April 17, 1964

ON THE MOLECULAR BASIS OF MEMORY

by Leo Szilard

The Salk Institute for Biological Studies
La Jolla, California

Introduction

The subject matter of this paper is a hypothetical molecular process on which the capability of the Central Nervous System to record and to recall an experience might conceivably be based. It may be open to doubt whether one knows enough about the biological processes which can take place in living cells to be able to say anything with reasonable assurance about the molecular processes that the brain employs. Still, with luck, one might perhaps guess the general nature of these processes. To what extent we may have succeeded in doing so, remains to be seen.

In another paper we shall discuss to what extent a simple neural network could account for Pavlov's experiments on the conditioning of the salivary reflex of the dog,[1] on the basis of the molecular processes here postulated. That discussion will cover the establishment as well as extinction of the conditioned response, and also the phenomena which Pavlov had designated as conditioned inhibition, differentiation, and negative induction.

Congenitally Determined Neurons and Memory Neurons

We assume that the neurons in the Central Nervous System which respond preferentially to different stimuli differ from each other in their "chemical specificity". We divide the neurons of the Central Nervous System into two broad classes. The class of "congenitally determined" neurons and the class of "memory" neurons.

(1) I.P. Pavlov, Conditioned Reflexes, Oxford University Press, 1927.

and the efficacy of the synapse bridging the two neurons ~~will be~~ is reduced by the same factor. Accordingly, such a "transprinted" neuron E may be caused to fire in spite of ~~the~~ receiving inhibitory signals ~~which it may receive~~ from the neuron $\bar{E}$*.

The transprintable neurons E get derepressed if the inhibitory neuron $\bar{E}$* is inhibited as a result of signals sent out by a neural network which we ~~shall~~ designate as the "Derepressor." ~~It~~ This will be known as the Derepressor sends out signals which are sufficiently strong ~~those will~~ to excite the inhibitory inter-neuron E**, which in turn will inhibit the inhibitory neuron $\bar{E}$*, ~~and will thereby relieve the repression of the transprintable neurons E.~~

The Derepressor network may receive an input signal from the neuron F and it may also receive an input signal from neurons E, via the inter-neuron $\bar{FI}$. These two input signals inhibit each other ~~, however,~~ within the Derepressor, however, and they cancel out if the intensity of both input signals is about the same. Accordingly, the Derepressor will send out strong signals only if the intensities of these two input signals differ from each other substantially.(1) In ~~our~~ a second paper we shall describe a simple "neural" network which ~~behaves~~ would function in this fashion.

---

(1) Footnote: The function we are attributing ~~to~~ for the Derepressor network is somewhat similar to that which has been ascribed to certain neurons, located in the superior olive, that receive signals from both ears. It seems to be the peculiarity of these neurons that they send out a signal if they receive a signal from one ear alone, whereas contemporaneous signals from both ears cancel out. It is believed that one may detect the directionality of sound by such a cell network and may thus be able to pick out what one hears in one ear, despite the large common background of noise in both ears.

As may be seen later, the Deprepressor network ~~will~~ may be expected to send out strong signals if food is introduced into the mouth of an unconditioned dog and ~~it~~ to ~~will~~ send out strong signals also ~~on an occasion~~ when a dog, whose salivary reflex has been fully conditioned,

We assume that the same holds also ~~—mutatis mutandis—~~ for the synapses of inhibitory neurons, except that in this case the "transmitter substance" which diffuses across the synaptic gap into the post-synaptic neuron lowers, rather than raises, the level of excitation of the post-synaptic, excitatory or inhibitory, neuron.

~~Importance of the "threshold".~~

In our models of neural networks one neuron may receive simultaneously input signals from a number of different excitatory neurons, which contact the neuron through one or several synapses each. Were we to assume that the intensity of the signal sent out by a post-synaptic neuron is a linear function of the intensities of the input signals, then ~~our model would be unable~~ to account for the ability of the dog - demonstrated by Pavlov - to learn to discriminate (differentiate) in the conditioned salivary reflex between the compound stimulus ~~component~~ and its components. (This point will be discussed in our second paper.)

