Thoughts and Thinking

ROGER ELLMAN
The-Origin Foundation, Inc.
320 Gemma Circle, Santa Rosa, CA
UNITED STATES OF AMERICA
RogerEllman@The-Origin.org http://www.The-Origin.org

Abstract: - Most discussions of the brain focus on the neural structure and its biological functioning. But, it is the "software", how the neural components logically interact, that produces the higher level results that we experience in our minds. The development begins with universals and then proceeds through perception, learning, and the processing of universals to:  concepts, thoughts, thinking, purposive behavior, and memory.

Key-Words: - Universals, Perception, Thoughts, Thinking, Neural Logic

1 Universals

The universal is the common characteristic of all elements of a set: E-ness, moving-ness, shirt-ness in the following three examples:
- the letter E, whether capital or lower case, hand written or mechanically produced, large or small, alone or among other symbols, etc.;
- all motion regardless of speed or direction;
- all shirts regardless of details.

Perception, correlating examples, with their universals, involves an input, a data processor, and an output. The processor is a device that operates on the input data to correlate examples with universals. The output is data representing the correlation.

A single input may be a sample of a number of different universals. For example a particular letter E might belong to all of the universals: E, upper case, small, hand written, in ink, red, moving, etc.

For the present the input is data from an idealized eye. The image appears on the eye’s retina as an array of “on” and “off” sensors.

Initially we will use an array of 16 sensors and the system of Figure 1 to refer to a sensor.

![Image](Fig. 1)

The half of the array of Figure 1 that is labeled A will be called “A”. The other half will be called “not A” and be written \( \overline{A} \). We can then refer to element #11 of Figure 1, for example, as being in (1) \( A \) and \( B \) and \( C \) and \( D \).

The letters \( A \), \( B \), etc., are variables. The only allowed values are 1 and 0 (“on” and “off”).

Instead of writing "and" as in equation (1) the notation \( ABCD \) will be used or, for clarity, \( A \cdot B \cdot C \cdot D \).

To refer to more than one element of the array at a time “or” will be used, written as “+”. To refer to elements #10 and #11 of Figure 1a the reference is (2) \( ABCD + AB \overline{C}D = B \overline{C}D \).

Let us consider recognizing a cross, a horizontal line crossing a vertical line. We seek an output of “1” for any, and only, such crosses.

First consider some examples as in Figure 2.

![Image](Fig. 2)

The elements making up each example are:

(3) (a) \( ABCD + ABCD + ABCD + ABCD + ABCD \)
(b) \( ABCD + ABCD + ABCD + ABCD + ABCD \)
(c) \( ABCD + ABCD + ABCD + ABCD + ABCD \)
(d) \( ABCD + ABCD + ABCD + ABCD + ABCD \).

Using the commutative and associative principals and factoring reduces (3)(a) to any one of

(4) \( AB[ C + D ] + CD[A + B] \) or \( AC[B + D] + BD[A + D] \) or \( AD[B + C] + BC[A + D] \).

and similarly for (3)(b, c, and d)

This cross is the “and” of any two variables with the “or” of the other two, that then “or”-ed with the
The eye can deal with about 27,000,000 patterns. The number of patterns in a binary digital array is about 7,000,000, the retinal rods and cones. Instead of 16 sensory elements the human eye has 2 Complex Perception Systems

Instead of 16 sensory elements the human eye has about 7,000,000, the retinal rods and cones. The number of patterns in a binary digital array is

(5) Number of possible patterns

2[number of elements in the array]

The eye can deal with about 27,000,000 patterns.

Taking 1,000 as an approximation to 210 = 1,024

(6) 27,000,000 = (210)700,000 = 1 followed by 2,100,000 zeros.

If the eye saw 10 patterns per second it would take 30 years to see 10,000,000,000 patterns (1 followed by 10 zeros, not 2,100,000 zeros) and inconceivable eons to see all of the patterns possible.

When an image appears on the retina, a family of signals from the sensors goes to the nervous system. The first level of processing identifies basic universals in the input image such as: corners, edges, shapes, motion, etc., like the above cross.

The possible number of such universals is quite large, enough to constitute a complete description of the input, consisting of all its universals with their location and orientation in the image. The input is converted from an array of points in one-to-one correspondence with the original to an array of characteristics of the image.

