naturally common or rare, it would have significantly increased the survival rate of the organisms behaving in that manner. They would likely have become the only type organisms surviving into their future and contributing evolved characteristics to their successors. Avoidance of danger, harm, destruction, must have become an operating principal of simple, early neural networks very early in their existence.

The pattern of

- too much sensory input producing

- greater sensory neuron firing rates, producing

- greater motor neuron excitation, producing

- action, motion that changes the situation,

naturally must have become an evolved survival characteristic of simple neural systems at a very early stage in their development.

In only quite slightly more sophisticated yet simple early neural systems the same response would develop to *too little* sensor input. The `not` Boolean operation is an essential of the logic of neural nets and neurons have both excitatory (normal) and inhibitory (*not*) input dendrites. The *not* of a *too little* firing rate would be a rapid firing rate, a *too much* signal. Of course some cases of *too little* can be just as dangerous as those of *too much*. For example, we humans react strongly to *too little* good air to breath.

We humans retain those same early-developed mechanisms. If one puts his finger on an oven at room temperature he can keep it there all day if he wishes. But if the oven is burning hot, then the moment that the finger touches the oven it is quickly withdrawn, withdrawn automatically without any thinking about it. The *too much* neural signals from the finger's sensory neurons directly trigger the arm motion motor neurons by interconnection in the spinal column without the brain as neural logic intermediary. It is a sensor-motor direct connection when the sensor signal is *too much*. For a room temperature oven the touch sensor signals go to the brain for processing.

Further, we humans exhibit examples of only a moderately more sophisticated neural response to *too much*. The eye, for example, automatically and very quickly shuts, shuts quite quickly, when an object is moving rapidly toward the eye. Our brain is not involved. We have shut our eyes before even being aware consciously that there is a problem.

That response is not a direct sensor-motor type of action. Significant neural network processing is needed to convert the raw visual picture into information that says

- a strange object is in <u>motion</u> in the visual field

- its <u>trajectory</u> is such that it will endanger the eye

- it is moving at rapid <u>speed</u>

- therefore immediate, quick, protective action is needed.

Most likely, in the eye that neural logic is performed in the several layers of neurons underlying the retina. It would appear to be too rapid a response to take place in the brain.

Eyes developed long after the "early, simple neural networks". But the long established character of those early such networks, that of treating excess as dangerous and of automatically taking action to correct the situation, appears developed into a greatly more sophisticated version in the eye's response to a detected danger. A complex set of universals represented by a significant number of neurons as a set producing *too much* signals collectively is involved in that action.

With the evolution of species, as neural networks became larger, more sophisticated and more complex, the evolved type species' operation of the *too much* response became more sophisticated, that is:

- larger,

> They involved greater numbers of neurons and in more numerous universals which described more complex thoughts.

- more sophisticated,

  They included logic to determine whether a response is really needed, to consider alternative responses, and even to put together patterns of responses.

- and more complex.

  They developed the ability to deal with multiple *too much* signals at the same time, the ability to arrange the corresponding multiple responses, to relate and prioritize the responses, and so forth.

Such behavior is the setting of goals and the making of choices among alternative courses of action. It is *purposive behavior*.

The *too much* signal and the reaction that it triggers ranges from the very simple sensor-motor type cases (the hot oven) through the significant neural processing type cases (the eye shutting) to more and more sophisticated motivations and resulting actions. Just as our thoughts are the patterns of which neurons are firing at a particular moment, so our conscious purposive behavior, our performance in life at home, on the job, as parents, in love, and so forth, is our responses to highly sophisticated and complex sets of neural *too much* (and *not-ed too little*) signals.

The signals involve, are related to, are the equivalent of, are *pain and pleasure* (*pleasure* is *not-pain*). When the signals involve material sensor input the consequent responses normally involve physical action, that is material response to material sensor input. When the signals involve non-sensor input, that is abstract thoughts, concepts, the consequent responses normally involve non-material actions, that which we refer to as *intentions and desires*. Of course, combinations of material and non-material responses are frequently the case.

Very early in the evolution of neural systems those systems evolved to treat extremes of neuron firing rates, low or high, as being: bad, a sign of danger, something to be avoided, triggers of corrective action. At the sophisticated level we now refer to the effect of excessively non-moderate neural firing rates on us as meaning that the related material or abstract objects (described by the universals the neurons of which are so immoderately firing) are:

- painful,
- unintended
- undesired
- unpleasant.

The opposite, neuron firing rates that are neither too great nor too small, that is moderate rates then signify

- comfortable
- intended
- desired
- pleasant.

It could be said that we spend our lives seeking to have our neurons firing at a rate well between the *too much* of a too rapid rate and the *too much* of a too slow rate. One could say that a state of moderate neuron firing rates is what we call *happiness, pleasure, contentment, joy*.

Or, perhaps, the greatest pleasure or joy, the best sensation, corresponds to neural firing rates that are as near to *too much* as possible without being so strong as to mandate corrective action. Our human experience would tend to indicate that we behave that way in some cases, that we crave excitement so long as it does not go over the boundary into the dangerous.

Or, perhaps, for different kinds of good, pleasure, and so forth, different neural firing rates apply -- contentment corresponding to a moderate rate, great joy to a rate near *too much*. Perhaps, the neural network involves a mix of different correspondences between neural firing rates and various subjective feelings of *good* for the variety of different such feelings. And, perhaps, that mix and the associated firing rates change throughout the individual's life as the neural system has more and more living experience, more learning and adjustment of its thresholds, as it evolves with the person's mental and emotional growth. And perhaps the precise state of the system is a little different for each individual -- each having a unique set of responses.

### *RESPONSE TO THE "TOO MUCH" SIGNAL*

The above has primarily discussed only the input aspect of the evolutionarily developed *too much* signal and behavior. Of equal importance, with the signal of rapid neuron firings conveying *too much* type information, is the action that the neural system takes when such signals appear. In the simple early neural systems the response was some kind of motor neuron (motion) action initiated by a direct sensor-motor neural connection.

In more sophisticated neural systems, such as that of the eye's response to danger from a rapidly approaching object, the response is, again, a motor response, the closing of the eye; however in this case it does not take place by direct sensor-motor connection. Rather, an analysis of the sensory data takes place and a motor signal is sent to close the eye if that analysis indicates that such a response is called for. The response is part of the action of a complex neural system that, quite in addition to the actual analysis of the visual image for the purposes of vision, takes such actions as adjusting the lens of the eye to optimum focus of the image on the light sensors, the retina, and opening or closing the eye's iris to admit more or less light as the circumstances call for.

In the sense of the highly evolved system's behavior being a highly evolved response to *too much* signals, it must nevertheless be a kind of response that is intended to remove or relieve the cause of the *too much* signal. That is, no matter how highly evolved and abstract the neural system, its response to inputs signifying pain, bad, unintended, or undesired must be a response that tends to or is intended to relieve or improve the situation, to remove or reduce the cause of the *too much* signal.

Thus sophisticated systems, such as those of we humans, respond, for example, to the frustrated desire for a sweet to eat by exciting our motor neurons to cause our walking to the cupboard, selecting a cookie, closing the cupboard and eating. Even more, in general they produce our performance of the routine of living: arising in the morning, eating, going to and performing the tasks of the day, and so forth.

But, sophisticated systems are really very complex. They can learn and act not only on that it is bad or painful (a too much signal) to fail to experience for example:

- a luscious desert sweet (a signal being as close to too much as possible without being excessive),

- or a desirable result (pleasantly high neural firing rates associated with buying the new sports car one has wanted);

but, even more, they can mandate, for example:

- revenge to relieve the pain of an affront or a loss,

- or then a declining to take the desired revenge in order to relieve the pain that revenge is too contrary to the neural system's own standards of character and behavior.

Let us consider just how immensely complex a neural network the size of the human brain is. (Not that we humans are necessarily the apex of possibilities. Larger and more sophisticated neural networks are possible, both naturally and artificially. We humans are merely the most sophisticated such systems currently known to us.)

In *The Origin and Its Meaning*[1] and in other works and analyses it is estimated that the total number of particles (protons) in the entire universe is on the order of $10^{84}$. Equations 4-1 and 4-2 estimate that only one percent of the human brain is enough neural capacity to support $10^{30,000,000}$ different thoughts, memories. That is, one percent of one human brain, of just the mind of one person not all people, supports

*1,000 ... [ten thousand pages of zeros] ... 000.*

thoughts, ideas, memories versus there being only

*1,000 ... [one line of zeros] ... 000.*

particles in the entire universe.

Nature abounds with examples of change in quantity, that is change in the amount of something, producing a resulting change in quality, that is sharp and distinct change in the something's behavior and characteristics. Some examples are: the transitions from ice to liquid water to steam as the amount of heat in a body of water increases, the critical mass of a nuclear bomb, the tree that succeeds in growing taller than its neighbors over-shadowing and stunting or killing them by taking their share of light for itself, and so forth.

So, likewise, the vastness of our neural networks results in their exhibiting: our ego, our sense of self, our rational powers and our powers of pure abstract thought.

Nature, material reality, abounds in very complex forms, for example the shape of a coast line or mountain range, the shape variation among the individual leaves of the same tree, or the distinct individual shape of each cloud. In general such forms come about through the repetitive action of relatively simple processes with minor variations from repetition to repetition. Fractal mathematics is the study of such processes.

Taking a very simple case, consider the geometric result of starting with the pattern of a simple cross as in Figure 5-2(a), below. From it two legs will be removed, as in Figure 5-2(b), so that the results can be displayed on the available page. Then the pattern is iterated. At each such iteration each individual component of the prior pattern is replaced with the entire prior pattern as in Figure 5-2(c).

```
       o                             o
(a)  o o o                      (b)  o o
       o

                                     o
(c)                                  o o
                                     o   o
                                     o o o o
                                     o       o
                                     o o     o o
                                     o   o   o   o              and ad
                                     o o o o o o o o       ...  infin-
                  o                  o               o           itum
                  o o                o o             o o
                  o   o              o   o           o   o
                  o o o o            o o o o         o o o o
       o          o       o          o       o       o       o
       o o        o o     o o        o o     o o     o o     o o
       o   o      o   o   o   o      o   o   o   o   o   o   o   o
       o o o o    o o o o o o o o    o o o o o o o o o o o o o o o o
         1st          2nd                     3rd
      Replace-      Replace-               Replace-
         ment          ment                   ment
```

*Figure 5-2*

Then consider the effect if the form that replaces every circle, that of Figure 5-2(b), above, were to vary randomly such, as for example, among the forms of Figure 5-3, below.

```
                      o                     o
o         o           o          o o        o o        o o
o o       o o o       o o        o o        o o        o o o
```

*Figure 5-3*

Now suppose that each of the circles above is a neuron in a neural net. Furthermore, the interconnections between them are not an identical pattern repeated precisely as Figure 5-2, above, but rather are variations on a general pattern of: (1) more interconnections to near neurons falling off gradually to fewer connections to more distant ones and (2) randomly selected choice of the particular neuron for each connection. Then, still further, introduce different thresholds for different neurons and random selection of inhibitory versus excitatory interconnections.

Then contemplate what the above figure, which has 81 circles at the 3rd Replacement, would look like in three dimensions rather than two and with the added complexity of all of the above modifications and with $10^8$ circles (the number of neurons in $1/1000^{th}$ of the human brain) instead of only $81$ circles. The result would be an immensely complex system, one so complex as to be beyond our ability to truly imagine or visualize it yet a system comparable to about $0.1\%$ of our human brain.

The power and capability of our intelligent rational system, our neural network, is so vast that we are unable to really comprehend it in terms of those numbers and its vastness. Except -- that we experience every moment the wonderful things that it does and that it can do. We realize the power of such a neural net in our daily living experience and we are able to understand the basic underlying mechanisms and arrangements that produce that result:

- the operation of individual neurons,

- the implementation of Boolean majority logic by networks of neurons,

- the learning that takes place in such networks because the level of each neuron's threshold affects the specific form of the logic implemented by that neuron,

- which learning takes place automatically -- repetition modifying thresholds,

- the overall effect being the recognition of universals,

- combined with a system of response to excessive input signals to neurons and neural subsystems that causes action to be taken that tends to reduce the excessive input if it is dangerous and that seeks to maintain it if it is desirable.

And that system continuously subtly changes with the changes in thresholds due to our thinking and our sensory experiences. Our thoughts, each the momentary signaling, the momentary neuron firings, of a particular large and complex set of universals, some being descriptions of material reality and some being abstract concepts, follow on in a train of thinking as successive such thoughts associate through commonality of the majority of their universals and the addition of some other universals that were not included along with the dropping of some of those that were included.

So accordingly, the complexity of our neural networks results in their exhibiting the complex behavior and capabilities that we humans exhibit: our ego, our sense of self, our rational powers and our powers of pure abstract thought.