~~In the meantime~~, we shall ~~in any case~~ assume ~~hereafter~~ that the intensity of the output signal of a post-synaptic neuron is an S-shaped function of the sum, extended over all synapses, of the "excitatory inputs" of the individual synapses. If we designate this sum as the "~~integrated~~ excitatory input", then we may say that the intensity of the output signal of a post-synaptic neuron is an S-shaped function of the "~~integrated~~ excitory input". ~~then we may say that the intensity of the output signal of the post-synaptic neuron is an S-shaped function of the "integrated excitatory input".~~ ~~We may also say that~~ roughly ~~speaking~~, the post-synaptic neuron has ~~something like~~ a "threshold" which the "~~integrated~~ excitatory input" must exceed in order to evoke an output signal of substantial intensity.

The Transprinting of Neurons

We divide neurons of the central nervous system into two broad classes: the "congenitally determined" neurons and the "memory" neurons. The neurons which attain their full chemical specificity of their cell membrane during the development of the

We assume that the same holds also - mutatis mutandis - for the synapses of inhibitory neurons, except that in this case the "transmitter substance" which diffuses across the synaptic gap into the post-synaptic neuron lowers, rather than raises, the level of excitation of the post-synaptic, excitatory or inhibitory, neuron.

In our models of neural netwroks one ~~excitatory~~ neuron may receive simultaneoulsy input signals from a number of different excitatory neurons, which contact the neuron through one or several synapses each. Were we to assume that the intensity of the signal sent out by a post-synaptic neuron is a linear function of the intensities of the input signals, then our model would be unable to account for the ability of the dog- demonstrated by Pavlov - to learn to discriminate (differentiate) in the conditioned salivary reflex between the compound stimulus, ~~that has say an auditory and visual~~ component, ~~on the one hand, and auditory and visual~~ components ~~on the other hand~~. ~~This~~ compound ~~stimulus~~ will be discussed in our second paper.

In the meantime, we shall in any case assume hereafter that the output signal ~~intensity~~ of a post-synaptic neuron is an S-shaped function ~~with the summation~~ of the sum, extended over all synapses, of the excitatory inputs of the individual synapses. If we designate this sum as the "integrated ~~excitatory~~ input", then we may say that the intensity of the output signal of ~~the~~ post-synaptic neuron is an S-shaped function of the "integrated excitory input", ~~then we may say that the intensity of the output signal of the post-synaptic neuron is an S-shaped function of the integrated excitatory input.~~ ~~Speaking roughly, we may, accordingly,~~ say that the post-synaptic neuron has something like a threshold which the integrated input must exceed in order to evoke an output signal of substanital intensity.

We assume that there is a class of "congenitally-determined" neurons which are capable of participating in the transprinting of a memory neuron and that if a "congenitally-determined" neuron of this class fires, then those parts of its cell membrane (covering the boutons of the branch fibres of its axon), which form the active zones of the pre-synaptic membranes become permeable for ~~its~~ the specific membrane proteins. We also assume that when a memory neuron fires, then those parts of the cell membrane, (covering its cell body and its dendrites) which constitute the active zones of the post-synaptic membranes, become permeable for the specific membrane proteins. Accordingly, if a "congenitally-determined" neuron of this class contacts a memory neuron through a synapse and if both neurons fire "simultaneously" so that for a period of time both the pre-synaptic and the post-synaptic membrane is permeable for the specific membrane proteins, ~~of the pre-synaptic memory neuron will diffuse through the pre-synaptic and the post-synaptic membrane into the~~ then the specific membrane proteins of the pre-synaptic congenitally-determined neuron will diffuse through the pre-synaptic and the post-synaptic membrane into the post-synaptic memory neuron. We postulate that if a specific membrane protein penetrates in this fashion into a memory neuron it induces in the memory neuron the complementary specific membrane protein -- just as an antigen induces its antibody, if it penetrates into certain lymphatic cells of the rabbit.

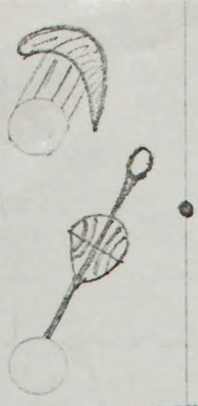
Insert on P. 6.