Referring to the 7,000,000 retina sensors as #A, #B, ... for all 7,000,000 of them, an image on that retina is the Boolean “and” of the on sensors and the “not” of the off sensors.

(7) Sample Image = ABCDEF... [7,000,000 letters]

A group of images, each in the equation (7) form, is the “or”-ing together of the expression for each.

(8) Sample Group of Images =

= ABCDEF... + ABCDEF... + ABCDEF... + ...

Such an expression is the universal of that group of images; any image of the group matches part of the expression and any not a member fails to do so.

This kind of expression can be implemented electronically with logic gates and flip-flops. But, the expressions are cumbersome and require too many logic gates and flip-flops.

More serious, this procedure is only correct for images used in its original set up. It is unable to correctly treat images never before experienced.

Referring to equation (8), if every input exhibiting the universal of interest has sensor #B = “on” and every one that does not exhibit that universal has #B = “off”, both regardless of the other sensors, then sensor #B, alone, represents the universal.

The nature of universals is simplified expressions, a neglect of specifics in favor of broad commonality, expressions simpler than equations (7) and (8).

But how does a system extract a simplified universal from a group of sample inputs?

3 Neural-Type Logic Devices

Neural-type logic differs from Boolean logic in two respects: its operator is the “majority” operation and it uses constants in addition to variables. The majority variables and constants values are +1 or –1. We use the notation M(1st variable, 2nd variable, ...), where the output is +1 if a majority of the variables are +1 and there is no output if not.

Majority logic can implement Boolean logic. For example (using Boolean values): (9) M(A,B,C) = AB + AC + BC

M(A,B,1) = A·B + A·1 + B·1 = A + B
M(A,B,0) = A·B + A·0 + B·0 = A·B
M(A,B,C) = AB + BC + AC

In a neural system, a neuron’s outputs are other neurons’ inputs. Some inputs act on the neuron in an excitatory fashion (natural variables) and some act in an inhibitory fashion (“not”-ed variables).

Neurons have an internal threshold. If it is zero the neuron’s logical construct is just the majority of its inputs.

But, if it is greater than zero, then for the neuron to have an output of +1 the number of excitatory inputs (+1’s) must be the threshold amount greater than the number of inhibitory inputs (–1’s). The effect is as of
as many constants equal to $-1$ as the level of the threshold. Likewise, a threshold less than zero corresponds to there being as many constants equal to $+1$ as the level of the threshold.

The neuron "remembers" the value of its threshold. The special power of the neuron is that its threshold can be changed, changing its constants; changing its Boolean logical construct. The change of neural thresholds to achieve a desired result is learning.

A network in which the outputs of some neurons are inputs to other neurons creates complex majority logic structures. Networks of neurons use the neuron's majority logic, modified by the individual neuron's thresholds, to represent complex Boolean logical descriptions which are the simplified logical representation of specific universals. Complex neural networks are universals representation systems.

To see how this can be done we operate a neural network with its input source the above 16 element array. The output of each of its 16 sensors is interconnected as inputs to a number of neurons and then the outputs of those neurons connected to the inputs of one final neuron. The output of that final, single, neuron is the act of this entire neural network.

But, which sensors should be connected to which inputs, excitatory or inhibitory, of which neurons? Let us assume, for here, that that has been solved.

We now teach the network to recognize the universal cross-ness, to output +1 if the input image has cross-ness and nothing otherwise.

(A) Project an example image onto the array.
(B) Note whether it has cross-ness or not.
(C) Note the network output (is it +1 or null?).
(D) Evaluate the network performance:

<table>
<thead>
<tr>
<th>Input Image</th>
<th>Output</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>cross</td>
<td>+1</td>
<td>Correct</td>
</tr>
<tr>
<td>cross</td>
<td>null</td>
<td>Wrong</td>
</tr>
<tr>
<td>not cross</td>
<td>+1</td>
<td>Wrong</td>
</tr>
<tr>
<td>not cross</td>
<td>null</td>
<td>Correct</td>
</tr>
</tbody>
</table>

(E) Change each neuron's threshold as follows.