## *PART 6 -- FREE WILL AND PREDESTINATION, ARTIFICIAL INTELLIGENCE AND CONCLUSION*

The problem of predestination versus free will has plagued philosophy, religion and science from their very beginnings. Predestination means that the course of all events, great and small, is already determined, is pre-destined, and cannot be altered. Free will refers to the freedom of each rational being to make choices among alternative paths of events. The two are clearly in direct conflict.

This problem is quite severe because both the logical case in support of predestination and the logical case in support of free will are each quite strong -- even though the two are directly contradictory.

Pro - Predestination

- Any religion that involves a creator-god finds itself forced to attribute to that creator full knowledge and understanding of its creation, the universe. That must include all events throughout all time. For those to be known to the creator-god they must be fixed and determined for all time. There can be no choosing among alternatives.

- Put another way, if some aspect of the creator-god's creation remains unknown to the creator-god until he "waits to see what happens" then the god is not infinite, not all powerful, not all knowing. Even if the god is deemed "outside of time, timeless" so that all events in all of time are mutually present to him the immutability of those events is necessary to the conception of the god.

- Thus, one can have no predestination and a defective god or predestination and a non-defective god but one cannot have the best of both.

- This same problem applies to science. Fundamental to science is that the universe behaves according to "laws", that the universe consistently and reliably does the same thing given the same prior state of conditions. If that were not so there would be no point to science because its results would be meaningless and would have no use, no application.

- But, if the universe always follows those laws then everything that happens in the universe is predictable and certain according to the operation of those laws. Thus, science is forced to accept predestination as a requisite of the validity of its work as science.

(The science of the 20th Century found what appeared to be a way around that problem, which is addressed shortly, below.)

Pro - Free Will

· We each and all "feel" that we have free will. We make choices and those choices result in different events occurring as we experience it. In a practical sense no rational person could be convinced on the basis of his personal experience that he lacks free will.

· Free will is essential to the social functioning of mankind. Without free will, with everything pre-determined, with no choices having any effect, then

- there is no responsibility,

- there is no motivation,

- we are mere automatons blindly following the program of fate.

Society and individuals cannot function without responsibility and motivation, without the imperative to avoid bad and seek good, to maintain survival, and to achieve progress. Those are essential to man and society.

So:

On the one hand predestination cannot be avoided because the universe does behave consistently according to the patterns of behavior that we call "laws", and

On the other hand we individuals and our societies must have free will because we feel that we do and we cannot function without it.

This profound dilemma is one of the reasons that 20th Century physics so readily adopted its system of uncertainty, probabilistic mechanics, and statistical behavior. That system allowed science to have "laws" but not laws that required predestination. The certainty of hard, solid laws was replaced with laws involving probability and statistical chance. That way predestination could be dropped and free will could be given its necessary sway.

But uncertainty, probabilistic mechanics, and statistical behavior do not really solve the dilemma. If the universe operates according to physical laws then the reality of predestination is unavoidable. And if the universe does not operate according to its physical laws then it does not operate, and does not exist, at all.

Then what about free will -- the free will that each of us "knows" that he has and without which society cannot function ?

The fact of the matter is that our functioning in our lives as individuals and society's functioning overall do not depend on the existence of free will in us individually nor as the members of society. The only requisite is the perception of free will, that we think that we have free will. If we each think, believe, inherently know, that we have free will then we function accordingly whether in fact we overall objectively do have free will or we function in a totally predetermined state.

We are all convinced that we have free will because our life experience so demonstrates to us. Therefore, we do have free will and it has the expected affect on our individual and social behavior.

As it so happens, a purely mechanical universe, our universe, is operating from its original starting condition, according to a fixed and immutable set of rules of its behavior (which we call physical laws), so that every single event and action, moment by moment, cosmic and microscopic, universal and personal, is a predictable, theoretically calculable, inevitable consequence of the prior existing state and the operation of those same physical laws.

That has no effect on our free will because:

- we are convinced that we have free will and we live using it,

- no one, not one of us, not all of us, ever could actually perform the calculations and extrapolations to discover the course, the fixed and immutable course, of future events.

- wherefore the future is as unknown and, to us, not pre-determined as if that were truly the case.

If this is difficult to accept, there is a way out, a rationalization of the situation. It is the calculation done in the prior part for one percent of the human mind now expanded to all of that mind.