Like congenitally-determined neurons, transprinted neurons may also particpate in the transprinting of a transprintable neuron.

Insert on P. 6.

Like congenitally-determined neurons, transprinted neurons may also particpate in the transprinting of a transprintable neuron.

We assume that there is a class of "congenitally-determined" neurons which are capable of participating in the transprinting of a memory neuron and that if a "congenitally-determined" neuron of this class fires, then those parts of its cell membrane (covering the boutons of the branch fibres of its axon), which form the active zones of the pre-synaptic membranes become permeable for the specific membrane proteins. ~~We also~~ Similarly we assume that when a memory neuron fires, then those parts of the cell membrane, (covering its cell body and its dendrites) which constitute the active zones of the post-synaptic membranes, become permeable for the specific membrane proteins. Accordingly, if a "congenitally-determined" neuron of this class contacts a memory neuron through a synapse and if both neurons fire "simultaneously", so that for a period of time both the pre-synaptic and the post-synaptic membrane is permeable for the specific membrane proteins, then the specific membrane proteins of the pre-synaptic "congenitally-determined" neuron will diffuse through the pre-synaptic and the post-synaptic membrane into the post-synaptic memory neuron. We postulate that if a specific membrane protein penetrates in this fashion into a memory neuron it induces in the memory neuron the complementary specific membrane protein --just as an antigen induces its antibody, if it penetrates into certain lymphatic cells of the rabbit.



Emmanuel

Stanford ; Parzen
" Chernoff } Statistics
Hermann
X

John Tukey ; Princeton
Apr. 19

~~[illegible]~~

{ Bahadur
{ Billingsley

Jimmy Savage
Center for Adv. Studies of
Palo-Alto
321-2052
(415)

Emmanuel

Stanford ; Parzen
" Chernoff } Statistics
Hermann
X

John Tukey ; Princeton
Apr. 19

Bahadur
Billingsley

Jimmy Savage
Center for adv. Studies of
Palo-Alto
321-2052
(415)

Reference to Dr. Brunovski's work in this section

p. 4

JVS
Aug 18, 1964

~~The Efficacy of the Synapse~~

We assume that in the Central Nervous System the neurons which respond preferentially to different stimuli differ from each other in their chemical specificity. We may divide the neurons in the Central Nervous System into two broad classes; the class of the congenitally determined neurons and the class of the memory neurons. The congenitally determined neurons attain their full chemical specificity through differentiation which they undergo during the development of the individual, either during embryonal life or at the latest during the early postnatal period. The memory neurons may attain their full chemical specificity anytime during the life of the adult, through a process that may be regarded as a kind of differentiation, induced by congenitally determined neurons on occasions when they are activated by some sensory or mental stimulus. We shall refer to this process of mentally-induced differentiation as transprinting. Accordingly, we will refer to these memory neurons before they attain their full chemical specificity as transprintable neurons and ~~after they have attained their full chemical specificity~~ as transprinted neurons.

We may as well state at this point just exactly what we mean when we use the term chemical specificity.

In the fruit fly the amount of DNA per somatic cell is just about enough to account for 10,000 genes.

In mammals the amount of DNA per somatic cell is about 100 times larger. ~~We assume that~~ each gene is capable of producing a specific protein molecule and that as a result of the differentiation which a somatic cell which the Central Nervous System undergoes during development ~~by the end of the early postnatal period~~, the neurons ~~in the Central Nervous System~~ which differ from each other in their preferential response to different stimuli

20

In the fruit fly the amount of DNA per somatic cell is just about enough to account for 10,000 genes. In mammals the amount of DNA per somatic cell is about 100 times larger.

We assume that somatic cells differ from each other because a different set of the genes they contain produces their specific protein molecules at a high rate. We assume in particular that, as a result of the differentiation, which the neurons of the Central Nervous System undergo during development, those congenitally determined neurons which sensory signals, contain a different set of genes which produce their specific protein molecules at a high rate, i.e. in those neurons, a different set of neuro-specific proteins will be maintained at a high level of concentration.