<table>
<thead>
<tr>
<th>Neuron Result</th>
<th>Change</th>
<th>Makes in</th>
</tr>
</thead>
<tbody>
<tr>
<td>Output</td>
<td>Threshold</td>
<td>Future</td>
</tr>
<tr>
<td>+1 Correct</td>
<td>down</td>
<td>likely</td>
</tr>
<tr>
<td>null Correct</td>
<td>up</td>
<td>likely</td>
</tr>
<tr>
<td>+1 Wrong</td>
<td>up</td>
<td>unlikely</td>
</tr>
<tr>
<td>null Wrong</td>
<td>down</td>
<td>unlikely</td>
</tr>
</tbody>
</table>

This process inevitably changes the network so that it is increasingly likely to correctly identify crosses versus non-crosses. The network’s learning is accomplished by logical adjustments to each neuron's threshold level. Such adjustments change the Boolean logical construct effected by the neuron's majority operation with the constants represented by its threshold. The Boolean logical operation the neural network performs gradually changes, approaching that of the universal being learned.

4 Aspects of This System

This perception mechanism requires an external teacher to direct the threshold changes after each input. A teacher-free system is needed.

We learn things by repetition. Explanation and experience help, but learning only happens when the lesson is repeated correctly, sufficiently.

- In real world nervous systems there is no external teacher adjusting thresholds,
- learning does occur in such neural systems,
- learning can only take place by means of threshold changes, and
- things are learned by repetition and emphasis;

It must be that the threshold changes occur by an inherent process in the neurons.

When a neuron "fires" its threshold decreases slightly so that a similar "firing" is more likely at the next similar input. And, when a it "does not fire" its threshold increases slightly so that a similar "non-firing" is more likely at the next similar input.

That is learning by repetition. That would tend to change the thresholds of the involved neurons in the direction that makes the same outcome more likely.

If we lapse in regular practice or rehearsal of something learned we begin to forget it. That corresponds to the gradual decay of thresholds if they are not regularly reinforced by repetition.

The neuron is an electrochemical device. It propagates electrical potentials generated and transmitted within it chemically. Its "firing" delivers electrochemical energy, temporarily depleting the neuron’s electrochemical threshold. Likewise, the absence of "firings" enables the neuron’s on-going metabolism, to accumulate more threshold.

Another issue is the system’s operation’s synchronization. It has been implicitly assumed that the data are processed at each neuron simultaneously.

Such simultaneity is unlikely. The various neural dendrites and axons are of different lengths and the signals’ travel times must be different.

Digital computers use a synchronizing clock that generates a train of pulses which are “and”-ed with the input to the flip-flops. Regardless of between pulses, only the conditions at the time of each pulse...
participate in the computer’s operation.

The brain 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.

A biological logic system is not synchronized. Much of the time a neuron puts out no signal. Then it “fires” a +1 signal. That is an input in other neurons so that they experience a +1 or −1 input.

Instead of synchronization neurons operate on accumulation. After its last firing the somewhat depleted neuron accumulates new electrochemical energy from its +1 inputs less its −1 inputs, all in random timing. When that accumulation exceeds its threshold the neuron fires again.

The neurons’ accumulation facilitates another characteristic of living neural systems. Although they are essentially binary, transmitting +1 pulses, the pulses convey information about “how much”. Whether a sensor detects touch or sound or light, the information conveyed to the neural system by sensor outputs is both “what” and “how much”.

“What” depends on the type sensor. “How much” is the rate of “firings”. The greater the excitation inputs the greater is the rate of neuron “firing”.

Neural majority logic plus adjusted thresholds leads to representing the related universal. With accumulation the system also responds to “amount of signal” dealing with time as well as structure.

The operation of neural systems depends on two aspects: the interconnections between neurons and the thresholds. The interconnections are fixed; but, how should they be for optimum performance?

Initially, we have no idea. Nature had no idea, initially, either. The best choice is randomness.

A neural system must deal with a great variety of inputs. It is not possible to design in advance for all of the possibilities. The only way for a system to optimize its dealing with the unknown is random interconnections, which avoid a bias in any direction.

Then Darwinian variation and natural selection enter. Some “random” interconnection systems turn out to perform better than others. The process tends over time to select optimal interconnection systems.

But "optimal" depends on the situation. Vision systems have existed hundreds of millions of years, sufficient time for optimal sets of retinal universal processing interconnections to develop.