> One human brain, just the mind of any one individual person [*a person's operating free will*] involves on the order of a capability of
>
> `1,000, ... [one million pages of zeros] ... ,000.`
>
> possible thoughts, ideas, memories.
>
> On the other hand the total number of particles in the universe [*with behavior predetermined by physical laws*] is on the order of
>
> `1,000, ... [one line of zeros] ... ,000.`
>
> about $10^{84}$ mere particles.

*ARTIFICIAL INTELLIGENCE*

The terminology "artificial intelligence" refers, of course, to a man-made rather than naturally occurring device or entity that exhibits "intelligence", that is, that exhibits the type of behavior that has been described as the behavior of complex, sophisticated neural networks in the preceding sections. There is nothing "artificial" about "artificial intelligence" except that it resides in a hand crafted, manufactured product not a natural biological being.

The development and construction of such rational systems, based on the operating principles of our own (and all animals') neural networks is quite within the range of the possible. It is also desirable in that the result could be much more effective automatic machines and processes. However, such development and construction would be a very difficult and immense task, and it eventually will raise some ethical questions.

In a sense, the development of such systems has already started. Witness to that is the arrival of "artificial intelligence" as a field of scientific and engineering specialty and one of the results of that activity, simple neural networks as commercial products for specialized, limited applications.

But, the task of developing a neural system able to perform at the human level, or even at the chimpanzee or dog level, will be a large project beside which other large projects such as the pyramids of Egypt, the development of the nuclear fission bomb and the development of space travel pale to relative insignificance.

There is so much to be learned before significant useful complex neural systems can be built. Research must reproduce the optimization, that nature obtained over a billion years of evolutionary trial and error, of the neural layers located at the sensory organs that directly perform the initial processing of sensory data, that for the eye and the ear especially. Similarly, the structures of the main abstract neural system (those found in the human cerebrum) involve many evolved solutions that research must re-develop.

Then there is the matter of instinct, which corresponds to "built-in", "pre-set" thresholds. But, which thresholds ? And, how much should the pre-set value be ? The questions seem endless.

Then, when one has constructed a neural network the network must be taught. In effect it must learn a learning corresponding to what we, or animals, learn during the development from birth to adulthood.

Man-made neural systems have a great advantage, however. It is not necessary to laboriously educate each individual system. Once a unit has been made it can be reproduced with its learning to that point in time included. The prototype model can be built so that every threshold can be "read out" at any time. The production models can be manufactured "as adults", that is with all of the prototype's learning pre-implanted as initial threshold settings.

One can imagine a prototype being trained to do house keeping. After the training is complete the production models are built with that learning built in. Another unit of the same prototype is trained in child care, another in vehicle mechanical maintenance and repair, and on and on. Especially useful would be emergency rescue units able to go and function where human rescuers cannot.

But, this is all well into the future. What would seem to be a reasonable, achievable, near term objective would be a neural network, a man-made "intelligent" system able to interpret human speech -- not able to understand the meaning of the speech, merely able to type out, for example, an accurate and grammatically correct word-for-word transcription of what the human said to it.

All of this eventually leads to ethical problems. These are of two types. The first is that in making machines to do our work we must not make intelligent slaves. When our machines achieve a sense of self and an ability to suffer, there we must stop to re-evaluate our actions.

The second ethical problem is more awkward. We ultimately should be able to make machines with more powerful neural networks than our own. What that means in terms of "pecking order", "right to survival", "who controls whom", remains to be worked out.

## *CONCLUSION*

The purpose of this paper has been to show how intelligence operates in a broad and fundamental sense, to develop the concepts of how an intelligent system works not to present a detailed design ready for implementation. By analogy, a steam engine has not been designed but the concept has been developed in the description: water boiled to make steam which acts on a piston the longitudinal motion of which is converted by a drive wheel into rotary motion and the action of which is enhanced by condensing the steam on the outlet side of the piston. (The condenser was Watt's great contribution).

And, just as all kinds of steam engines can be designed, built and usefully employed, so also all kinds of neural networks can be designed built and usefully employed from the underlying principles presented in the preceding sections.

That being the case, intelligence, even the level of intelligence we humans exhibit, is a natural phenomenon evolved by nature.

## *References*

**[1]** This paper is extracted from R. Ellman, *The Origin and Its Meaning*, The-Origin Foundation, Inc., http://www.The-Origin.org, 1997.