If we were to assume, as we could for the moment, without losing too much flexibility at this point, that the number of neuro-specific proteins which are elevated is the same in all the congenitally determined neurons and if we designate this number with n, then the number of possible congenitally determined neurons which will differ from each other in their chemical specificity would be given by the binomial coefficient $B = \binom{N}{n}$, where N designates the number of neuro-specific genes.

For n = 30 and $N = 10^4$, the binomial would amount to

This number would far exceed the total number of neurons in the human brain, which is generally estimated to be about 10^{10}. This overlap number disappears, however, when we consider that many of these would have a number of elevated neuro-specific proteins in common.

If we were to assume, as we could without losing too much flexibility at this point, that the number of elevated neuro-specific proteins were the same in all congenitally determined neurons, the number of different congenitally-determined neurons would be given by the binomial coefficient $\binom{N}{n}$ where N designates the number of neuro-specific genes and n designates the number of neuro-specific proteins which are elevated in the congenitally determined neuron.

For N = 1,000 and n = 30, the binomial would amount to about which would far exceed the total number of neurons in the human brain which is generally estimated to be about 10^{10}.

We shall designate as the overlap number m of two neurons the number of elevated neuro-specific proteins which are contained in the set of elevated neuro-specific proteins of both ~~neurons~~ of them.

As we shall see below, according to the notions here adopted, the signals coming from two congenitally determined neurons can be confused by the Central Nervous System if their overlap number is a substantial fraction of the number of neuro-specific proteins which are elevated in the congenitally determined neurons.

The number of possible congenitally determined neurons ~~given by the binomial quoted above~~ would be drastically below the value given by the ~~drastically~~ cut down if we ~~limit ourselves to demand neurons so~~ that no two neurons shall have an overlap number m which exceeds a ~~certain small fraction, perhaps 10%~~ of the number n.

a certain small number m_0, i.e. a number which is a small fraction, say 10% of n.

As we shall see below, according to the notions here adopted, the signals coming from two congenitally determined neurons can be ~~fairly easily~~ confused if the number of the neuro-specific proteins which their sets of elevated neuro-specific proteins have in common is substantial. Let us then define the number of neuro-specific proteins which the sets of elevated neuro-specific proteins of ~~the~~ two neurons have in common, as the overlap number m, of the two neurons. We shall assume that the overlap number, m, of two congenitally determined neurons which respond to two different sensory signals between which the individual can easily discriminate does not exceed a certain ~~percentage, perhaps 10% of n, so that~~ thus, for instance in the case of n = 30, ~~we~~ . ~~If we designate~~ two congenitally determined neurons for which the overlap number does not exceed some such limit as being substantially different, then we can ask how many substantially different neurons are possible for ~~any~~ N, n and m.

The mathematical problem which is involved may be formulated as follows:

~~Should~~ there be n different objects, how many different sets composed of n objects are there, if no two sets must have an overlap number of m or larger than m?

~~A solution of~~ this problem, which does not seem ~~capable of a~~ trivial ~~solution,~~ has been ~~solved~~ by Dr. Bruno Bronowski, of this Institute, and will be reported elsewhere.

The overlap of (activated) genes in cells

We have a set of N genes (say $10^3 \le N \le 10^4$), of which I will call one particular gene g.

We consider cells in which n of these genes have been activated; say, $n \ge 30$.

We ask what is the number $\{N, n, k\}$ of the largest set of such cells such that no two cells have more than k activated genes in common; say, $k = 3$.

I divide this set of cells into two subsets: the subset of cells which contain (actively) the particular gene g, whose number is $\{N-1, n-1, k-1\}$; & the subset of cells which do not contain (actively) the gene g, which numbers $\{N-1, n, k\}$. Therefore

$$\{N, n, k\} = \{N-1, n-1, k-1\} + \{N-1, n, k\}$$

that is

$$\boxed{\{N, n, k\} - \{N-1, n, k\} = \{N-1, n-1, k-1\}.} \quad (1)$$

I now use the basic recurrence formula (1), to climb from $k = 0$ to any k — in particular, to $k = 3$.

First, then, $\{N, n, 0\}$ is the largest integer in $\frac{N}{n}$, obviously; that is,

$$\{N, n, 0\} = \left[\frac{N}{n}\right] \simeq \frac{N}{n} \qquad \text{——— (2)}$$

Next, for $k=1$, from (1)

$$\{N, n, 1\} - \{N-1, n, 1\} = \{N-1, n-1, 0\} \simeq \frac{N-1}{n-1}$$
$$\{N-1, n, 1\} - \{N-2, n, 1\} = \{N-2, n-1, 0\} \simeq \frac{N-2}{n-1}$$
$$\cdots\cdots\cdots\cdots\cdots\cdots$$
$$\{2n, n, 1\} - \{2n-1, n, 1\} = \{2n-1, n-1, 0\} \simeq \frac{2n-1}{n-1}$$

& $\{2n-1, n, 1\} = 1$

Whence, by addition,

$$\{N, n, 1\} = \sum_{i=2n-1}^{i=N-1} \{i, n-1, 0\} + 1 \simeq \frac{1}{n-1}\left(\overline{N-1} + \overline{N-2} + \cdots + 2\overline{n-1}\right) + 1$$

$$\simeq \frac{1}{2}\cdot\frac{1}{n-1}\cdot(N+2n-2)(N-2n)$$

$$\simeq O\left(\frac{N^2}{n}\right) \quad \text{if } N \gg n, \text{ as here} \qquad \text{——— (3)}$$

Next, for $k=2$, from (1), it follows by the same procedure that

$$\{N, n, 2\} = \sum_{i=2n-2}^{i=N-1} \{i, n-1, 1\} + 1$$

$$\simeq \frac{1}{2}(N-2n+1)\, O\left(\frac{N^2}{n}\right)$$

$$\simeq O\left(\frac{N^3}{n}\right) \quad \text{if } N \gg n, \text{ as here} \qquad \text{——— (4)}$$

Finally, by the same procedure, it follows from the basic recurrence formula (1) that for $k=3$,

$$\left\{ N, n, 3 \right\} = \sum_{i=2n-3}^{i=N-1} \left\{ i, n-1, 2 \right\} + 1$$

$$\simeq \frac{1}{2}(N-2n+2)\, O\left(\frac{N}{n}^3\right)$$

$$\simeq O\left(\frac{N^4}{n}\right) \quad \text{if } N \gg n, \text{ as here} \qquad (5)$$

The general formula is

$$\boxed{\left\{ N, n, k \right\} = \sum_{i=2n-k-1}^{i=N-1} \left\{ i, n-1, k-1 \right\}.} \qquad (6)$$

The estimate (3) must be roughly right; the estimates (4) & (5) need to be checked more carefully — I have sketched the method without the detail, writing in haste as I do.

The merit of (1)&(6) & the method is, of course, that they allow exact calculation for different N, n & k very simply. Note that the order of N is important; the exact value of n seems unimportant (apart from its order); & the precise value of k is far more important than the ratio n/k.

J Bronowski 9 April 64

We may designate as the efficacy of a synapse the magnitude in the rise of acetylcholine concentration which results from a given rise in signal intensity. The efficacy of the synapse thus defined is then inversely proportional to the "cholinesterase" concentration prevailing at the post-synaptic membrane and this means that the efficacy of the synapse is proportional to the overlap fraction, δ, of the pre-synaptic and post-synaptic neurons.

The Transprinting of Neurons

According to the notions here adopted, an adult can remember and learn because there are present in his Central Nervous System neurons which have not attained their full chemical specificity as yet. and these "memory" neurons can attain their full chemical specificity during his lifetime, through the process of "transprinting". Thus, transprinting may be regarded as a certain kind of belated differentiation through which a memory neuron attains its full chemical specificity.

We know very little about how the differentiation of the "congenitally determined" neurons comes about during development.

For all we know, differentiation might have something in common with enzyme induction in bacteria. In both cases a gene which is potentially capable of producing an enzyme, either does not produce that enzyme, or produces it at a very low rate --until something happens that causes the formation of the enzyme. Thus, we know for instance, that in bacteria the level of the enzyme β-galactosidase can be raised about a thousandfold by adding lactose to the growth medium and there is reason to think that lactose is the natural inducer of this enzyme.

However, the changed rate of enzyme production will persist in the growing bacterial culture, only as long as the inducer remains in the growth medium. As soon as the inducer is removed the rate of enzyme production reverts to normal, i.e. in a bacterial culture, the bacteria do not remember for long having been exposed to the inducer.

In contrast to this when ~~so much~~ a somatic cell undergoes differentiation during embryonic development, there appear in the cell a number of specific proteins raised to a high level of concentration and ~~therefore~~ thereafter the cell, as well as its decendants, will contain these specific proteins at a high concentration.

In order to account for this phonomenon of persistence - which is absent in enzyme induction in bacteria - one may assume that once a concentration of such a protein molecule is raised in a somatic cell ~~so much acceleration~~ above the threshold, thereafter the rate of production of this protein molecule remains high and its concentration is maintained at a high level. This implies the operation of some sort of a locking mechanism. It implies that a specific protein molecule must be able to maintain its concentration at a high level, acting either directly or through the intermediary of certain key compounds, where each key compound would have to be specific for a ~~key~~ specific transmission specific protein. H. S. Anker suggested in 1960 that this kind of locking mechanism might be the biochemical basis of memory .(.. H. S. Anker, NATURE, 188 p.938 1960 ...). We do not know what these key compounds might be in the case of the transmission specific proteins of the congenitally determined neurons, and for all we know there might be the transmission of specific proteins themselves. ~~But~~ However the only thing we need to assume about these key compounds here, is that in a congentially determined neuron in which certain transmission-specific proteins are maintained at a high level of concentration, the corresponding key compounds are also maintained at a high level of concentration.

According to the notions here adopted, the "unit of recorded memory" which may be ~~recorded~~ created within one neuron is not a bit but something that contains considerably more information than a bit. There is a class of dreams, first described by Freud and discussed by him in great detail, which is centered on what he calls a "Traumgedanke". In a dream of this particular type the same Traumgedanke appears in a number of different representations, and if ~~within one night~~ within one night, several dreams of this type follow each other, the same Traumgedanke is likely to be represented in every one of them.

It is conceivable that a Freudian dream is generated because one single neuron gets excited during the night and remains excited for an extended period of time. If the individual wakes up, ~~becomes conscious~~ and the dream enters his consciousness, the neuron involved may cease to be excited.

We consider a Traumgedanke to be a Gestalt in terms of the concept ~~initially~~ introduced by Wertheimer and Koehler and, ~~according to our notions~~, it is a most complex Gestalt that can ~~still~~ be stored in a single neuron. Our model cannot account for a greater capacity of storage than one Gestalt per neuron and one may ask whether this limitation might not get us into trouble from the start.

If an individual were able to retain information of the complexity of a Traumgedanke conveyed to him every few seconds, 24 hours a day, over a period of one hundred years, then the units of recorded memory ~~stored by such an individual~~ would ~~irreversably~~ tie down just about 10^9 neurons of ~~his~~ brain. This would still be ~~10~~ times less than the number of neurons in the brain of man, which is usually quoted to be 10^{10}.

April 1st Version

ON THE MOLECULAR BASIS OF LONG-TERM MEMORY

by Leo Szilard

The Salk Institute for Biological Studies
La Jolla, California

Introduction

The subject matter of this paper is a molecular process on which the capability of the Central Nervous System to record and to recall an experience might be based. The chances are that we do not know enough about the biological processes which may take place in the living cell to be able to guess what molecular process the brain may employ. But even if we were to fail to present a model for the process that the brain employs, we might, with luck, still succeed in presenting a model that would be capable of accounting for a highly efficient system of recording and recalling an experience, on the basis of a mechanism which living cells might conceivably employ.

Such a model may be regarded as "efficient" if the number of neurons which a "unit of recorded memory" would tie down, is comparatively small and if the neural networks required for the recall of an experience in the right circumstances, would be comparatively simple.

I propose to present here a hypothetical molecular process and I believe that the model which it represents is "efficient" in the meaning of the term defined above. If it should turn out that the neural networks that one would need to postulate in order to account for the higher mental functions of man, then --even if otherwise worse came to worst, the final verdict would still be "se non è vero è ben trovato" and this is more than could be said for many of the other models which might be concocted.