But, what is optimal brain operation for astronauts, steelworkers, gourmet chefs? Man experiencing so many different geographies, weathers, etc., confronts an unpredictable thinking need. The most likely success is one that is not biased in any direction.

At high levels of neural systems randomness is still the optimum design even though specific subsystems, can be optimized in evolved ways.

Our brains’ 10^8 neurons can express [2]^{10^8} universals, an immense number. Individual neurons cannot participate in just one universal.

Each neuron’s output inputs to multiple universals and their particular Boolean logic expressions, all simultaneously. Each universal is a different logic expression but its participating neurons participate in many other universals. While the same set of neurons might appear in a number of universals most universal’s neuron set is different from others.

The involvement of individual neurons in many universals simultaneously has a great advantage. It creates the capability for thinking.

5 Concepts, Thoughts, Thinking

The distinction between an input example and a universal is important.

A universal does not exist external to the neural system. It exists in the neural system as a configuration of neurons and their thresholds that can discriminate the presence or absence of that universal in an input example.

The input example exists in the world external to the neural system. It does not exist within the neural system except briefly as the “firing” of neurons representing the universals in which it participates.

A universal is relatively permanent in the neural system being the "wiring" configuration of neural interconnections and their thresholds. The input example is a momentary set of neural “firings”.

A concept is a set of universals that are related. A thought is a specific example of a concept. A concept is the neural logic structure of its universals. A thought is the firing, in that logic structure, of a subset of its neurons for an example that satisfies the universal. Their firing is the thought.

Thinking is sequences of thoughts, each thought a set of neuron “firings”, the set changing each “firing” and having in common parts of their universals’.

For example if thought #1 consists of universals [a, b, d, f, g], #2 consists of that plus [k], and #3 is #2 less [d], the three thoughts in that sequence are "a line of thought", thoughts sharing gradually shifting parts of their Boolean logical expression: thinking.
A thought is the “firing” of a specific set of neurons. Subsets of it are also parts of other related thoughts. At the moment of the current thought those other thoughts are not active because the current neuron “firings” are not exactly their required set.

The current thought could, with “help”, lead to activation of one or more other thoughts sharing a portion of logic with the current thought. That “help” would be something activating other related subsets of neurons. Because of the sharing of neurons between the two thoughts, the successive thoughts will tend to follow in terms of thinking.

Neurons cannot participate in contradictory multiple universals because a threshold could not form. In general the set of universals in which a neuron participates are a somewhat related family.

Subsets of the current thought’s logic participate in many other potential thoughts; thinking can proceed in several directions from the current thought.

With the extensive interconnection of neurons, and the recirculation of output firings as inputs elsewhere, and with constant sensor input, “help’s” to progress to a next thought are always present. A current thought inevitably results in an immediately following next thought ad infinitum. It’s direction depends on which “help” is stronger, dominant.

The transitions from thought to thought are changes of some of the individual universals that comprise the current thought. The process is thinking but not yet purposive thinking. It takes place in neural systems over a wide range of complexity: man, dogs, birds, snakes and beetles.

Each thought modifies the system in an iterative process of evolving universals and sequences of specific examples (thoughts) where the examples modify the universals and the universals determine the various directions the sequence takes.

7 Motivation

A sensory neuron’s “firing” rate tells: how loud a sound, how bright a light, the magnitude of a touch. A “too much” signal can mean danger. In the simple neural networks of early organisms some responded to “too much” inputs by action to avoid the cause.

The most simple early neural system was a sensor neuron connected directly to a motor one. It could, for example, produce withdrawal of an exposed body part or waving of flagella (to swim away).

That response increased the survival rate of the organisms. They became the dominant survivors into their future, their successors inheriting the response. Such avoidance of danger became an operating principal of early neural networks.

We retain those early-developed mechanisms. One can touch a cool oven but if it is hot the moment it is touched the hand is automatically withdrawn. The “too much” signals from the hand trigger the motion by motor neurons at interconnections in the spine.

We exhibit a more sophisticated neural response to “too much”. The eye shuts and the head averts if a threatening object is detected. The reaction is automatic before we are conscious of the problem. That is not a direct sensor-motor action. Significant neural processing is needed to convert the raw visual picture into “an object is moving rapidly on a threatening path” and to generate averting the head.

Eyes developed long after “early, simple neural networks”. But, their long established behavior, treating excess as danger and automatically acting to correct the situation, appears developed into a sophisticated version in the response to eye danger. A complex set of universals, a significant number of neuronal sets, processing “too much” signals collectively are involved in that action.

With evolution as neural networks became larger and more complex the operation of their “too much” response became more sophisticated, that is:

- It involved more neurons in more numerous universals which described more complex thoughts.
- It developed dealing with multiple “too much” signals to organize and prioritize multiple responses.
- It included logic to decide if a response is really needed and to consider alternative responses.

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 (hot oven), to the major neural processing type (eye shutting and head averting) to increasingly sophisticated motivations and resulting actions.

As our thoughts are the patterns of which neurons are firing at a particular moment, so our conscious purposive behavior, our actions in life at home, on the job, as parents, in love are our responses to highly sophisticated and complex sets of neural “too much” (and “not-ed” “too little”) signals.

Those signals involve, are related to, are the equivalent of, are pain and pleasure (pleasure being 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, the consequent responses normally involve non-material actions, involving intentions and desires.

Early in the evolution of neural systems they evolved to treat extremes of neuron firing rates, low or high, as being something to be avoided – triggers of corrective action. At a sophisticated level we now refer to the effect of excessively non-moderate sensor neural firing rates as meaning that the related material objects are painful or unpleasant.

At a high sophisticated level we now refer to the effect of excessively non-moderate non-sensor neural firing rates as meaning that the related mental concepts are unintended or undesired, fail to agree with a goal.

The opposite, moderate neuron firing rates that are neither too great nor too small signify: comfortable, intended, desired, goal achieved.

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

Or, perhaps, the greatest joy, the best sensation, corresponds to neural firing rates that are as near to “too much” as possible without mandating corrective action. Our experience would tend to indicate that we behave that way, that we crave excitement so long as it does not go over the boundary into the damaging.

Or, perhaps, for different kinds of good or pleasure 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 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.

9 Response to Motivation

Of equal importance to the input aspect of the evolved “too much” signal is the neural system’s action when such signals appear. In early neural systems that was a motor action tending to relieve the cause of the “too much” signal.

A highly evolved system’s response to “too much” signals also tends to relieve the cause of the signals. No matter how highly evolved 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 the situation, to remove or reduce the cause of the “too much” signal.

Thus we respond to desire for a cookie by exciting our motor neurons to cause our walking to the cupboard, selecting a cookie, and eating. Even more, in general responses to “too much” signals produce our performance of the routine of living: arising, eating, performing the day’s tasks, etc.

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

- a luscious desert;
- or 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 declining revenge to relieve the pain that it is contrary to one’s standard of moral character.

That can be done by a system as complex as follows. Our brain has 100,000,000,000 neurons. Assigning 10% to sensor and motor activities and providing for 100 subsystems, any one sub-system has 10^8 neurons and could represent 2^{100,000,000} different patterns.

A page can hold about 3,000 zeros. It takes 10,000 pages of zeros just to write out 2^{100,000,000} – to write it down not to express its value. (Writing "1000" takes 4 digits, but its value is 1,000.)

10 Memory

A memory is a sequence of thoughts, of neural firings, repeating a prior thought sequence. To the extent that it is repeated its thresholds become more set; the memory becomes more permanent.

The remembering of it, is via the associations of thought universals as in the original thinking. For access we must think of a thing associated with it that will trigger the sequence of thoughts.

Memories reside distributed over a large number of neurons, not in some separate “file cabinet” of the brain. The only difference between a remembering and a thinking is whether the pattern of thoughts is new or is retraced; however, at retrace some of the original thresholds have become changed.
11 Conclusion

As compared to discussions of the brain and its mental processes that focus on the neural structure of the brain and its biological functioning, the present analysis has addressed the "software" of the mind, the "software" that operates on the brain's biological neural structure. It is how the neural components logically interact that produces the higher level results that we experience in our minds. The development began with universals and then proceeded through perception, learning, and the processing of universals to ultimately the concepts and operations that constitute thoughts and thinking, behavior and memory.

Overall that is an immense topic to treat and to attempt to comprehend; however, an initial step, an opening of the subject, has been accomplished.

References: