Reprinted from the PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES
Vol. 51, No. 6, pp. 1092–1099. June, 1964.

*ON MEMORY AND RECALL**

BY LEO SZILARD

THE SALK INSTITUTE FOR BIOLOGICAL STUDIES, LA JOLLA, CALIFORNIA

Communicated April 27, 1964

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

The Efficacy of a Synapse Bridging Two Neurons.—Our neural network models involve excitatory neurons and inhibitory neurons (of the kind which exert a postsynaptic inhibitory effect).

Let us consider an excitatory neuron which contacts through a synapse another neuron. If such an excitatory neuron sends a volley of nerve impulses to this synapse, then a certain quantity of an excitatory "transmitter substance" is released in the vicinity of the presynaptic membrane which diffuses across a gap—the synaptic cleft—into the postsynaptic neuron and raises the level of excitation of that neuron by a certain amount. We shall designate this excitatory transmitter substance as "acetylcholine" (in quotes). The "acetylcholine" which diffuses into the postsynaptic neuron is destroyed, in the vicinity of the postsynaptic membrane, by an enzyme which we shall designate as "choline esterase."

The rate at which "acetylcholine" is released in the vicinity of the presynaptic membrane is a function of the frequency of the nerve impulses which reach the synapse, and we shall designate this rate as the "signal intensity." For the sake of simplicity, we shall assume that the signal intensity is for all synapses the same function of the frequency of the nerve impulses which are fed into the synapse.

The rate at which "acetylcholine" is destroyed in the postsynaptic neuron is proportional to the product of the concentration of "acetylcholine" and the concentration of the enzyme "choline esterase" in the vicinity of the postsynaptic

membrane. Therefore, if, at a given point in time, nerve impulses of a certain frequency begin to arrive at a synapse, the "acetylcholine" concentration will begin to rise and will asymptotically approach a limit in the vicinity of the postsynaptic membrane, which is proportional to the signal intensity and inversely proportional to the concentration of "choline esterase," prevailing in the vicinity of the postsynaptic membrane. The "acetylcholine" concentration which is asymptotically approached at the postsynaptic membrane constitutes the "excitatory input," which is received from the synapse by the postsynaptic neuron. On this basis we may then say that, for any given signal intensity, the excitatory input received from a given synapse by the postsynaptic neuron is inversely proportional to the "choline esterase" concentration prevailing in the vicinity of the postsynaptic membrane of that synapse.

We assume that the enzyme "choline esterase" is inactivated at the postsynaptic membrane in different synapses at different rates and that this rate of inactivation is determined by the chemical specificities of the two neurons which are bridged by the synapse. We shall assume, for the sake of simplicity, however, *that the enzyme "choline esterase" is produced at the same rate in all excitatory neurons.*

We designate as the efficacy of the synapse the excitatory input which a postsynaptic neuron receives from that synapse, *per unit of signal intensity.* On the basis of the above assumptions we may then say that the efficacy of the synapse is proportional to the rate at which "choline esterase" is inactivated at the postsynaptic membrane which in turn is determined by the chemical specificities of the two neurons which are bridged by the synapse.

The Rate of Inactivation of "Choline Esterase."—We assume that neurons which differ from each other in their response-specificity contain a different set of certain specific proteins in their cell membrane. We shall refer hereafter to these proteins as the "specific membrane proteins."

We postulate that to each specific membrane protein, there exists a complementary specific membrane protein and that a specific membrane protein molecule can combine with its complementary counterpart, just as an antibody molecule can combine with a molecule of its antigen. Accordingly, two complementary specific membrane proteins may behave as if they were, so to speak, each other's antibodies, as well as each other's antigens.

When an antibody molecule combines with an antigen molecule, it undergoes an allosteric transition, and an antibody molecule, when it is thus "dimerized," can bind complement. We assume that quite similarly a molecule of a specific membrane protein, when it combines with its complementary counterpart, undergoes an allosteric transition, and, when it is thus "dimerized," it can bind—and not only bind but also inactivate—the enzyme "choline esterase."

The gap (synaptic cleft) between the presynaptic membrane and the postsynaptic membrane is estimated to be about 200 Å wide. According to the notions here adopted, there must be, however, a number of places within the active zones of the two synaptic membranes at which this gap is narrowed down, so that the presynaptic membrane and the postsynaptic membrane are in physical contact. We assume that at such a point of contact, a molecule of a specific membrane protein, located in the postsynaptic membrane, can "dimerize" across the synaptic gap, with its complementary counterpart, located in the presynaptic membrane.

Let us now consider two neurons, A and B, which are bridged by a synapse. Neuron A is characterized by a set (*a*) of specific membrane proteins, which are present in its cell membrane, and neuron B is characterized by another set, (*b*). *We shall designate as the "overlap number" of these two neurons the number of specific membrane proteins contained within the set* (*a*) *which have their complementary counterpart contained within the set* (*b*) [*or vice versa*].

From this overlap number we may compute the efficacy of the synapse which bridges these two neurons. In order to simplify this computation we shall assume that the area of the active zone of the synaptic membrane is the same for all synapses, and also assume that the concentration of each specific membrane protein in the cell membrane is the same for any given neuron. On the basis of these simplifying assumptions, we may then say that the number of "dimers" contained within the active zone of the membrane of a synapse, which bridges neuron A and neuron B, is given either by the ratio of the "overlap number" to the total number of specific membrane proteins of neuron A, or by the ratio of the "overlap number" to the total number of specific membrane proteins of neuron B—*whichever ratio is smaller.* We shall designate *the smaller* one of these two ratios as the "overlap fraction" of the neurons A and B. *Accordingly, we may then say that the efficacy of a synapse bridging two neurons is proportional to the overlap fraction of the two neurons. This is the first fundamental postulate of our model.*

We assume that the same holds true also for the synapses of our inhibitory neurons, except that in this case the "transmitter substance" which diffuses across the synaptic gap into the postsynaptic neuron lowers, rather than raises, the level of excitation of the postsynaptic neuron.

The Transprinting of Neurons.—We divide neurons of the central nervous system into two broad classes: the "congenitally determined" neurons and the "memory" neurons. We designate neurons which attain their full chemical specificity of their cell membrane during the development of the individual (mostly during embryonal life and at the latest during the early postnatal period) as "congenitally determined neurons." If all the neurons of the central nervous system were of this sort, then the individual would not be able to learn and his behavior would be wholly governed by the inborn reflexes. According to the notions here adopted, an adult can learn, and recall what he has learned, because his central nervous system contains memory neurons and each of these can, once in his lifetime, acquire an additional set of specific membrane proteins, through a process which we designate as "transprinting."

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 fibers of its axon), which form the active zones of the presynaptic membranes become permeable for the specific membrane proteins. 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 postsynaptic 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 presynaptic and the postsynaptic membrane is

permeable for the specific membrane proteins, then the specific membrane proteins of the presynaptic "congenitally determined" neuron will diffuse through the presynaptic and the postsynaptic membrane into the postsynaptic 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 a certain lymphatic cells of the rabbit.*[1] If several such presynaptic neurons fire simultaneously with the memory neuron, *then the memory neuron will on such an occasion acquire the sets of specific membrane proteins which are complementary to the sets of all of these presynaptic neurons.*[2] *This is the process of transprinting. Its occurrence as an "all or none" process constitutes our second fundamental postulate.*

We shall refer to memory neurons before they are transprinted as transprintable neurons and thereafter we shall refer to them as transprinted neurons. Like congenitally determined neurons, transprinted neurons may also participate in the transprinting of a transprintable neuron.

If a neuron participates in the transprinting of a transprintable neuron, then we may expect this neuron and the transprinted neuron to have a large overlap fraction and, accordingly, we may expect synapses bridging these two neurons to have a high efficacy.

The Conditioned Response.—In order to illustrate how transprinting may take place, we shall use as an example the classical (Pavolovian) conditioning of the salivary reflex of the dog.[3] We shall indicate on this occasion, however, only rather sketchily what takes place during conditioning.

When "food" is introduced into the mouth of a dog, the dog responds with salivation. This is the inborn, or unconditioned, response. Let us now expose the dog to a compound stimulus which has an auditory as well as a visual component, and let us—before the compound stimulus is turned off—place food into the mouth of the dog. If, after several such "conditioning exposures," the dog is then presented for the first time with the compound stimulus, unreinforced on this occasion by the introduction of food into its mouth, the dog may be expected to salivate. This is the conditioned response.

We assume that there is a neuron F in the central nervous system, characterized by the set (f), which preferentially responds to the stimulus of "food in the mouth." *Moreover, we shall assume in particular that the signal to which the neuron F responds is the onset of this stimulus.* As shown in Figure 1, the neuron F is connected through a synapse to an effector neuron, which innervates the salivary gland. This effector neuron is characterized by the set $(\bar{f})$, where $(\bar{f})$ denotes a set of specific membrane proteins which is complementary to the set (f). Because the overlap fraction of the neuron A and the effector neuron is *one*, the synapses which bridge these two neurons have a high efficacy. Therefore, placing food in the dog's mouth may be expected to cause the dog to salivate.

In order to account for the conditioned response, we postulate the existence of a number of groups of transprintable neurons E. Each of these groups may consist of several hundred neurons E, all of which have the following in common: The neuron F contacts through a synapse each of the neurons E and in turn each neuron E contacts, through a synapse, an interneuron $\bar{\text{F}}\text{I}$ [characterized by the set $(f) + (\bar{i})$] which in turn contacts, through a synapse, the effector neuron.

Until something happens which is "significant" from the point of view of the

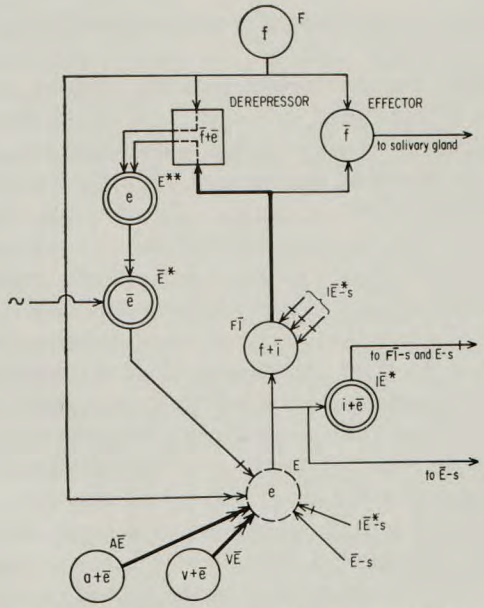


FIG. 1.—Excitatory neurons are represented by circles, and inhibitory neurons are represented by double circles. Excitatory synapses are represented by simple arrows, except if they belong to neurons which are capable of transprinting, in which case they are represented by double arrows. Inhibitory synapses are represented by arrows with a crossbar. The transprintable neuron $\bar{E}$ is represented by a dotted circle.

salivary reflex, all the transprintable neurons E are repressed, because they are inhibited by signals which are continuously being sent out by the inhibitory neuron $\bar{E}^*$ [characterized by the set $(\bar{e})$]. This inhibition is assumed to be strong enough to prevent a transprintable neuron E from firing, even if it should receive a substantial aggregate "excitatory input," because the overlap fraction of the inhibitory neuron $\bar{E}^*$ and of the transprintable neuron E is *one*. It should be noted, however, that after the neuron E is transprinted and acquires a set of specific membrane proteins which is composed of a large number of such proteins, then its overlap fraction with the inhibitory neuron $\bar{E}^*$ is reduced by a substantial factor, and the efficacy of the synapse bridging the two neurons is also reduced by the same factor. Accordingly, such a "transprinted" neuron E may be caused to fire in spite of receiving inhibitory signals from the neuron $\bar{E}^*$.

The transprintable neurons E get *derepressed* if the inhibitory neuron $\bar{E}^*$ is inhibited by signals emanating from a neural network designated as the "derepressor." This will happen if the derepressor sends out signals which are sufficiently strong to excite the inhibitory interneuron E**, which in turn will inhibit the inhibitory neuron $\bar{E}^*$. The derepressor network may receive an input signal from the neuron F and it may also receive an input signal, via the interneuron $\bar{F}I$, from neurons E. These two input signals counteract each other 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.* In our second

paper we shall describe a very simple "neural network" which would function in this fashion.

As will be seen later, the derepressor network may be expected to send out strong signals if food is introduced into the mouth of an unconditioned dog and to send out strong signals also when a dog, whose salivary reflex has been fully conditioned to a certain stimulus, is presented with that stimulus, without having, on this occasion, food placed into its mouth. The depressor network will not send out signals, however, if the fully conditioned dog is presented with the correct stimulus and food is placed into its mouth. Accordingly, no additional neurons E would be transprinted as the result of such "routine exposures."

It is probably generally true that a sensory experience is recorded only if there is "significance" attached to that experience. In our model of the conditioned salivary reflex there is significance attached to the stimulus "food in the mouth" for an unconditioned (or not fully conditioned) dog, while for a fully conditioned dog, there is significance attached to the conditioned stimulus, but only if that stimulus is *not* accompanied by the signal "food in the mouth."

We shall try to indicate next in what manner a conditioned salivary response may be established to a compound stimulus which has a visual and an auditory component. To this end we assume that in the central nervous system there is a neuron $A\bar{E}$ which responds preferentially to the auditory component of the compound stimulus, and another neuron $V\bar{E}$ which responds preferentially to the visual component. These two neurons are characterized by the sets $(a) + (\bar{e})$ and $(v) + (\bar{e})$, respectively. We assume that the number of different specific membrane proteins contained in the neurons $A\bar{E}$ and $V\bar{E}$, which we designate by $n(A\bar{E})$ and $n(V\bar{E})$, respectively, are large compared to the number of different specific membrane proteins contained in the transprintable neuron E, which we designate by $n(E)$. Accordingly, we have: $n(A\bar{E}) > n(E)$, and $n(V\bar{E}) > n(E)$.

We assume that, out of a group of several hundred neurons E, a certain fraction is contacted through a synapse by the neuron $A\bar{E}$, a certain fraction is contacted through a synapse by the neuron $V\bar{E}$, and a certain fraction is contacted by both the neuron A-$\bar{E}$ and the neuron $V\bar{E}$. Because the neuron $A\bar{E}$ as well as the neuron $V\bar{E}$ has an appreciable—even though small—overlap fraction with the transprintable neurons E, we may assume that if either of these two neurons fires, or if both of them fire, at a time when the transprintable neurons E are derepressed, then one or more transprintable neurons E will fire also and will on that occasion be transprinted by the neurons $A\bar{E}$ or $V\bar{E}$ or both. If at the same time the neuron F fires also, then these same neurons E will be transprinted also by the neuron F. The firing of neuron F alone would, however, not cause the neurons E to fire, even when the neurons E happen to be derepressed, because the neurons F and the transprintable neurons E have zero overlap.

If the unconditioned dog is for the first time exposed to the compound stimulus, and at the same time food is introduced into its mouth, the derepressor network will send out a strong signal and one, or several, of the transprintable neurons E will be caused to fire. These neurons E will then be transprinted, on this occasion, with the sets (f), (a), and (v). If this dog is exposed, for the second time, to the compound stimulus and at the same time food is introduced into its mouth, then the signal sent out by the derepressor network will be somewhat weaker than the

first time. This is so because the neurons E which have been transprinted at the time of the first conditioning exposure, and which will be excited at the time of the second exposure, have a large overlap fraction with the interneuron $F\bar{I}$ and will, therefore, send a signal to the derepressor network, which counteracts the signal received by this network from the neuron F. As the conditioning process is continued and the dog is repeatedly subjected to such "conditioning" exposures, the neurons E which are transprinted with the sets $(\bar{f})$, $(\bar{a})$, and $(\bar{v})$ will increase in number. Finally, the derepressor network will no longer send out a signal when the dog is exposed to the compound stimulus and at the same time "food" is introduced into its mouth. Such a dog is then fully conditioned and continuing the conditioning exposures would not make the conditioning any deeper.

Let us now expose such a fully conditioned dog to the compound stimulus, unreinforced on this occasion by the introduction of food into its mouth. The neurons E which have been transprinted during the previous conditioning exposures with the set $(\bar{f})$ as well as the sets $(\bar{a})$ and $(\bar{v})$ will now be caused to fire. Because of the substantial overlap of the transprinted neurons E, which contain the set $(\bar{f})$, with the interneuron $F\bar{I}$, the firing of the neurons E will lead to the firing of the interneuron $F\bar{I}$ and this in turn will lead to the firing of the effector neuron. Accordingly, on the occasion of this unreinforced exposure of the dog to the compound stimulus, the dog will salivate. This is the conditioned response.

Incidentally, on this occasion, when the interneuron $F\bar{I}$ fires, it will cause the derepressor network to send out a strong signal because this network does not on this occasion receive a signal from the neuron F. Accordingly, on this occasion, one or more neurons E will get transprinted with the sets $(\bar{a})$ or $(\bar{v})$ or both, but none of them will be transprinted with the set $(\bar{f})$. Therefore, if the dog is repeatedly exposed to the compound stimulus in such a fashion, i.e., without reinforcement, then the number of neurons E which are transprinted with the sets $(\bar{a})$ or $(\bar{v})$ or both, but not with the set $(\bar{f})$, will increase on each such occasion. The overlap fraction of these transprinted neurons E with the interneurons $F\bar{I}$ is zero, and therefore the excitation of these transprinted neurons E would not contribute to the excitation of the effector neuron. Their activation would, however, contribute to the excitation of the inhibitory neurons $I\bar{E}^*$, with which the neurons E have an appreciable overlap. This is the reason why the accumulation of neurons E which are transprinted with $(\bar{a})$ or $(\bar{v})$ or both, but not with $(\bar{f})$, will extinguish the previously established conditioned response to the compound stimulus.

Note to the First Model.—One more thing needs to be said at this point: it seems to be a fact that if we establish a conditioned salivary response in the dog to a compound stimulus, which has an auditory as well as a visual component, and if we subsequently extinguish the response, say to the visual component, we thereby automatically extinguish the conditioned response to the auditory component also. It can be shown that, in order to account for this fact, we must assume in our model that the central nervous system contains, in addition to a number of neurons E which are characterized by the set (e), about an equal number of neurons $\bar{E}$ which are characterized by the complementary set $(\bar{e})$, and that the neurons $\bar{E}$, characterized by one of these two sets, must contact through synapses the neuron E characterized by the complementary set (and vice versa). Presumably, this would mean that quite generally neurons characterized by complementary sets of specific

membrane proteins must be present in about equal numbers in the central nervous system of the individual.

The Second Model.—We may escape this complication (if a complication it is) by assuming that *every specific membrane protein is complementary to itself.* According to this second alternative model, any set of specific membrane proteins is then identical with the complementary set of specific membrane proteins. Thus we may write (e) in place of $(\bar{e})$ and (f) in place of $(\bar{f})$, etc. Accordingly, in our second model, *the overlap number of two sets is then defined as the number of specific membrane proteins which the two sets have in common, and when transprinting takes place, the transprinted neuron incorporates the sets of specific membrane proteins of the transprinting neurons.* Whatever functions the neural network, represented in Figure 1, would be capable of fulfilling on the basis of our first model, it would fulfill on the basis of our second model also, and the remainder of our discussion will be couched in terms of this second model, rather than the first one.

The Orderliness of the Inborn Neural Code.—According to the notions here adopted, we assume that two neurons in the central nervous system, which preferentially respond to two different sensory stimuli that "resemble" each other, must have a large overlap number. We assume that in the code of the congenitally determined neurons there is an orderly transition to smaller and smaller overlap numbers, as we go from one neuron to other neurons which differ from it more and more in their response-specificity. If it were otherwise, our model could not account for the phenomenon of the "generalization of stimuli" in the conditioned salivary reflex of the dog, first described by Pavlov.

Postscript.—If our two fundamental postulates are correct, then it ought to be possible to devise a neural network which would fully account for the phenomena exhibited by the conditioned responses of the autonomic nervous system. (The network described by Figure 1 represents a first attempt in this direction.) If one wanted to see, however, whether higher mental functions could be explained on the basis of our two fundamental postulates, then one would first have to invent adequate neural networks. Thus, if one wanted to see whether one could explain on this basis the mental functions which man is capable of performing, but the primates are not, one would perhaps have to invent the very same networks which are contained in the brain of man, but not in the brain of the primates. Clearly, this would be no mean task.

The "mental capacity" of suitable neural network models, operating on the basis of our two fundamental postulates, might be very high. For instance, the recording of information such as may be contained in a simple sentence would have to tie down only one "transprintable" neuron. Thus, if one were to expose an individual to a simple sentence every 4 seconds, 24 hours a day, and if, on each such occasion, one would tie down one transprintable neuron, then one would tie down just about 10^9 neurons over a period of 100 years. This is one tenth of the number of neurons believed to be contained in the human brain.

* This work was supported by a research grant, administered by The University of Chicago, of the General Medical Sciences Division of the National Institutes of Health.

[1] Anker, H. S., *Nature*, **188**, 938 (1960); Szilard, Leo, these PROCEEDINGS, **46**, 293–302 (1960).

[2] No neuron may, however, incorporate into its cell membrane the complementary counterpart of a specific membrane protein which its cell membrane already contains.

[3] Pavlov, I. P., *Conditioned Reflexes* (Oxford University Press, 1927).

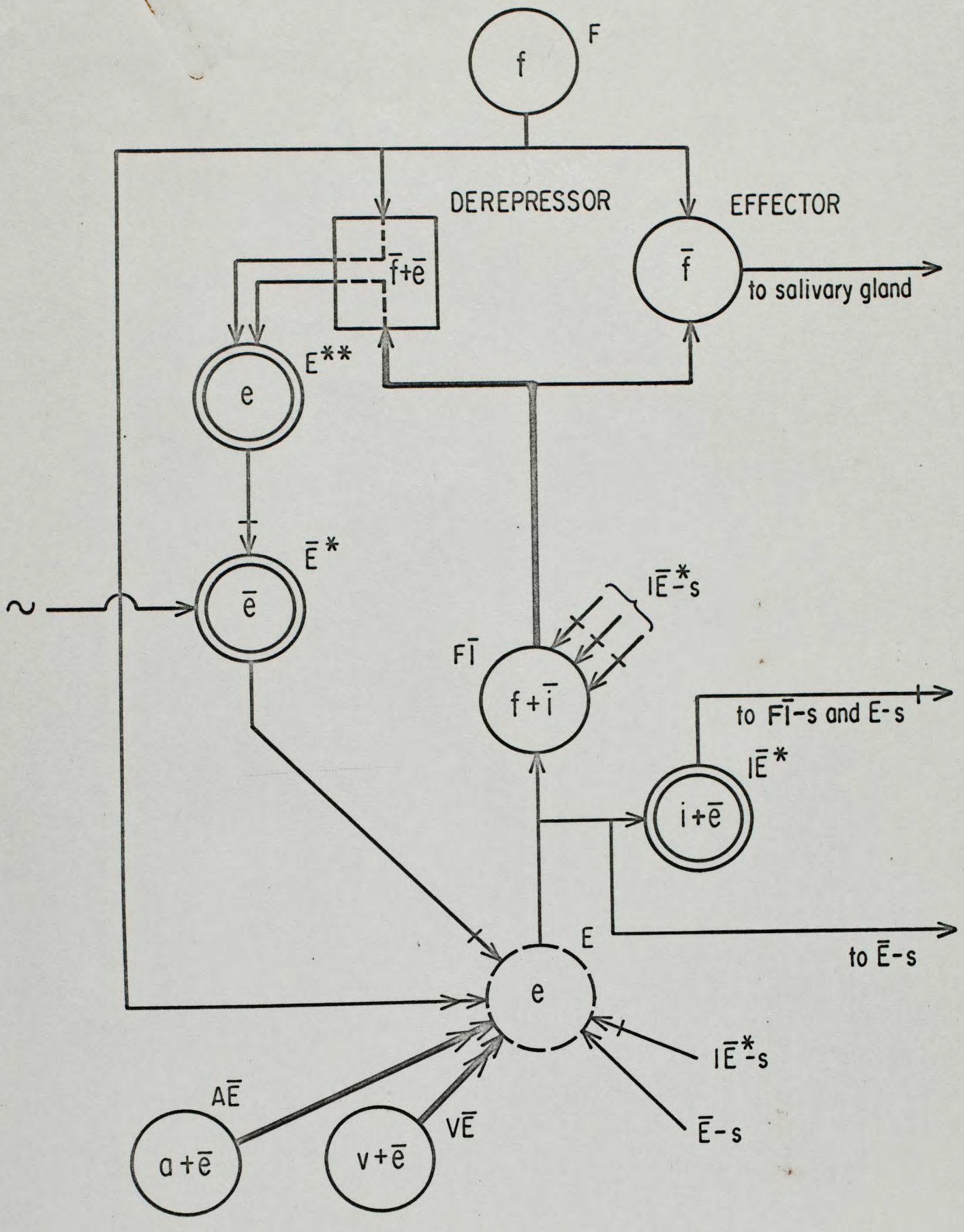


Figure 1: Excitatory neurons are represented by circles and inhibitory neurons are represented by double circles. Excitatory synapses are represented by simple arrows, except if they belong to neurons which are capable of transprinting, in which case they are represented by double arrows. Inhibitory synapses are represented by arrows with a crossbar. The transprintable neuron E is represented by a dotted circle.

uncorrected

NOT FOR RELEASE

Preprint of an article scheduled to appear in the June issue of the Proceedings of the National Academy of Sciences.

April 23, 1964

On Memory and Recall* - Part I

by Leo Szilard

The Salk Institute for Biological Studies
La Jolla, California

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

The Efficacy of a Synapse Bridging Two Neurons

Our neural network models involve excitatory neurons and some inhibitory neurons (of the kind which exert a post-synaptic inhibitory effect).

Let us consider an excitatory neuron which contacts through a synapse another neuron. If such an excitatory neuron sends a volley of nerve impulses to this synapse, then a certain quantity of an excitatory "transmitter substance" is released in the vicinity of the pre-synaptic membrane which diffuses across a gap --the synaptic cleft --into the post-synaptic neuron and raises the level of excitation of that neuron by a certain amount. We shall designate this excitatory transmitter substance as "acetylcholine" --in quotes. The "acetylcholine" which diffuses into the post-synaptic neuron is destroyed, in the vicinity of the post-synaptic membrane, by an enzyme which we shall designate as "choline esterase" --in quotes.

The rate at which "acetylcholine" is released in the vicinity of the pre-synaptic membrane is a function of the frequency of the nerve impulses which reach the synapse and we shall designate this rate as the "signal intensity". For the sake of

*This work was supported by a research grant, administered by The University of Chicago, of the General Medical Division of the National Institutes of Health.

simplicity, we shall assume that the "signal intensity" is for all synapses the same function of the frequency of the nerve impulses which are fed into the synapse.

The rate at which "acetylcholine" is destroyed in the post-synaptic neuron is proportional to the product of the concentration of "acetylcholine" and the concentration of the enzyme "choline esterase" in the vicinity of the post-synaptic membrane. Therefore, if, at a given point in time, nerve impulses of a certain frequency begin to arrive at a synapse, the "acetylcholine" concentration will begin to rise and will asymptotically approach a limit in the vicinity of the post-synaptic membrane, which is proportional to the "signal intensity" and inversely proportional to the concentration of "choline esterase", prevailing in the vicinity of the post-synaptic membrane. The "acetylcholine" concentration which is asymptotically approached at the post-synaptic membrane constitutes the "excitatory input", which is received from the synapse by the post-synaptic neuron. On this basis we may then say that, for any given signal intensity, the excitatory input received from a given synapse by the post-synaptic neuron is inversely proportional to the "choline esterase" concentration prevailing in the vicinity of the post-synaptic membrane of that synapse.

We assume that the enzyme "choline esterase" is inactivated at the post-synaptic membrane in different synapses at different rates and that this rate of inactivation is determined by the chemical specificities of the two neurons which are bridged by the synapse. We shall assume, for the sake of simplicity, however, <u>that the enzyme "choline esterase" is produced at the same rate in all excitatory neurons</u>.

We designate as the "efficacy" of the synapse the "excitatory input" which a post-synaptic neuron receives from that synapse, <u>per unit of "signal intensity"</u>. On the basis of the above assumptions we may then say that the efficacy of the synapse is inversely proportional to the rate at which "choline esterase" is inactivated at the post-synaptic membrane which, in turn, is determined by the chemical specificities of the two neurons which are bridged by the synapse.

The Rate of Inactivation of "Choline esterase"

We assume that neurons which differ from each other in their response-specificity contain a different set of certain specific proteins in their cell membrane. We shall refer hereafter to these proteins as the "specific membrane proteins".

We postulate that to each "specific membrane protein", there exists a complementary specific membrane protein and that a specific membrane protein molecule can combine with its complementary counterpart, just as an antibody molecule can combine with a molecule of its antigen. Accordingly, two complementary specific membrane proteins may behave as if they were, so to speak, each other's antibodies, as well as each other's antigens.

When an antibody molecule combines with an antigen molecule it undergoes an allosteric transition and an antibody molecule, when it is thus "dimerized", can bind complement. We assume that quite similarly a molecule of a "specific membrane protein", when it combines with its complementary counterpart, undergoes an allosteric transition and, when it is thus "dimerized", it can bind --and not only bind but also inactivate -- the enzyme "choline esterase".

The gap (synaptic cleft) between the pre-synaptic membrane and the post-synaptic membrane is estimated to be about 200 Å wide. According to the notions here adopted, there must be, however, a number of places within the active zones of the two synaptic membranes at which this gap is narrowed down, so that the pre-synaptic membrane and the post-synaptic membrane are in physical contact. We assume that at such a point of contact, a molecule of a "specific membrane protein", located in the post-synaptic membrane, can "dimerize" across the synaptic gap, with its complementary counterpart, located in the pre-synaptic membrane. The number of such "dimers", contained within the active zone of the synaptic membranes, would then determine the rate at which the enzyme "choline esterase" is inactivated at the post-synaptic membrane.

Let us now consider two neurons, A and B, which are bridged by a synapse. Neuron A is characterized by a set (a) of specific membrane proteins, which are present in its cell membrane and neuron B is characterized by another set, (b). We shall designate as the "overlap number" of these two neurons the number of specific membrane proteins contained within the set (a) which have their complementary counterpart contained within the set (b) [or vice versa].

From this overlap number we may compute the "efficacy" of the synapse which bridges these two neurons. In order to simplify this computation we shall assume that the area of the active zone of the synaptic membrane is the same for all synapses, and also assume, that the concentration of each "specific membrane protein" in the cell membrane is the same for any given neuron. On the basis of these simplifying assumptions, we may then say that the number of "dimers" contained within the active zone of the membrane of a synapse, which bridges neuron A and neuron B, is determined either by the ratio of the overlap number to the total number of specific membrane proteins of neuron A, or by the ratio of the overlap number to the total number of specific membrane proteins of neuron B --whichever ratio is smaller. We shall designate the smaller one of these two ratios as the "overlap fraction" of the neurons A and B. Accordingly, we may then say that the "efficacy" of a synapse bridging two neurons is proportional to the overlap fraction of the two neurons. This is the first fundamental postulate of our model.

We assume that the same holds true also for the synapses of our 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 neuron.

The Transprinting of Neurons:

We divide neurons of the Central Nervous System into two broad classes: the "congenitally-determined" neurons and the "memory" neurons. We designate the neurons

which attain their full chemical specificity of their cell membrane during the development of the individual (mostly during embryonal life and at the latest during the early post-natal period) as "congenitally-determined" neurons. If all the neurons of the Central Nervous System were of this sort then the individual would not be able to learn and his behavior would be wholly governed by the inborn reflexes. According to the notions here adopted, an adult can learn, and recall what he has learned, because his Central Nervous System contains "memory neurons" and each of these can, once in a lifetime, acquire an additional set of specific membrane proteins, through a process which we designate as "transprinting".

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 specific membrane proteins. We further assume that if, on such an occasion, a specific membrane protein penetrates into the memory neuron from a pre-synaptic neuron, then 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.

We assume that there is a class of "congenitally-determined" neurons which are capable of participating in the transprinting of a memory neuron. We assume that if such a "congenitally-determined" neuron 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 specific membrane proteins.

Let us now consider what happens if a "congenitally-determined" neuron of this sort, which contacts a memory neuron through a synapse, fires "simultaneously", with this memory neuron, so that for a period of time both the pre-synaptic and post-synaptic membranes are permeable for the specific membrane proteins. If this takes place then the specific membrane proteins of the pre-synaptic neuron will penetrate into the memory neuron and induce in the memory neuron their complementary counterparts.

If several such pre-synaptic neurons fire simultaneously with the memory neuron, then the memory neuron will on such an occasion acquire the sets of specific membrane proteins which are complementary to the sets of all of these pre-synaptic neurons.[1] This is the process of transprinting. Its occurrence as an "all or none" process, constitutes our second fundamental postulate.

We shall refer to memory neurons before they are transprinted, as "transprintable" neurons and, thereafter, we shall refer to them as "transprinted" neurons. "Transprinted"neurons are like "congenitally-determined" neurons inasmuch as they may transprint 'transprintable" neurons.

If a neuron participates in the transprinting of a "transprintable" neuron then we may expect this neuron and the "transprinted" neuron to have a large overlap fraction and, accordingly, we may expect synapses bridging these two neurons to have a high efficacy.

The Conditioned Response

In order to illustrate how transprinting may take place, we shall use as an example, the classical (Pavlovian) conditioning of the salivary reflex of the dog.[2] We shall indicate on this occasion, however, only rather sketchily what takes place during conditioning.

When "food" is introduced into the mouth of a dog, the dog responds with salivation. This is the inborn, or unconditioned, response. Let us now expose the dog to a compound stimulus which has an auditory, as well as a visual, component and let us -- before the compound stimulus is turned off --place "food" into the mouth of the dog. If, after several such "conditioning exposures", the dog is then presented for the first time with the compound stimulus, unreinforced, on this occasion, by the introduction of "food" into its mouth, the dog may be expected to salivate. This is the conditioned response.

---

(1) No neuron may, however, incorporate into its cell membrane the complementary counterpart of a specific membrane protein which its cell membrane already contains.

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

We assume that there is a neuron F in the Central Nervous System, characterized by the set (f), which preferentially responds to the stimulus of "food in the mouth". Moreover, we shall assume in particular that the signal to which the neuron F responds is the "onset" of this stimulus. As shown in Figure 1, the neuron F is connected through a synapse to an Effector neuron, which innervates the salivary gland. This Effector neuron is characterized by the set $(\bar{f})$, where $(\bar{f})$ denotes a set of specific membrane proteins which is complementary to the set (f). Because the overlap fraction of the neuron A and the Effector neuron is one, the synapses which bridge these two neurons have a high efficacy. Therefore, placing "food" into the dog's mouth may be expected to cause the dog to salivate.

In order to account for the conditioned response, we postulate the existence of a number of groups of transprintable neurons E. Each of these groups may consist of several hundred neurons E, all of which have the following in common. The neuron F contacts through a synapse each of the neurons E and in turn each neuron E contacts, through a synapse, an inter-neuron $F\bar{I}$ [characterized by the set $(f) + (\bar{i})$] which in turn contacts, through a synapse, the Effector neuron.

Until something happens which is "significant" from the point of view of the salivary reflex, all the transprintable neurons E are repressed, because they are inhibited by signals which are continuously being sent out by the inhibitory neuron E* [characterized by the set $(\bar{e})$]. This inhibition is assumed to be strong enough to prevent a transprintable neuron E to fire, even if it should receive a substantial aggregate "excitatory input", because the overlap fraction of the inhibitory neuron $\bar{E}^*$ and of the transprintable neuron E is one. It should be noted, however, that after the neuron E is transprinted and acquires a set of specific membrane proteins which is composed of a large number of such proteins, then its overlap fraction with the inhibitory neuron $\bar{E}^*$ is reduced by a substantial factor and the efficacy of the synapse bridging the two neurons is also reduced by the same factor. Accordingly, such a

"transprinted" neuron E may be caused to fire in spite of receiving inhibitory signals from the neuron $\bar{E}^*$.

The transprintable neurons E get de-repressed if the inhibitory neuron $\bar{E}^*$ is inhibited by signals emanating from a neural network designated as the "Derepressor". This will happen if the Derepressor sends out signals which are sufficiently strong to excite the inhibitory inter-neuron E**, which in turn will inhibit the inhibitory neuron $\bar{E}^*$, The Derepressor network may receive an input signal from the neuron F and it may also receive an input signal, via the inter-neuron $F\bar{I}$, from neurons E. These two input signals counteract each other 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. In our second paper we shall describe a very simple "neural network" which would function in this fashion.

As will be seen later, the Derepressor network may be expected to send out strong signals if "food" is introduced into the mouth of an unconditioned dog and to send out strong signals also when a dog, whose salivary reflex has been fully conditioned to a certain stimulus, is presented with that stimulus, without having, on this occasion, "food"placed into its mouth. The Derepressor network will not send out signals, however, if the fully conditioned dog is presented with the correct stimulus and "food" is placed into its mouth. Accordingly, no additional neurons E would be transprinted as the result of such "routine exposures".

It is probably generally true that a sensory experience is recorded only if there is "significance" attached to that experience. In our model of the conditioned salivary reflex there is "significance" attached to the stimulus "food in the mouth" for an unconditioned (or not fully conditioned) dog, while for a fully conditioned dog, there is "significance" attached to the conditioned stimulus, but only if that stimulus is not accompanied by the signal "food in the mouth".

We shall try to indicate next in what manner a conditioned salivary response may be established to a compound stimulus which has a visual and an auditory component. To this end we assume that in the Central Nervous System there is a neuron $A\bar{E}$ which responds preferentially to the auditory component of the compound stimulus, and another neuron $V\bar{E}$ which responds preferentially to the visual component. These two neurons are characterized by the sets (a) + ($\bar{e}$) and (v) + ($\bar{e}$) respectively. We assume that the number of different specific membrane proteins contained in the neurons $A\bar{E}$ and $V\bar{E}$ which we designate by $n(A\bar{E})$ and $n(V\bar{E})$, respectively, are large compared to the number of different specific membrane proteins contained in the transprintable neuron E, which we designate by n(E). Accordingly, we have: $n(A\bar{E}) > n(E)$, and $n(V\bar{E}) > n(E)$.

We assume that, out of a group of several hundred neurons E, a certain fraction is contacted through a synapse by the neuron $A\bar{E}$, a certain fraction is contacted through a synapse by the neuron $V\bar{E}$ and a certain fraction is contacted by both the neuron $A\bar{E}$, as well as the neuron $V\bar{E}$. Because the neurons $A\bar{E}$, as well as $V\bar{E}$ have an appreciable --even though small --overlap fraction with the transprintable neurons E, we may assume that if either of these two neurons fires, or if both of them fire, at a time when the transprintable neurons E are de-repressed, then one or more transprintable neurons will fire also and will on that occasion be transprinted by the neurons $A\bar{E}$ or $V\bar{E}$ or both. If, at the same time, the neuron F fires also, then these same neurons E will be transprinted also by the neuron F. The firing of neuron F alone would, however, not cause the neurons E to fire, even when the neurons E happen to be de-repressed because the neurons F and the transprintable neurons E have zero overlap.

If the unconditioned dog is for the first time exposed to the compound stimulus, and at the same time "food" is introduced into its mouth, the Derepressor network will send out a strong signal and one, or several, of the transprintable neurons E will be caused to fire. These neurons E will then be transprinted, on this occasion, with the the sets ($\bar{f}$), ($\bar{a}$) and ($\bar{v}$). If this dog is exposed for the second time to the compound

stimulus and at the same time "food" is introduced into its mouth, then the signal sent out by the Derepressor network will be somewhat weaker than the first time. This is so because the neurons E which have been transprinted, at the time of the first conditioning exposure, and which will be excited at the time of the second exposure, have a large overlap fraction with the inter-neuron $F\bar{I}$ and will, therefore, send a signal to the Derepressor network which counteracts the signal received by this network from the neuron F. As the conditioning process is continued and the dog is repeatedly subjected to such conditioning exposures, the neurons E which are transprinted with the sets $(\bar{f})$, $(\bar{a})$ and $(\bar{v})$ will increase in number. Finally, the Derepressor network will no longer send out a signal when the dog is exposed to the compound stimulus and at the same time "food" is introduced into its mouth. Such a dog is then fully conditioned and continuing the conditioning exposures would not make the conditioning any deeper.

Let us now expose such a fully conditioned dog to the compound stimulus, unreinforced on this occasion by the introduction of "food" into its mouth. The neurons E which have been transprinted during the previous conditioning exposures, with the set $(\bar{f})$, as well as the sets $(\bar{a})$ and $(\bar{v})$ will now be caused to fire. Because of the substantial overlap of the transprinted neurons E, which contain the set $(\bar{f})$, with the inter-neuron $F\bar{I}$, the firing of the neurons E will lead to the firing of the inter-neuron $F\bar{I}$ and this in turn will lead to the firing of the Effector neuron. Accordingly, on the occasion of this exposure of the dog to the compound stimulus, the dog will salivate. This is the conditioned response.

Incidentally, on this occasion when the inter-neuron $F\bar{I}$ fires, it will cause the Derepressor network to send out a strong signal because this network does not receive on this occasion a signal from the neuron F. Accordingly, on this occasion, one or more neurons E will get transprinted with the sets (a) or $(\bar{v})$ or both, but none of them will be transprinted with the set $(\bar{f})$. Therefore, if the dog is repeatedly exposed to

the compound stimulus in such a fashion, i.e. without reinforcement, then the number of neurons E which are transprinted with the sets $(\bar{a})$ or $(\bar{v})$ or both, but not with the set $(\bar{f})$, will increase on each such occasion. The overlap fraction of these transprinted neurons E with the inter-neurons $F\bar{I}$ is zero and, therefore, the excitation of these transprinted neurons E would not contribute to the excitation of the Effector neuron. Their activation would, however, contribute to the excitation of the inhibitory neurons $I\bar{E}$, with which the neurons E have an appreciable overlap. This is the reason why the accumulation of neurons E which are transprinted with $(\bar{a})$ or $(\bar{v})$ or both, but not with $(\bar{f})$, will extinguish the previously established conditioned response to the compound stimulus.

One more thing needs to be said at this point: It seems to be a fact that if we establish a conditioned salivary response in the dog to a compound stimulus, which has an auditory as well as a visual component, and if we subsequently extinguish the response, say to the visual component, we thereby automatically extinguish the conditioned response to the auditory component also. It can be shown that, in order to account for this fact, we must assume that the Central Nervous System contains, in addition to a number of neurons E which are characterized by the set (e), about an equal number of neurons $\bar{E}$ which are characterized by the complementary set $(\bar{e})$, and that the neurons E, characterized by one of these two sets, must contact through synapses the neuron $\bar{E}$, characterized by the complementary set (and vice versa). Presumably, this would mean that quite generally neurons characterized by complementary sets of specific membrane proteins must be present in about equal numbers in the Central Nervous System of each individual.

The Second Model

We may escape this complication (if a complication it is), by assuming that every specific membrane protein is complementary to itself. According to this second model, any set of specific membrane proteins is then identical with the complementary set of specific membrane proteins. Thus we may write (e) in place of $(\bar{e})$ and (f) in

place of ($\bar{f}$), etc. Accordingly, in our second model, the overlap number of two sets is then defined as the number of specific membrane proteins which the two sets have in common and when transprinting takes place, the "transprinted" neuron incorporates the sets of specific membrane proteins of the transprinting neurons. Whatever functions neural networks, of the kind represented in Figure 1, would be capable of fulfilling on the basis of our first model, they would fulfill on the basis of our second model also, and the remainder of our discussion will be couched in terms of this second model rather than the first one.

The Orderliness of the Inborn Neural Code

According to the notions here adopted, we assume that two neurons in the Central Nervous System, which preferentially respond to two different sensory stimuli that "resemble" each other, must have a large overlap number. We assume that in the code of the congenitally-determined neurons there is an orderly transition to smaller and smaller overlap numbers, as we go from one neuron to other neurons which differ from it more and more in their response-specificity. If it were otherwise, our model could not account for the phenomenon of the "generalization" of stimuli in the conditioned salivary reflex of the dog, first described by Pavlov.

Postcript

If our two fundamental postulates are correct, then it ought to be possible to devise a neural network which would fully account for the phenomena exhibited by the conditioned responses of the autonomous nervous system. (The network described by Figure 1 represents a first attempt in this direction). If one wanted to see, however, whether higher mental functions could be explained on the basis of our two fundamental postulates, then one would first have to invent adequate neural networks. Thus, if one wanted to see whether one could explain on this basis the mental functions which man is capable of performing, but the primates are not, one would perhaps have to invent the very same networks which are contained in the brain of man, but not in the

brain of the primates. Clearly, this would be no mean task.

The "mental capacity" of suitable neural network models, operating on the basis of our two fundamental postulates, might be very high. For instance, the recording of information such as may be contained in a "simple sentence" would have to tie down only one "transprintable" neuron. Thus, if one were to expose an individual to a simple sentence every four seconds, twenty-four hours a day, and if, on each such occasion, one would tie down one "transprintable" neuron, then one would tie down just about 10^9 neurons over a period of one hundred years. This is about one tenth of the number of neurons believed to be contained in the human brain.

THE END

Not For Release

Preprint of an article scheduled to appear in
the June issue of the Proceedings of the
National Academy of Sciences

On Memory and Recall* - Part I

by Leo Szilard

The Salk Institute for Biological Studies
La Jolla, California

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

The Efficacy of a Synapse Bridging Two Neurons: The neural network models which we shall be using here are based on excitatory neurons, as well as inhibitory neurons of one particular kind, the kind having post-synaptic inhibitory action.

Let us consider an excitatory neuron which contacts through a synapse another neuron. If such an excitatory neuron sends a volley of nerve impulses to this synapse, then a certain quantity of an excitatory "transmitter substance" is released in the vicinity of the pre-synaptic membrane which diffuses across a gap - the synaptic cleft - into the post-synaptic neuron and raises the level of excitation of that neuron by a certain amount. We shall designate this excitatory transmitter substance as "acetylcholine" - in quotes. The "acetylcholine" which diffuses into the post-synaptic neuron is destroyed, in the vicinity of the post-synaptic membrane, by an enzyme which we shall designate as "choline esterase" - in quotes.

The rate at which "acetylcholine" is released in the vicinity of the pre-synaptic membrane is a function of the frequency of the nerve impulses which reach the synapse and we shall designate this rate as the "signal intensity". For the sake

---

*This work was supported by a research grant, administered by the University of Chicago, of the Division of General Medical Sciences of the National Institutes of Health.

of simplicity, we shall assume that the "signal intensity" is for all synapses the same function of the frequency of the nerve impulses which are fed into the synapse.

The rate at which "acetylcholine" is destroyed in the post-synaptic neuron is proportional to the product of the concentration of "acetylcholine" and the concentration of the enzyme "choline esterase" in the vicinity of the post-synaptic membrane. Therefore, if, at a given point in time, nerve impulses of a certain frequency begin to arrive at a synapse, the "acetylcholine" concentration will begin to rise and will asymptotically approach a limit in the vicinity of the post-synaptic membrane, which is proportional to the "signal intensity" and inversely proportional to the concentration of "choline esterase", prevailing in the vicinity of the post-synaptic membrane. The "acetylcholine" concentration which is asymptotically approached at the post-synaptic membrane constitutes the "excitatory input", which is received from the synapse by the post-synaptic neuron. On this basis we may then say that, for any given "signal intensity", the excitatory input received from a given synapse by the post-synaptic neuron is inversely proportional to the "choline esterase" concentration prevailing in the vicinity of the post-synaptic membrane of that synapse.

We assume that the enzyme "choline esterase" is inactivated at the post-synaptic membrane in different synapses at different rates and that this rate of inactivation is determined by the chemical specificities of the two neurons which are bridged by the synapse. We shall assume, for the sake of simplicity, however, that the enzyme "choline esterase" is produced at the same rate in all excitatory neurons.

We designate as the "efficacy" of the synapse the "excitatory input" which a post-synaptic neuron receives from that synapse per unit of "signal intensity". On the basis of the above assumptions we may then say that the efficacy of the synapse is proportional to the rate at which "choline esterase" is inactivated at the post-synaptic membrane which, in turn, is determined by the chemical specificities of the two neurons which are bridged by the synapse.

The Rate of Inactivation of "Choline esterase": We assume that neurons which differ from each other in their response-specificity contain a different set of certain specific proteins in their cell membrane. We shall refer hereafter to these proteins as the "specific membrane proteins".

We postulate that to each "specific membrane protein", there exists a complementary specific membrane protein and that a specific membrane protein molecule can combine with its complementary counterpart, just as an antibody molecule can combine with a molecule of its antigen. Accordingly, two complementary specific membrane proteins may behave as if they were, so to speak, each other's antibodies, as well as each other's antigens.

When an antibody molecule combines with an antigen molecule it undergoes an allosteric transition and an antibody molecule, when it is thus "dimerized", can bind complement. We assume that quite similarly a molecule of a "specific membrane protein", when it combines with its complementary counterpart, undergoes an allosteric transition and, when it is thus "dimerized", it can bind - and not only bind but also inactivate - the enzyme "choline esterase".

The gap (synaptic cleft) between the pre-synaptic membrane and the post-synaptic membrane is estimated to be about 200 Å wide. According to the notions here adopted, there must be, however, a number of places within the active zones of the two synaptic membranes at which this gap is narrowed down, so that the pre-synaptic membrane and the post-synaptic membrane are in physical contact. We assume that at such a point of contact, a molecule of a "specific membrane protein", located in the post-synaptic membrane, can "dimerize" across the synaptic gap, with its complementary counterpart, located in the pre-synaptic membrane. The number of such "dimers", contained within the active zone of the synaptic membranes, would then determine the rate at which the enzyme "choline esterase" is inactivated at the post-synaptic membrane.

Let us now consider two neurons, A and B, which are bridged by a synapse. Neuron A is characterized by a set, (a), of specific membrane proteins, which are present

in its cell membrane and neuron B is characterized by another set, (b). We shall designate as the "overlap number" of these two neurons the number of specific membrane proteins contained within the set (a) which have their complementary counterpart contained within the set (b) [or vice versa].

From this overlap number we may compute the "efficacy" of the synapse which bridges these two neurons. In order to simplify this computation we shall assume that the area of the active zone of the synaptic membrane is the same for all synapses, and also assume, that the concentration of each "specific membrane protein" in the cell membrane is the same for any given neuron. On the basis of these simplifying assumptions, we may then say that the number of "dimers" contained within the active zone of the membrane of a synapse, which bridges neuron A and neuron B, is determined either by the ratio of the "overlap number" to the total number of specific membrane proteins of neuron A, or by the ratio of the "overlap number" to the total number of specific membrane proteins of neuron B - whichever ratio is smaller. We shall designate the smaller one of these two ratios as the "overlap fraction" of the neurons A and B. Accordingly, we may then say that the "efficacy" of a synapse bridging two neurons is proportional to the overlap fraction of the two neurons. This is the first fundamental postulate of our model.

We assume that the same holds true also for the synapses of our 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 neuron.

The Transprinting of Neurons: We divide neurons of the Central Nervous System into two broad classes: the "congenitally-determined" neurons and the "memory" neurons. We designate the neurons which attain their full chemical specificity of their cell membrane during the development of the individual (mostly during embryonal life and at the latest during the early post-natal period) as "congenitally-determined" neurons. If all the neurons of the Central Nervouse System were of this sort then

the individual would not be able to learn and his behavior would be wholly governed by the inborn reflexes. According to the notions here adopted, an adult can learn, and recall what he has learned, because his Central Nervous System contains "memory neurons" and each of these can, once in a lifetime, acquire an additional set of specific membrane proteins, through a process which we designate as "transprinting".

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. 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.(1) If several such pre-synaptic neurons fire simultaneously with the memory neuron, then the memory neuron will on such an occasion acquire the sets of specific membrane proteins which are complementary to the sets of all of these pre-synaptic neurons.(2) This is the process of transprinting. Its occurrence as an "all or none" process, constitutes our second fundamental postulate.

(1) H.S. Anker, Nature, 188 938, 1960
Leo Szilard, Proc. Nat. Acad. Sc. March, 1960

We shall refer to memory neurons before they are transprinted, as "transprintable" neurons, and, thereafter, we shall refer to them as "transprinted" neurons. "Transprinted" neurons resemble "congenitally-determined" neurons, in that they may transprint "transprintable" neurons.

If a neuron participates in the transprinting of a "transprintable" neuron then we may expect this neuron and the "transprinted" neuron to have a large overlap fraction and, accordingly, we may expect synapses bridging these two neurons to have a high efficacy.

The Conditioned Response: In order to illustrate how transprinting may take place, we shall use as an example, the classical (Pavlovian) conditioning of the salivary reflex of the dog.(3) We shall indicate on this occasion, however, only rather sketchily what takes place during conditioning.

When "food" is introduced into the mouth of a dog, the dog responds with salivation. This is the inborn, or unconditioned, response. Let us now expose the dog to a compound stimulus which has an auditory, as well as a visual, component and let us - before the compound stimulus is turned off - place "food" into the mouth of the dog. If, after several such "conditioning exposures", the dog is then presented for the first time with the compound stimulus, unreinforced on this occasion by the introduction of "food" into its mouth, the dog may be expected to salivate. This is the conditioned response.

We assume that there is a neuron F in the Central Nervous System, characterized by the set (f), which preferentially responds to the stimulus of "food in the mouth". Moreover, we shall assume, in particular, that the signal to which the neuron F responds is the "onset" of this stimulus.

As shown in Figure 1, the neuron F is connected through a synapse to an Effector neuron, which innervates the salivary gland. This Effector neuron is characterized

---

(2) No neuron may, however, incorporate into its cell membrane the complementary counterpart of a specific membrane protein which its cell membrane already contains.

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

by the set $(\bar{f})$, where $(\bar{f})$ denotes a set of specific membrane proteins which is complementary to the set (f). Because the overlap fraction of the neuron A and the Effector neuron is one, the synapses which bridge these two neurons have a high efficacy. Therefore, placing "food" into the dog's mouth may be expected to cause the dog to salivate.

In order to account for the conditioned response, we postulate the existence of a number of groups of transprintable neurons E. Each of these groups may consist of several hundred neurons E, all of which have the following in common. The neuron F contacts through a synapse each of the neurons E and in turn each neuron E contacts, through a synapse, an inter-neuron $F\bar{I}$ [characterized by the set $(f)+(\bar{i})$] which in turn contacts, through a synapse, the Effector neuron.

Until something happens which is "significant" from the point of view of the salivary reflex, all the transprintable neurons E are repressed, because they are inhibited by signals which are continuously being sent out by the inhibitory neurons $\bar{E}^*$ [characterized by the set $(\bar{e})$]. This inhibition is assumed to be strong enough to prevent a transprintable neuron E to fire, even if it should receive a substantial aggregate "excitatory input", because the overlap fraction of the inhibitory neuron $\bar{E}^*$ and of the transprintable neuron E is one. It should be noted, however, that after the neuron E is transprinted and acquires a set of specific membrane proteins which is composed of a large number of such proteins, then its overlap fraction with the inhibitory neuron $\bar{E}^*$ is reduced by a substantial factor and the efficacy of the synapse bridging the two neurons is also reduced by the same factor. Accordingly, such a "transprinted" neuron E may be caused to fire in spite of receiving inhibitory signals from the neuron $\bar{E}^*$.

The transprintable neurons E get de-repressed if the inhibitory neuron $\bar{E}^*$ is inhibited by signals emanating from a neural network designated as the "Derepressor". This will happen if the Derepressor sends out signals which are sufficiently strong to excite the inhibitory inter-neuron E**, which in turn will inhibit the inhibitory

neuron $\overline{E}$*. The Derepressor network may receive an input signal from the neuron F and it may also receive an input signal, via the inter-neuron $F\overline{I}$, from neurons E. These two input signals counteract each other 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. In our second paper we shall describe a very simple "neural network" which would function in this fashion.

As will be seen later, the Derepressor network may be expected to send out strong signals if "food" is introduced into the mouth of an inconditioned dog and to send out strong signals also when a dog, whose salivary reflex has been fully conditioned to a certain stimulus, is presented with that stimulus, without having, on this occasion, "food" placed into its mouth. The Derepressor network will not send out signals, however, if the fully conditioned dog is presented with the correct stimulus and "food" is placed into its mouth. Accordingly, no additional neurons E would be transprinted as the result of such "routine exposures".

It is probably generally true that a sensory experience is recorded only if there is "significance" attached to that experience. In our model of the conditioned salivary reflex there is "significance" attached to the stimulus "food in the mouth" for an unconditioned (or not fully conditioned) dog, while for a fully conditioned dog, there is "significance" attached to the conditioned stimulus, but only if that stimulus is not accompanied by the signal "food in the mouth".

We shall try to indicate next in what manner a conditioned salivary response may be established to a compound stimulus which has a visual and an auditory component. To this end we assume that in the Central Nervous System there is a neuron $A\overline{E}$ which responds preferentially to the auditory component of the compound stimulus, and another neuron $V\overline{E}$ which responds preferentially to the visual component. These two neurons are characterized by the sets $(a)+(\overline{e})$ and $(v)+(\overline{e})$ respectively. We assume that the number of different specific membrane proteins contained in the neurons $A\overline{E}$ and $V\overline{E}$, which we designate by $n(A\overline{E})$ and $n(V\overline{E})$, respectively, are large

compared to the number of different specific membrane proteins contained in the transprintable neuron E, which we designate by n(E). Accordingly, we have: $n(\overline{AE}) > n(E)$, and $n(\overline{VE}) > n(E)$.

We assume that, out of a group of several hundred neurons E, a certain fraction is contacted through a synapse by the neuron $\overline{AE}$, a certain fraction is contacted through a synapse by the neuron $\overline{VE}$ and a certain fraction is contacted by both the neuron $\overline{AE}$, as well as the neuron $\overline{VE}$. Because the neuron $\overline{AE}$, as well as the neuron VE has an appreciable - even though small - overlap fraction with the transprintable neurons E, we may assume that if either of these two neurons fires, or if both of them fire, at a time when the transprintable neurons E are de-repressed, then one or more transprintable neurons E will fire also and will on that occasion be transprinted by the neurons $\overline{AE}$ or $\overline{VE}$ or both. If, at the same time, the neuron F fires also, then these same neurons E will be transprinted also by the neuron F. The firing of neuron F alone would, however, not cause the neuron E to fire, even when the neurons E happen to be de-repressed, because the neurons F and the transprintable neurons E have zero overlap.

If the unconditioned dog is for the first time exposed to the compound stimulus, and at the same time "food" is introduced into its mouth, the Derepressor network will send out a strong signal and one, or several, of the transprintable neurons E will be caused to fire. These neurons E will then be transprinted, on this occasion, with the sets $(\overline{f})$, $(\overline{a})$ and $(\overline{v})$. If this dog is exposed, for the second time, to the compound stimulus and at the same time "food" is introduced into its mouth, then the signal sent out by the Derepressor network will be somewhat weaker than the first time. This is so because the neurons E which have been transprinted, at the time of the first conditioning exposure, and which will be excited at the time of the second exposure, have a large overlap fraction with the inter-neuron $\overline{FI}$ and will, therefore, send a signal to the Derepressor network which counteracts the signal received by this network from the neuron F. As the conditioning process is continued and the dog is repeatedly subjected to such "conditioning exposures", the neurons E which are trans-

printed with the sets $(\overline{f})$, $(\overline{a})$ and $(\overline{v})$ will increase in number. Finally, the Derepressor network will no longer send out a signal when the dog is exposed to the compound stimulus and at the same time "food" is introduced into its mouth. Such a dog is then fully conditioned and continuing the conditioning exposures would not make the conditioning any deeper.

Let us now expose such a fully conditioned dog to the compound stimulus, unreinforced on this occasion by the introduction of "food" into its mouth. The neurons E which have been transprinted during the previous conditioning exposures, with the set $(\overline{f})$, as well as the sets $(\overline{a})$ and $(\overline{v})$ will now be caused to fire. Because of the substantial overlap of the transprinted neurons E, which contain the set $(\overline{f})$, with the inter-neuron $\overline{FI}$, the firing of the neurons E will lead to the firing of the inter-neuron $\overline{FI}$ and this in turn will lead to the firing of the Effector neuron. Accordingly, on the occasion of this unreinforced exposure of the dog to the compound stimulus, the dog will salivate. This is the conditioned response.

Incidentally, on this occasion when the inter-neuron $\overline{FI}$ fires, it will cause the Derepressor network to send out a strong signal because this network does not receive on this occasion a signal from the neuron F. Accordingly, on this occasion, one or more neurons E will get transprinted with the sets $(\overline{a})$, or $(\overline{v})$ or both, but none of them will be transprinted with the set $(\overline{f})$. Therefore, if the dog is repeatedly exposed to the compound stimulus in such a fashion, i.e. without reinforcement, then the number of neurons E which are transprinted with the sets $(\overline{a})$ or $(\overline{v})$ or both, but not with the set $(\overline{f})$, will increase on each such occasion. The overlap fraction of these transprinted neurons E with the inter-neurons $\overline{FI}$ is zero and therefore, the excitation of these transprinted neurons E would not contribute to the excitation of the Effector neuron. Their activation would, however, contribute to the excitation of the inhibitory neurons $\overline{IE}^*$, with which the neurons E have an appreciable overlap. This is the reason why the accumulation of neurons E which are transprinted with $(\overline{a})$ or $(\overline{v})$ or both, but not with $(\overline{f})$, will extinguish

the previously established conditioned response to the compound stimulus.

Note to the First Model:

One more thing needs to be said at this point: It seems to be a fact that if we establish a conditioned salivary response in the dog to a compound stimulus, which has an auditory as well as a visual component, and if we subsequently extinguish the response, say to the visual component, we thereby automatically extinguish the conditioned response to the auditory component also. It can be shown that, in order to account for this fact, we must assume in our model that the Central Nervous System contains, in addition to a number of neurons E which are characterized by the set (e), an about equal number of neurons $\overline{E}$ which are characterized by the complementary set $(\overline{e})$, and that the neurons E, characterized by one of these two sets, must contact through synapses the neurons $\overline{E}$, characterized by the complementary set (and vice versa). Presumably, this would mean that quite generally neurons characterized by complementary sets of specific membrane proteins must be present in about equal numbers in the Central Nervous System of each individual.

The Second Model: We may escape this complication (if a complication it is), by assuming that every specific membrane protein is complementary to itself. According to this second, alternative, model, any set of specific membrane proteins is then identical with the complementary set of specific membrane proteins. Thus we may write (e) in place of $(\overline{e})$ and (f) in place of $(\overline{f})$, etc. Accordingly, in our second model, the overlap number of two sets is then defined as the number of specific membrane proteins which the two sets have in common and when transprinting takes place, the "transprinted" neuron incorporates the sets of specific membrane proteins of the transprinting neurons. Whatever functions neural networks, of the kind represented in Figure 1, would be capable of fulfilling on the basis of our first model, they would fulfill on the basis of our second model also, and the remainder of our discussion will be couched in terms of this second model, rather than the first one.

The Orderliness of the Inborn Code: According to the notions here adopted, we assume that two neurons in the Central Nervous System, which preferentially respond to two different sensory stimuli that "resemble" each other, must have a

large overlap number. We assume that in the code of the "congenitally-determined" neurons there is an orderly transition to smaller and smaller overlap numbers, as we go from one neuron to other neurons which differ from it more and more in their response-specificity. If it were otherwise, our model could not account for the phenomenon of the "generalization of stimuli" in the conditioned salivary reflex of the dog, first described by Pavlov.

Postscript: If our two fundamental postulates are correct, then it ought to be possible to devise a neural network which would fully account for the phenomena exhibited by the conditioned responses of the autonomous nervous system. (The network described by Figure 1 represents a first attempt in this direction). If one wanted to see, however, whether higher mental functions could be explained on the basis of our two fundamental postulates, then one would first have to invent adequate neural networks. Thus, if one wanted to see whether one could explain on this basis the mental functions which man is capable of performing, but the primates are not, one would perhaps have to invent the very same networks which are contained in the brain of man, but not in the brain of the primates. Clearly, this would be no mean task.

The "mental capacity" of suitable neural network models, operating on the basis of our two fundamental postulates, might be very high. For instance, the recording of information such as may be contained in a "simple sentence" would have to tie down only one "transprintable" neuron. Thus, if one were to expose an individual to a simple sentence every four seconds, twenty-four hours a day, and if, on each such occasion, one would tie down one "transprintable" neuron, then one would tie down just about 10^9 neurons, over a period of one hundred years. This is about one tenth of the number of neurons to be contained in the human brain.

THE END

NOT FOR RELEASE

Preprint of an article scheduled to appear in the June issue of the Proceedings of the National Academy of Sciences.

April 23, 1964

On Memory and Recall* - Part I

by Leo Szilard

The Salk Institute for Biological Studies
La Jolla, California

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

The Efficacy of a Synapse Bridging Two Neurons

Our neural network models involve both excitatory neurons and inhibitory neurons, each of one particular kind.

Let us consider an excitatory neuron which contacts through a synapse another neuron. If such an excitatory neuron sends a volley of nerve impulses to this synapse, then a certain quantity of an excitatory "transmitter substance" is released in the vicinity of the pre-synaptic membrane which diffuses across a gap --the synaptic cleft-- into the post-synaptic neuron and raises the level of excitation of that neuron by a certain amount. We shall designate this excitatory transmitter substance as "acetylcholine" --in quotes. The "acetylcholine" which diffuses into the post-synaptic neuron is destroyed, in the vicinity of the post-synaptic membrane, by an enzyme which we shall designate as "choline esterase" --in quotes.

---

*This work was supported by a research grant, administered by The University of Chicago, of the General Medical Division of the National Institutes of Health.

The rate at which "acetylcholine" is released in the vicinity of the pre-synaptic membrane is a function of the frequency of the nerve impulses which reach the synapse and we shall designate this rate as the "signal intensity". For the sake of simplicity, we shall assume that the "signal intensity" is for all synapses the same function of the frequency of the nerve impulses which are fed into the synapse, i.e. ~~any~~ a given frequency of the nerve impulses would produce for all synapses the same "signal intensity".

The rate at which "acetylcholine" is destroyed in the post-synaptic neuron is proportional to the product of the concentration of "acetylcholine" and the concentration of the enzyme "choline esterase" in the vicinity of the post-synaptic membrane. Therefore, if, at a given point in time, nerve impulses of a certain frequency begin to arrive at a synapse, the "acetylcholine" concentration will begin to rise and will asymptotically approach a limit in the vicinity of the post-synaptic membrane, which is proportional to the "signal intensity" and inversely proportional to the concentration of "choline esterase", prevailing in the vicinity of the post-synaptic membrane. The "acetylcholine" concentration which is asymptotically approached at the post-synaptic membrane ~~is~~ constitutes the "excitatory input", which is received from the synapse by the post-synaptic neuron. On this basis we may then say that, for any given signal intensity, the excitatory input received from a given synapse by the post-synaptic neuron is inversely proportional to the "choline esterase" concentration prevailing in the vicinity of the post-synaptic membrane of that synapse.

We assume that the enzyme "choline esterase" is inactivated at the post-synaptic membrane at different rates in different synapses and that ~~the~~ this rate of inactivation is determined by the chemical specificities of the two neurons which are bridged by the synapse. We shall assume, for the sake of simplicity, however, that the enzyme "choline esterase" is produced at the same rate in all excitatory neurons.

We designate as the "efficacy" of the synapse the "excitatory input" which a post-synaptic neuron receives from that synapse, per unit of "signal intensity."

— April 26

~~intensity" as the "efficacy" of the synapse.~~ On the basis of the above assumptions we may then say that the efficacy of the synapse is inversely proportional to the rate at which "choline esterase" is inactivated at the post-synaptic membrane ~~and,~~ which in turn accordingly, is determined by the chemical specificities of the two neurons which are bridged by the synapse.

The Rate of Inactivation of "Choline esterase"

We assume that neurons which differ from each other in their response-specificity contain a different set of certain specific proteins in their cell membrane. We shall refer hereafter to these proteins as the "specific membrane proteins".

We postulate that to each "specific membrane protein", there exists a complementary specific membrane protein and that a specific membrane protein molecule can combine with its complementary counterpart, just as an antibody molecule can combine with a molecule of its antigen. Accordingly, two complementary specific membrane proteins may behave as if they were, so to speak, each other's antibodies, as well as each other's antigens.

When an antibody molecule combines with an antigen molecule it undergoes an allosteric transition and an antibody molecule, when it is thus "dimerized", can bind complement. We assume that quite similarly a molecule of a "specific membrane protein", when it combines with its complementary counterpart, undergoes an allosteric transition and, when it is thus "dimerized", it can bind --and not only bind but also inactivate -- --the enzyme "choline esterase".

The gap (synaptic cleft) between the pre-synaptic membrane and the post-synaptic membrane is estimated to be about 200 Å wide. According to the notions here adopted, there must be, however, a number of places within the active zones of the two synaptic membranes at which this gap is narrowed down, so that the pre-synaptic membrane and the post-synaptic membrane are in physical contact. We assume that at such a point

of contact, a molecule of a "specific membrane protein," located in the post-synaptic membrane, can "dimerize" across the synaptic gap, with its complementary counterpart, located in the pre-synaptic membrane. The number of such "dimers", contained within the active zone of the synaptic membranes, ~~will~~ would then determine the rate at which the enzyme "choline esterase" is inactivated at the post-synaptic membrane.

Let us now consider two neurons, A and B, which are bridged by a synapse. Neuron A is characterized by a set (a) of specific membrane proteins, ~~designated by (a),~~ which are present in its cell membrane ~~at an elevated level of concentration,~~ and neuron B is characterized by another set (b). We shall designate as the "overlap number" of these two neurons the number of specific membrane proteins contained within the set (a) which have their complementary counterpart contained within the set (b) [or vice versa].

From this overlap number we may compute the "efficacy" of the synapse which bridges these two ~~such~~ neurons. In order to simplify this computation we shall assume that the area of the active zone of the synaptic membrane is the same for all synapses, and also assume that, for any given neuron, the concentration of each "specific membrane protein" in the cell membrane is the same. On the basis of these simplifying assumptions, we may then say that the number of "dimers" contained within the active zone of the membrane of a synapse which bridges neuron A and neuron B, is determined either by the ratio of the overlap number to the total number of specific membrane proteins of neuron A, or by the ratio of the overlap number to the total number of specific membrane proteins of neuron B --whichever ratio is smaller. We shall designate the smaller one of these two ratios as the "overlap fraction" of the neurons A and B. ~~and~~ Accordingly, we may then say that the "efficacy" of a synapse bridging two neurons is proportional to the overlap fraction of the two neurons. This is the first ~~basic~~ fundamental postulate of our model.

We assume that the same holds true also for the synapses of our 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.

The Transprinting of Neurons:

We divide neurons of the Central Nervous System into two broad classes: the "congenitally-determined" neurons and the "memory" neurons. We designate the neurons which attain their full chemical specificity of their cell membrane during the development of the individual (mostly during embryonal life and at the latest during the early post-natal period) ~~we designate~~ as "congenitally-determined" neurons.

If all the neurons of the Central Nervous System were of this sort then the individual would not be able to learn and his behavior would be wholly governed by inborn reflexes.

According to the notions here adopted, an adult can learn, and recall what he has learned, because his Central Nervous System contains "memory neurons" and each of these can, once in his lifetime, acquire an additional set of specific membrane proteins, through a process which we designate as "transprinting".

We assume that when a memory neuron fires then those parts of the cell membrane, covering its cell body ~~of~~ and its dendrites, which constitute the active zones of the post-synaptic membranes, become permeable for specific membrane proteins. We further assume that if, on such an occasion, a specific membrane protein penetrates into the memory neuron from a pre-synaptic neuron, then it induces in the memory neuron the complementary specific membrane protein ~~(1)~~ --just as an antigen induces its antibody if it penetrates into certain lymphatic cells of the rabbit.~~(1)~~

We assume that there is a class of "congenitally-determined" neurons which

---

~~(1) H.S. Anker, Nature, 188, p. 938, 1960.~~

We assume that

are capable of participating in the transprinting of a memory neuron. If such a "congenitally-determined" neuron fires then those parts of ~~the~~ 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 specific membrane proteins.

Let us now consider what happens if a "congenitally-determined" neuron of this sort, which contacts a memory neuron through a synapse, fires "simultaneously", with this memory neuron, so that for a period of time both the pre-synaptic and post-synaptic membranes are permeable for the specific membrane proteins. If this takes place then the specific membrane proteins of the pre-synaptic neuron will penetrate into the memory neuron and induce in the memory neuron ~~the~~ their complementary ~~specific membrane proteins~~ counterparts. If several such pre-synaptic neurons fire simultaneously with the memory neuron, then the memory neuron will on such an occasion acquire the sets of specific membrane proteins which are complementary to the sets of all of these pre-synaptic neurons.(1) This is the process of transprinting. Its occurrence as an "all or none" process, constitutes our second fundamental postulate.

We shall refer to memory neurons before they are transprinted, as "transprintable" neurons and, thereafter, we shall refer to them as "transprinted" neurons. If a neuron participates in the transprinting of a "transprintable" neuron then we may expect this neuron and the "transprinted" neuron to have a large overlap fraction and, accordingly, we may expect synapses bridging these two neurons to have a high efficacy.

The Conditioned Response

In order to illustrate how transprinting may take place, we shall use as an example, the classical (Pavlovian) conditioning of the salivary reflex of the dog.(2)

---

(1) No neuron may, however, incorporate into its cell membrane the complementary counterpart of a specific membrane protein which its cell membrane already contained.

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

We shall indicate on this occasion, however, only rather sketchily what takes place during conditioning.

When "food" is introduced into the mouth of a dog, the dog responds with salivation. This is the inborn, or unconditioned, response.

Let us now expose the dog to a compound stimulus which has an auditory, as well as a visual, component and let us --before the compound stimulus is turned off --place "food" into the mouth of the dog. If, after several such "conditioning exposures", the dog is then presented for the first time with the compound stimulus, unreinforced (on this occasion) by the introduction of "food" into its mouth, the dog may be expected to salivate. This is the conditioned response.

We assume that there is a neuron F in the Central Nervous System, characterized by the set (f), which preferentially responds to the stimulus of "food in the mouth". Moreover, we shall assume in particular that the signal to which the neuron F responds is the "onset" of this stimulus. As shown in Figure 1, the neuron F is connected through a synapse to an Effector neuron, which innervates the salivary gland. This Effector neuron is characterized by the set $(\bar{f})$, where $(\bar{f})$ denotes a set of specific membrane proteins which is complementary to the set (f). Because the overlap fraction of the neuron A ~~with~~ and the Effector neuron is one, the synapses which bridge these two neurons have a high efficacy. Therefore, placing "food" into the dog's mouth may be expected to cause the dog to salivate.

In order to account for the conditioned response, we postulate the existence of a number of groups of transprintable neurons E. Each of these groups may consist of several hundred neurons E, all of which have the following in common:

The neuron F contacts through a synapse each of the neurons E and in turn each neuron E contacts, through a synapse, an inter-neuron $F\bar{I}$ [characterized by the set $(f) + (\bar{i})$] which in turn contacts, through a synapse, the Effector neuron.

Until something happens which is "significant" from the point of view of the salivary reflex, all the transprintable neurons E are repressed, because they are inhibited by signals which are continuously being sent out by the inhibitory neuron E* [characterized by the set $(\bar{e})$]. This inhibition is assumed to be strong enough to prevent a transprintable neuron E to fire, even if it should receive a substantial "aggregate excitatory input", because the overlap fraction of the inhibitory neuron $\bar{E}^*$ and of the transprintable neuron E is one. It should be noted, however, that after the neuron E is transprinted and acquires a set of specific membrane proteins which is composed of a large number of such proteins, then its overlap fraction with the inhibitory neuron $\bar{E}^*$ is reduced by a substantial factor and the efficacy of the synapse bridging the two neurons is also reduced by the same factor. Accordingly, such a "transprinted" neuron E may be caused to fire in spite of receiving inhibitory signals from the neuron $\bar{E}^*$.

The transprintable neurons E get de-repressed if the inhibitory neuron $\bar{E}^*$ is inhibited by signals emanating from a neural network designated as the "Derepressor". This will happen if the Derepressor sends out signals which are sufficiently strong to excite the inhibitory inter-neuron E**, which in turn will inhibit the inhibitory neuron $\bar{E}^*$.

The Derepressor network may receive an input signal from the neuron F and it may also receive an input signal, via the inter-neuron $F\bar{I}$, from neurons E. These two input signals counteract each other 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. In our second paper we shall describe a very simple "neural network" which would function in this fashion.

As will be seen later, the Derepressor network may be expected to send out strong signals if "food" is introduced into the mouth of an unconditioned dog and to send out strong signals also when a dog, whose salivary reflex has been fully conditioned, to a certain stimulus, is presented with that stimulus, without having "food" placed into its mouth on this occasion. The Depressor network will not send out signals, however, if the fully conditioned dog is presented with the correct stimulus and "food" is placed into its mouth. Accordingly, no additional neurons would E will be transprinted as the result of such "routine exposures".

It is probably generally true that a sensory experience is recorded only if there is "significance" attached to that experience. In our model of the conditioned salivary reflex there is "significance" attached to the stimulus "food in the mouth" for an unconditioned (or not fully conditioned) dog, while for a fully conditioned dog, there is "significance" attached to the conditioned stimulus, but only if provided that stimulus is not accompanied by the signal "food in the mouth".

We shall try to indicate next in what manner a conditioned salivary response may be established to a compound stimulus which has a visual and an auditory component. To this end we assume that in the Central Nervous System there is a neuron $A\bar{E}$ which respnds preferentially to the auditory component of the compound stimulus, and another neuron $V\bar{E}$ which responds preferentially to the visual component. These two neurons are characterized by the sets (a) + ($\bar{e}$) and (v) + ($\bar{e}$) respectively. We assume that the number of different specific membrane proteins contained in the neurons $A\bar{E}$ and $V\bar{E}$ which are designated by $n(A\bar{E})$ and $n(V\bar{E})$, respectively, are large compared to the number of different specific membrane proteins contained in the transprintable neuron E, which we designate by n(E). Accordingly, we have: $n(A\bar{E}) > n(E)$, and $n(V\bar{E}) > n(E)$.

We assume that, out of a group of several hundred neurons E, a certain fraction is contacted through a synapse by the neuron $A\bar{E}$, ~~a certain fraction is contacted~~

through a synapse by the neuron $A\bar{E}$, a certain fraction is contacted through a synapse by the neuron $V\bar{E}$ and a certain fraction is contacted through a synapse by both the neuron $A\bar{E}$, as well as through another synapse by the neuron $V\bar{E}$. Because the neurons $A\bar{E}$, as well as $V\bar{E}$ have an appreciable --even though small --overlap fraction with the transprintable neurons E, we may assume that if either of these two neurons fires, or if both of them fire, at a time when the transprintable neurons E are de-repressed, then one or more transprintable neurons will fire also and will on that occasion be transprinted by the neurons $A\bar{E}$ or $V\bar{E}$ or both. If at the same time when the neurons $A\bar{E}$ or $V\bar{E}$, or both, fire, the neuron F happens to fire also, then these same neurons E will be transprinted also by the neuron F. The firing of neuron F alone would, however, not cause the neurons E to fire, even when the neurons E happen to be de-repressed, because the neurons F and the transprintable neurons E have zero overlap.

If When the unconditioned dog is for the first time exposed to the compound stimulus, and at the same time "food" is introduced into its mouth, the Derepressor network will send out a strong signal and one, or several, of the transprintable neurons E will be caused to fire. These neurons E will then be transprinted, on this occasion, with the sets $(\bar{f})$, $(\bar{a})$ and $(\bar{v})$. If When this dog is exposed for the second time to the compound stimulus and at the same time "food" is introduced into its mouth, then the signal sent out by the Derepressor network will be somewhat weaker than the first time. This is so because the neurons E which have been transprinted, at the time of the first conditioning exposure, and which will be excited at the time of by the second exposure, have a large overlap fraction with the inter-neuron $F\bar{I}$ and will, therefore, send a signal to the Derepressor network, which counteracts the effect of the signal received by this network from the neuron F. As the conditioning process is continued and the dog is repeatedly subjected to such conditioning exposures, the neurons E which are transprinted with the sets $(\bar{f})$, $(\bar{a})$ and $(\bar{v})$ will increase in number. Finally, the

Derepressor network will no longer send out a signal when ~~a~~ the dog is exposed to the compound stimulus and at the same time "food" is introduced into its mouth. Such a dog is then fully conditioned and continuing the conditioning exposures would not make the conditioning any deeper.

Let us now expose such a fully conditioned dog to the compound stimulus, unreinforced on this occasion by the introduction of "food" into its mouth. The neurons E which have been transprinted during the previous conditioning exposures, with the set $(\bar{f})$, as well as the sets $(\bar{a})$ and $(\bar{v})$ will now be caused to fire. Because of the substantial overlap of the transprinted neurons E, which contain the set $(\bar{f})$, with the inter-neuron $F\bar{I}$, the firing of the neurons E will lead to the firing of the inter-neuron $F\bar{I}$ and this in turn will lead to the firing of the Effector neuron. Accordingly, on the occasion of this exposure of the dog to the compound stimulus, the dog will salivate. This is the conditioned response.

Incidentally, ~~when~~ on this occasion when the inter-neuron $F\bar{I}$ fires, it will cause the Derepressor network to send out a strong signal because this network does not receive on this occasion a signal from the neuron F. Accordingly, on this occasion one or more neurons E will ~~be~~ get transprinted with the sets $(\bar{a})$, or $(\bar{v})$ or both, but none of them will be transprinted with the set $(\bar{f})$. Therefore, if the dog is repeatedly exposed to the compound stimulus in such a fashion, i.e. without reinforcement, then the number of neurons E which are transprinted with the sets $(\bar{a})$ or $(\bar{v})$ or both, but not with the set $(\bar{f})$ will increase on each such occasion. The overlap fraction of these transprinted neurons E with the inter-neurons $F\bar{I}$ is zero and, therefore, the excitation of these transprinted neurons E would not contribute to the excitation of the Effector neuron. Their activation would, however, contribute to the excitation of the inhibitory neurons $I\bar{E}$, with which the neurons E have an appreciable overlap. This is the reason why the accumulation of neurons E which are transprinted with $(\bar{a})$ or $(\bar{v})$ or both, but not with $(\bar{f})$ will extinguish ~~leads to the extinguishing of~~ the previously established conditioned

response to the compound stimulus.

One more thing needs to be said ~~on this occasion~~ at this point: It seems to be a fact that if we establish a conditioned salivary response in the dog to a compound stimulus, which has an auditory ~~and~~ as well as a visual component, and if we subsequently extinguish the response, say to the visual component, we thereby automatically extinguish the conditioned response to the auditory component also. It can be shown that, in order to account for this fact, we must assume that the Central Nervous System contains, in addition to a number of neurons E which are characterized by the set (e), about an equal number of neurons $\bar{E}$ which are characterized by the complementary set $(\bar{e})$, and that the neurons E, characterized by one of these two sets, must contact through synapses, the neuron $\bar{E}$ characterized by the complementary set (and vice versa). Presumably, this would mean that, quite generally ~~that~~ neurons characterized by complementary sets of specific membrane proteins must be present in about equal numbers in the Central Nervous System of ~~the~~ each individual.

The Second Model

We may escape this complication (if a complication it is), by assuming that every specific membrane protein is complementary to itself. According to this second model, any set of specific membrane proteins is ~~thus~~ then identical with the complementary set of specific membrane proteins. Thus we may write (e) in place of $(\bar{e})$ and (f) in place of $(\bar{f})$, etc. Accordingly, in our second model, the overlap number of two sets is then defined as the number of specific membrane proteins which the two sets have in common. ~~Further,~~ And when transprinting takes place, the "transprinted" neuron incorporates the sets of specific membrane proteins of the transprinting neurons.

~~For the moment we shall leave aside the question to what extent this second model may be regarded as biologically plausible.~~ Whatever functions neural networks of the kind represented in Figure 1, would be capable of fulfilling on the basis of our

second model also, and the remainder of our discussion will be couched in terms of this second model rather than the first one.

The Orderliness of the Inborn Neural Code: According to the notions here adopted, we assume that two neurons in the Central Nervous System, which preferentially respond to two different sensory stimuli that "resemble" each other, must have a large overlap number. We assume that in the code of the congenitally-determined neurons there is an orderly transition to smaller and smaller overlap numbers, as we go from one neuron to other neurons which differ from it more and more in their response-specificity. If it were otherwise, our model could not account for the phenomenon of "generalization" of the conditioned salivary reflex of the dog, first described by Pavlov.

Postcript: If our two fundamental postulates are correct, then it ought to be possible to devise a neural network which would fully account for the phenomena exhibited by the conditioned responses of the autonomous nervous system. (The network described by Figure 1 represents ~~only~~ a first attempt in this direction.) ~~and it remains to be seen to what extent it can account for such phenomena as "inhibition" and "differentiation" (discrimination) of the conditioned response.~~ If one wanted to see, however, whether higher mental functions could be explained on the basis of our two fundamental postulates, then one would first have to invent adequate neural networks. Thus, if one wanted to see whether one could explain on this basis the mental functions which man is capable of performing, but the primates are not, one ~~might~~ would perhaps have to invent the very same networks which are contained in the brain of man, but not in the brain of the primates. Clearly, this would be no mean task.

The "mental capacity" of suitable neural network models, operating on the basis of our two fundamental postulates, might be very high, ~~and~~ For instance, the recording of information such as may be contained in a "simple sentence" would have to tie down only one "transprintable" neuron. Thus, if one were to expose an individual to a simple sentence every four seconds, twenty-four hours a day, and if on each such occasion one would

tie down one transprintable neuron, then one would tie down just about 10^9 neurons over a period of one hundred years. This is about one tenth of the number of neurons believed to be contained in the human brain.

THE END

Hold for proof Not for release.

Preprint of an article scheduled to appear in the June issue of the Proceedings of the National Academy of Sciences.

April 23, 1964

On Memory and Recall* - Part I

by Leo Szilard

The Salk Institute for Biological Studies
La Jolla, California

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

The Efficacy of a Synapse Bridging Two Neurons

The neural network models which we shall be using here are based on excitatory neurons, as well as inhibitory neurons of one particular kind, the kind having post-synaptic inhibitory action.

neuron. If such an excitatory neuron sends a volley of nerve impulses to this synapse, then a certain quantity of an excitatory "transmitter substance" is released in the vicinity of the pre-synaptic membrane which diffuses across a gap --the synaptic cleft --into the post-synaptic neuron and raises the level of excitation of that neuron by a certain amount. We shall designate this excitatory transmitter substance as "acetylcholine" --in quotes. The "acetylcholine" which diffuses into the post-synaptic neuron is destroyed, in the vicinity of the post-synaptic membrane, by an enzyme which we shall designate as "choline esterase" --in quotes.

The rate at which "acetylcholine" is released in the vicinity of the pre-synaptic membrane is a function of the frequency of the nerve impulses which reach the synapse and we shall designate this rate as the "signal intensity". For the sake of

*This work was supported by a research grant, administered by The University of Chicago, of the General Medical ~~Division~~ of the National Institutes of Health.

simplicity, we shall assume that the "signal intensity" is for all synapses the same function of the frequency of the nerve impulses which are fed into the synapse.

The rate at which "acetylcholine" is destroyed in the post-synaptic neuron is proportional to the product of the concentration of "acetylcholine" and the concentration of the enzyme "choline esterase" in the vicinity of the post-synaptic membrane. Therefore, if, at a given point in time, nerve impulses of a certain frequency begin to arrive at a synapse, the "acetylcholine" concentration will begin to rise and will asymptotically approach a limit in the vicinity of the post-synaptic membrane, which is proportional to the "signal intensity" and inversely proportional to the concentration of "choline esterase", prevailing in the vicinity of the post-synaptic membrane. The "acetylcholine" concentration which is asymptotically approached at the post-synaptic membrane constitutes the "excitatory input", which is received from the synapse by the post-synaptic neuron. On this basis we may then say that, for any given "signal intensity," the excitatory input received from a given synapse by the post-synaptic neuron is inversely proportional to the "choline esterase" concentration prevailing in the vicinity of the post-synaptic membrane of that synapse.

We assume that the enzyme "choline esterase" is inactivated at the post-synaptic membrane in different synapses at different rates and that this rate of inactivation is determined by the chemical specificities of the two neurons which are bridged by the synapse. We shall assume, for the sake of simplicity, however, that the enzyme "choline esterase" is produced at the same rate in all excitatory neurons.

We designate as the "efficacy" of the synapse the "excitatory input" which a post-synaptic neuron receives from that synapse, per unit of "signal intensity". On the basis of the above assumptions we may then say that the efficacy of the synapse is inversely proportional to the rate at which "choline esterase" is inactivated at the post-synaptic membrane which, in turn, is determined by the chemical specificities of the two neurons which are bridged by the synapse.

The Rate of Inactivation of "Choline esterase":

We assume that neurons which differ from each other in their response-specificity contain a different set of certain specific proteins in their cell membrane. We shall refer hereafter to these proteins as the "specific membrane proteins".

We postulate that to each "specific membrane protein", there exists a complementary specific membrane protein and that a specific membrane protein molecule can combine with its complementary counterpart, just as an antibody molecule can combine with a molecule of its antigen. Accordingly, two complementary specific membrane proteins may behave as if they were, so to speak, each other's antibodies, as well as each other's antigens.

When an antibody molecule combines with an antigen molecule it undergoes an allosteric transition and an antibody molecule, when it is thus "dimerized", can bind complement. We assume that quite similarly a molecule of a "specific membrane protein", when it combines with its complementary counterpart, undergoes an allosteric transition and, when it is thus "dimerized", it can bind --and not only bind but also inactivate -- the enzyme "choline esterase".

The gap (synaptic cleft) between the pre-synaptic membrane and the post-synaptic membrane is estimated to be about 200 Å wide. According to the notions here adopted, there must be, however, a number of places within the active zones of the two synaptic membranes at which this gap is narrowed down, so that the pre-synaptic membrane and the post-synaptic membrane are in physical contact. We assume that at such a point of contact, a molecule of a "specific membrane protein", located in the post-synaptic membrane, can "dimerize" across the synaptic gap, with its complementary counterpart, located in the pre-synaptic membrane. The number of such "dimers", contained within the active zone of the synaptic membranes, would then determine the rate at which the enzyme "choline esterase" is inactivated at the post-synaptic membrane.

Let us now consider two neurons, A and B, which are bridged by a synapse. Neuron A is characterized by a set (a) of specific membrane proteins, which are present in its cell membrane and neuron B is characterized by another set, (b). We shall designate as the "overlap number" of these two neurons the number of specific membrane proteins contained within the set (a) which have their complementary counterpart contained within the set (b) [or vice versa].

From this overlap number we may compute the "efficacy" of the synapse which bridges these two neurons. In order to simplify this computation we shall assume that the area of the active zone of the synaptic membrane is the same for all synapses, and also assume, that the concentration of each "specific membrane protein" in the cell membrane is the same for any given neuron. On the basis of these simplifying assumptions, we may then say that the number of "dimers" contained within the active zone of the membrane of a synapse, which bridges neuron A and neuron B, is determined either by the ratio of the "overlap number" to the total number of specific membrane proteins of neuron A, or by the ratio of the "overlap number" to the total number of specific membrane proteins of neuron B --whichever ratio is smaller. We shall designate the smaller one of these two ratios as the "overlap fraction" of the neurons A and B. Accordingly, we may then say that the "efficacy" of a synapse bridging two neurons is proportional to the overlap fraction of the two neurons. This is the first fundamental postulate of our model.

We assume that the same holds true also for the synapses of our 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 neuron.

The Transprinting of Neurons:

We divide neurons of the Central Nervous System into two broad classes: the "congenitally-determined" neurons and the "memory" neurons. We designate the neurons

which attain their full chemical specificity of their cell membrane during the development of the individual (mostly during embryonal life and at the latest during the early post-natal period) as "congenitally-determined" neurons. If all the neurons of the Central Nervous System were of this sort then the individual would not be able to learn and his behavior would be wholly governed by the inborn reflexes. According to the notions here adopted, an adult can learn, and recall what he has learned, because his Central Nervous System contains "memory neurons" and each of these can, once in a lifetime, acquire an additional set of specific membrane proteins, through a process which we designate as "transprinting".

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. 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.(1)

(1) H.S. Anker, Nature 188 938 1960
Leo Szilard Proc. Nat. Acad. Sc. 46, 293-302 1960

If several such pre-synaptic neurons fire simultaneously with the memory neuron, then the memory neuron will on such an occasion acquire the sets of specific membrane proteins which are complementary to the sets of all of these pre-synaptic neurons.(1) This is the process of transprinting. Its occurrence as an "all or none" process, constitutes our second fundamental postulate.

We shall refer to memory neurons before they are transprinted, as "transprintable" neurons and, thereafter, we shall refer to them as "transprinted" neurons. "Transprinted" neurons resemble "congenitally-determined" neurons in that they may transprint "transprintable" neurons.

If a neuron participates in the transprinting of a "transprintable" neuron then we may expect this neuron and the "transprinted" neuron to have a large overlap fraction and, accordingly, we may expect synapses bridging these two neurons to have a high efficacy.

The Conditioned Response

In order to illustrate how transprinting may take place, we shall use as an example, the classical (Pavlovian) conditioning of the salivary reflex of the dog.(2) We shall indicate on this occasion, however, only rather sketchily what takes place during conditioning.

When "food" is introduced into the mouth of a dog, the dog responds with salivation. This is the inborn, or unconditioned, response. Let us now expose the dog to a compound stimulus which has an auditory, as well as a visual, component and let us -- before the compound stimulus is turned off --place "food" into the mouth of the dog. If, after several such "conditioning exposures", the dog is then presented for the first time with the compound stimulus, unreinforced on this occasion by the introduction of "food" into its mouth, the dog may be expected to salivate. This is the conditioned response.

---

(1) No neuron may, however, incorporate into its cell membrane the complementary counterpart of a specific membrane protein which its cell membrane already contains.

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

We assume that there is a neuron F in the Central Nervous System, characterized by the set (f), which preferentially responds to the stimulus of "food in the mouth". Moreover, we shall assume, in particular, that the signal to which the neuron F responds is the "onset" of this stimulus. As shown in Figure 1, the neuron F is connected through a synapse to an Effector neuron, which innervates the salivary gland. This Effector neuron is characterized by the set $(\bar{f})$, where $(\bar{f})$ denotes a set of specific membrane proteins which is complementary to the set (f). Because the overlap fraction of the neuron A and the Effector neuron is one, the synapses which bridge these two neurons have a high efficacy. Therefore, placing "food" into the dog's mouth may be expected to cause the dog to salivate.

In order to account for the conditioned response, we postulate the existence of a number of groups of transprintable neurons E. Each of these groups may consist of several hundred neurons E, all of which have the following in common. The neuron F contacts through a synapse each of the neurons E and in turn each neuron E contacts, through a synapse, an inter-neuron $F\bar{I}$ [characterized by the set (f) + $(\bar{i})$] which in turn contacts, through a synapse, the Effector neuron.

Until something happens which is "significant" from the point of view of the salivary reflex, all the transprintable neurons E are repressed, because they are inhibited by signals which are continuously being sent out by the inhibitory neuron $\bar{E}^*$ [characterized by the set $(\bar{e})$]. This inhibition is assumed to be strong enough to prevent a transprintable neuron E to fire, even if it should receive a substantial aggregate "excitatory input", because the overlap fraction of the inhibitory neuron $\bar{E}^*$ and of the transprintable neuron E is one. It should be noted, however, that after the neuron E is transprinted and acquires a set of specific membrane proteins which is composed of a large number of such proteins, then its overlap fraction with the inhibitory neuron $\bar{E}^*$ is reduced by a substantial factor and the efficacy of the synapse bridging the two neurons is also reduced by the same factor. Accordingly, such a

"transprinted" neuron E may be caused to fire in spite of receiving inhibitory signals from the neuron $\bar{E}^*$.

The transprintable neurons E get de-repressed if the inhibitory neuron $\bar{E}^*$ is inhibited by signals emanating from a neural network designated as the "Derepressor". This will happen if the Derepressor sends out signals which are sufficiently strong to excite the inhibitory inter-neuron E**, which in turn will inhibit the inhibitory neuron $\bar{E}^*$, The Derepressor network may receive an input signal from the neuron F and it may also receive an input signal, via the inter-neuron $F\bar{I}$, from neurons E. These two input signals counteract each other 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. In our second paper we shall describe a very simple "neural network" which would function in this fashion.

As will be seen later, the Derepressor network may be expected to send out strong signals if "food" is introduced into the mouth of an unconditioned dog and to send out strong signals also when a dog, whose salivary reflex has been fully conditioned to a certain stimulus, is presented with that stimulus, without having, on this occasion, "food" placed into its mouth. The Derepressor network will not send out signals, however, if the fully conditioned dog is presented with the correct stimulus and "food" is placed into its mouth. Accordingly, no additional neurons E would be transprinted as the result of such "routine exposures".

It is probably generally true that a sensory experience is recorded only if there is "significance" attached to that experience. In our model of the conditioned salivary reflex there is "significance" attached to the stimulus "food in the mouth" for an unconditioned (or not fully conditioned) dog, while for a fully conditioned dog, there is "significance" attached to the conditioned stimulus, but only if that stimulus is not accompanied by the signal "food in the mouth".

We shall try to indicate next in what manner a conditioned salivary response may be established to a compound stimulus which has a visual and an auditory component. To this end we assume that in the Central Nervous System there is a neuron $A\bar{E}$ which responds preferentially to the auditory component of the compound stimulus, and another neuron $V\bar{E}$ which responds preferentially to the visual component. These two neurons are characterized by the sets $(a) + (\bar{e})$ and $(v) + (\bar{e})$ respectively. We assume that the number of different specific membrane proteins contained in the neurons $A\bar{E}$ and $V\bar{E}$, which we designate by $n(A\bar{E})$ and $n(V\bar{E})$, respectively, are large compared to the number of different specific membrane proteins contained in the transprintable neuron E, which we designate by $n(E)$. Accordingly, we have: $n(A\bar{E}) > n(E)$, and $n(V\bar{E}) > n(E)$.

We assume that, out of a group of several hundred neurons E, a certain fraction is contacted through a synapse by the neuron $A\bar{E}$, a certain fraction is contacted through a synapse by the neuron $V\bar{E}$ and a certain fraction is contacted by both the neuron $A\bar{E}$, as well as the neuron $V\bar{E}$. Because the neurons $A\bar{E}$, as well as $V\bar{E}$ have an appreciable --even though small --overlap fraction with the transprintable neurons E, we may assume that if either of these two neurons fires, or if both of them fire, at a time when the transprintable neurons E are de-repressed, then one or more transprintable neurons will fire also and will on that occasion be transprinted by the neurons $A\bar{E}$ or $V\bar{E}$ or both. If, at the same time, the neuron F fires also, then these same neurons E will be transprinted also by the neuron F. The firing of neuron F alone would, however, not cause the neurons E to fire, even when the neurons E happen to be de-repressed, because the neurons F and the transprintable neurons E have zero overlap.

If the unconditioned dog is for the first time exposed to the compound stimulus, and at the same time "food" is introduced into its mouth, the Derepressor network will send out a strong signal and one, or several, of the transprintable neurons E will be caused to fire. These neurons E will then be transprinted, on this occasion, with the the sets $(\bar{f})$, $(\bar{a})$ and $(\bar{v})$. If this dog is exposed for the second time to the compound

stimulus and at the same time "food" is introduced into its mouth, then the signal sent out by the Derepressor network will be somewhat weaker than the first time. This is so because the neurons E which have been transprinted, at the time of the first conditioning exposure, and which will be excited at the time of the second exposure, have a large overlap fraction with the inter-neuron $F\bar{I}$ and will, therefore, send a signal to the Derepressor network which counteracts the signal received by this network from the neuron F. As the conditioning process is continued and the dog is repeatedly subjected to such conditioning exposures, the neurons E which are transprinted with the sets $(\bar{f})$, $(\bar{a})$ and $(\bar{v})$ will increase in number. Finally, the Derepressor network will no longer send out a signal when the dog is exposed to the compound stimulus and at the same time "food" is introduced into its mouth. Such a dog is then fully conditioned and continuing the conditioning exposures would not make the conditioning any deeper.

Let us now expose such a fully conditioned dog to the compound stimulus, unreinforced on this occasion by the introduction of "food" into its mouth. The neurons E which have been transprinted during the previous conditioning exposures, with the set $(\bar{f})$, as well as the sets $(\bar{a})$ and $(\bar{v})$ will now be caused to fire. Because of the substantial overlap of the transprinted neurons E, which contain the set $(\bar{f})$, with the inter-neuron $F\bar{I}$, the firing of the neurons E will lead to the firing of the inter-neuron $F\bar{I}$ and this in turn will lead to the firing of the Effector neuron. Accordingly, on the occasion of this unreinforced exposure of the dog to the compound stimulus, the dog will salivate. This is the conditioned response.

Incidentally, on this occasion when the inter-neuron $F\bar{I}$ fires, it will cause the Derepressor network to send out a strong signal because this network does not receive on this occasion a signal from the neuron F. Accordingly, on this occasion, one or more neurons E will get transprinted with the sets $(\bar{a})$ or $(\bar{v})$ or both, but none of them will be transprinted with the set $(\bar{f})$. Therefore, if the dog is repeatedly exposed

the compound stimulus in such a fashion, i.e. without reinforcement, then the number of neurons E which are transprinted with the sets ($\bar{a}$) or ($\bar{v}$) or both, but not with the set ($\bar{f}$), will increase on each such occasion. The overlap fraction of these transprinted neurons E with the inter-neurons $F\bar{I}$ is zero and, therefore, the excitation of these transprinted neurons E would not contribute to the excitation of the Effector neuron. Their activation would, however, contribute to the excitation of the inhibitory neurons $I\bar{E}$, with which the neurons E have an appreciable overlap. This is the reason why the accumulation of neurons E which are transprinted with ($\bar{a}$) or ($\bar{v}$) or both, but not with ($\bar{f}$), will extinguish the previously established conditioned response to the compound stimulus.

Note to the first model

One more thing needs to be said at this point: It seems to be a fact that if we establish a conditioned salivary response in the dog to a compound stimulus, which has an auditory as well as a visual component, and if we subsequently extinguish the response, say to the visual component, we thereby automatically extinguish the conditioned response to the auditory component also. It can be shown that, in order to account for this fact, we must assume in our model that the Central Nervous System contains, in addition to a number of neurons E which are characterized by the set (e), about an equal number of neurons $\bar{E}$ which are characterized by the complementary set ($\bar{e}$), and that the neurons E, characterized by one of these two sets, must contact through synapses the neuron $\bar{E}$, characterized by the complementary set (and vice versa). Presumably, this would mean that quite generally neurons characterized by complementary sets of specific membrane proteins must be present in about equal numbers in the Central Nervous System of each individual.

The Second Model:

We may escape this complication (if a complication it is), by assuming that every specific membrane protein is complementary to itself. According to this alternative second model, any set of specific membrane proteins is then identical with the complementary set of specific membrane proteins. Thus we may write (e) in place of ($\bar{e}$) and (f) in

place of ($\bar{f}$), etc. Accordingly, in our second model, the overlap number of two sets is then defined as the number of specific membrane proteins which the two sets have in common and when transprinting takes place, the "transprinted" neuron incorporates the sets of specific membrane proteins of the transprinting neurons. Whatever functions neural networks, of the kind represented in Figure 1, would be capable of fulfilling on the basis of our first model, they would fulfill on the basis of our second model also, and the remainder of our discussion will be couched in terms of this second model, rather than the first one,

The Orderliness of the Inborn Neural Code:

According to the notions here adopted, we assume that two neurons in the Central Nervous System, which preferentially respond to two different sensory stimuli that "resemble" each other, must have a large overlap number. We assume that in the code of the congenitally-determined neurons there is an orderly transition to smaller and smaller overlap numbers, as we go from one neuron to other neurons which differ from it more and more in their response-specificity. If it were otherwise, our model could not account for the phenomenon of the "generalization" of stimuli in the conditioned salivary reflex of the dog, first described by Pavlov.

Postcript:

If our two fundamental postulates are correct, then it ought to be possible to devise a neural network which would fully account for the phenomena exhibited by the conditioned responses of the autonomous nervous system. (The network described by Figure 1 represents a first attempt in this direction). If one wanted to see, however, whether higher mental functions could be explained on the basis of our two fundamental postulates, then one would first have to invent adequate neural networks. Thus, if one wanted to see whether one could explain on this basis the mental functions which man is capable of performing, but the primates are not, one would perhaps have to invent the very same networks which are contained in the brain of man, but not in the

brain of the primates. Clearly, this would be no mean task.

The "mental capacity" of suitable neural network models, operating on the basis of our two fundamental postulates, might be very high. For instance, the recording of information such as may be contained in a "simple sentence" would have to tie down only one "transprintable" neuron. Thus, if one were to expose an individual to a simple sentence every four seconds, twenty-four hours a day, and if, on each such occasion, one would tie down one "transprintable" neuron, then one would tie down just about 10^9 neurons over a period of one hundred years. This is about one tenth of the number of neurons believed to be contained in the human brain.

THE END

Figure 1: Excitatory neurons are represented by circles and inhibitory neurons are represented by double circles. Excitatory synapses are represented by simple arrows, except if they belong to neurons which are capable of transprinting, in which case they are represented by double arrows. Inhibitory synapses are represented by arrows with a crossbar. The transprintable neuron E is represented by a dotted circle.

by phagocytes and such recognition

that it would very well be all or none process. This may or may not be true and discrimination between self and non-self may or may not ..... and fagocytes may or may not be able to discriminate between self and non-self. But, however, that may be and whether fagocytes are or are not able to discriminate between self and non-self, we shall postulate that such a recognition of self and non-self plays a major role in the nervous system. We shall not assume, however, this to be an all or none process, rather we shall attribute to the degree of selfness a major role, We shall not assume that this is a qualitative all or none process, rather we shall assume that it is a quantitative phenomenon, in which the degree of selfness plays a determining role. Moreover, we shall present a biochemical model which could account for this phenomenon.

Roughly speaking, our postulate says that if a neuron contributes to the excitation or inhibition of another neuron its contribution is proportional --all other things being equal -- to the "overlap" of the "chemical specificity" of the two neurons. We propose to formulate further below much more precisely just what we mean by "chemical specificity" and what we mean by "overlap" and we also propose to present on that occasion a biochemical model which could account for this postulate. The second major concept relates to the supposed ability of an antigen when injected into a rabbit to bring about an irreversible change in certain lymphotic cells and to induce these cells to "diffferentiate" in the sense that as a result of the exposure to the antigen these cells will thereafter produce a specific antibody to the antigen at a high rate. This concept of a "post-natal induced differentiation" which irreversibly alters the chemical specificity of the cell was introduced by me in a paper which appeared in 1960.

We know very little about how differentiation of cells comes about during embryonal development. There is a general belief that the differentiation of cells within a tissue may, in some way, be induced by cells of another tissue, provided they are adjacent to and presumably in physical contact with each other.

Differentiation has 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 throusandfold by adding lactose to the growth medium and there is reason to think that lactose is a 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 a somatic cell undergoes differentiation during embyonal development, there appear in the cell a number of ~~new~~ proteins raised to a high level of concentration and thereafter the cell, as well as its descendents, will contain these proteins at a high concentration.

In order to account for this phenomenon of "persistence" --which is absent in enzyme induction in bacteria --one may assume that in differentiation once the concentration of such a protein molecule is raised in a somatic cell above a certain threshold thereafter the rate of production of this protein molecule remains high. This implies the operation of some sort of locking mechanism. It implies that a specific protein molecule must in a sense be able to act as its own inducer; it could act as its own inducer either directly or in some indirect manner.

In a paper I published in 1960 I presented a biochemical mechanism which could explain how an antigen which penetrates into a lymphatic cell raises the concentration of an antibody molecule which is specific for this antigen above the threshhold, beyond which a locking mechanism would go into effect. The antigen would thus induce an irreversible differentiation in a lymphatic cell so that after the cell had been exposed to the antigen the concentration of the corresponding antibody would thereafter be maintained at a high level. If this were correct then the lymphotic system would process a certain kind of memory and this memory could account for the abundant production of the antibody if the antigen is injected into a pre-immunized rabbit for a second time (the socalled secondary response). Taking his departure from these considerations, H.S. Anker suggested in a Letter to Nature that a similar biochemical mechanism could conceivably account for long-term memory in the central nervous system. Irrespective of whether or not antigens are capable of inducing this kind of differentiation in the lymphotic cells of the adult, I propose to postulate here that induced differentiation takes place in the central nervous system of the adult and is responsibile for long-term memory in the central nervous system.

Roughly speaking, our second postulate says that there are neurons in the central nervous system of the adult, which have not reached their final chemical specificity when they have undergone differentiation during embryonal development or in the early postnatal period. Such neurons which have remained in a certain sense plastic, can in certain circumstances be induced to undergo further differentiation and to assume their final chemical specificity by other neurons which contributes to their excitation. It is assumed that such induced differentiation takes place over a period of a few minutes and that thereafter the plastic neuron having attained its final chemical specificity ceases to be plastic. It is assumed that such an induced differentiation

Such a neuron which has remained, so to apeak, plastic can under certain circumstances be induced by other neurons which contribute to its excitation to undergo further differentiation and thereby to assume its final chemical specificity. It is assumed that such induced differentiation takes place only if the excitation level of the plastic neuron reaches the point where it begins to fire volleys into its axon, that such differentiation takes place within a period of a few minutes and that thereafter the plastic neuron, having attained its final chemical specificity, ceases to be plastic. We shall designate the process that takes place during such an induced differentiation as a process of transprinting. Transprinting consists in the memory neuron acquiring some of the chemical specificities of some of the neurons which contributed to its excitation at the time of induced differentiation. As a result of such transprinting the final chemical specificity of the memory neuron may have a substantial overlap with the chemical specificity of some other neurons which contribute to its excitation at the time of the induced differentiation.

Thus, roughly speaking, our second postulate says that a neuron in the central nervous system of the adult, which is not as yet fully differentiated, may be transprinted during the lifetime of the adult by neurons which have attained their final chemical specificity and on such an occasion these neurons confer their final chemical specificity on the memory neuron. We propose further below to formulate more precisely just when and how a memory neuron is induced to differentiate but we shall not be able to give a detailed biochemical mechanism for this process.

We will take it for granted that after a memory neuron acquires its final chemical specificity this chemical specificity will persist by virtue of the same kind of locking mechanism which operates in the neurons which have attained their final chemical specificity on the memory neuron.

We propose further below to formulate more precisely just what and how a memory neuron is induced to differentiate but we shall not be able to give a detailed biochemical mechanism for this process.

We will take it for granted that after a memory neuron acquires its final chemical specificity this chemical specificity will persist by virtue of the same kind of locking mechanism which operates in the neurons which have attained their final chemical specificity during embryonal development or in the early postnatal period. This, however, does not tell us in what manner neurons which have attained their final chemical specificity can induce differentiation in neurons which have as yet remained plastic.

The phenomenon of enzyme induction does not provide us in this regard with any real clue, still we know in the case of enzyme induction in the case of bacteria at least this much. Where the inducer locates is present in the growth medium the gene which is specific for galactosidase produces an RNA molecule --the messenger RNA which is specific for galactosidase. The base sequence of this messenger RNA determines the amino acid sequence of the corresponding polypeptide chain and four of these polypeptide chains combine to form the enzyme galactosidase.

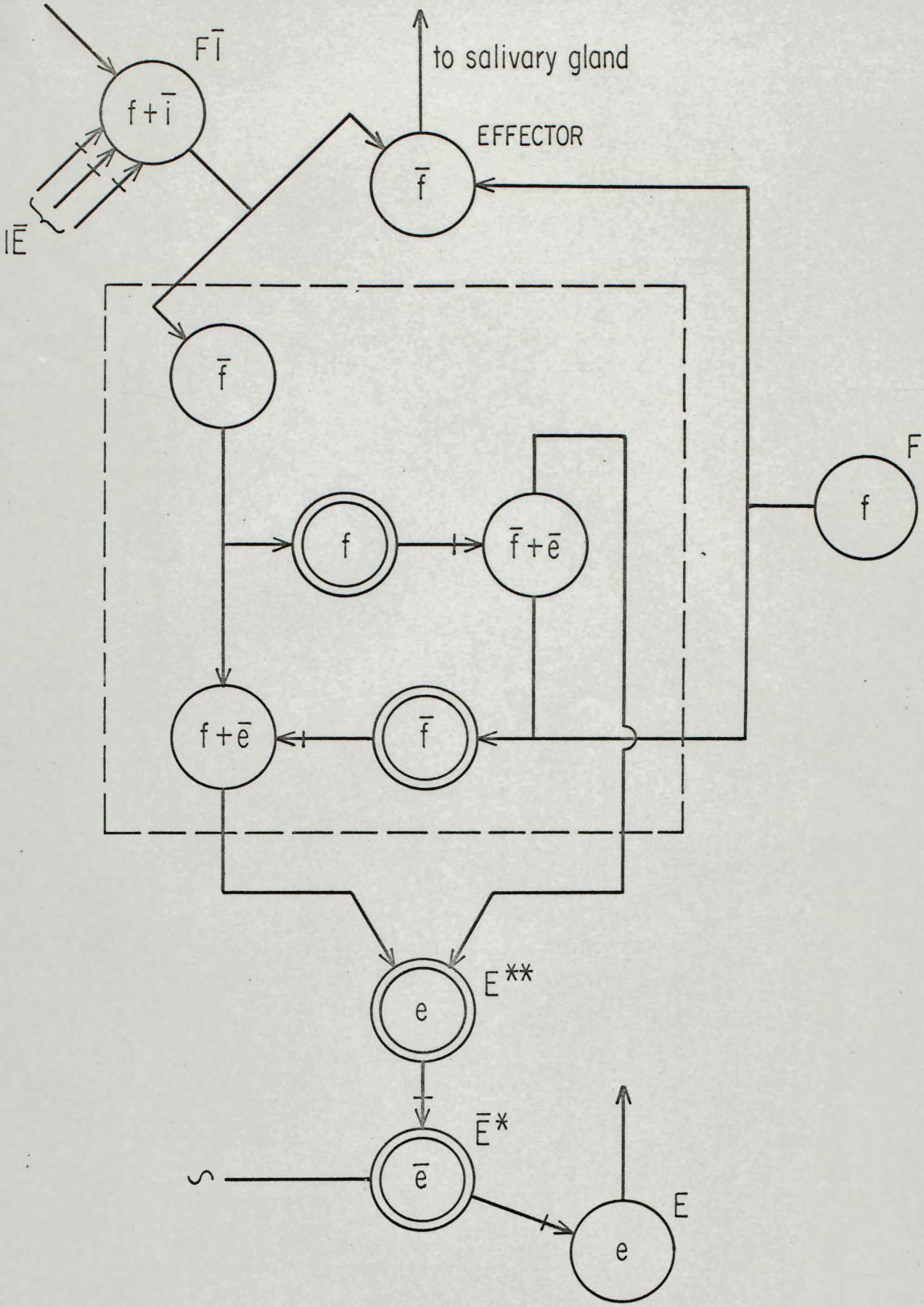
By analogy, we may surmise that in a neuron which has attained its final chemical specificity and in which a set of n different protein molecules are maintained at a high level of concentration, the n corresponding messenger RNA molecules are produced at a high rate. We do not know, however, what role these specific protein molecules themselves or other messenger RNA molecules, may play in the locking mechanism which is responsible for the persistence of these high levels.

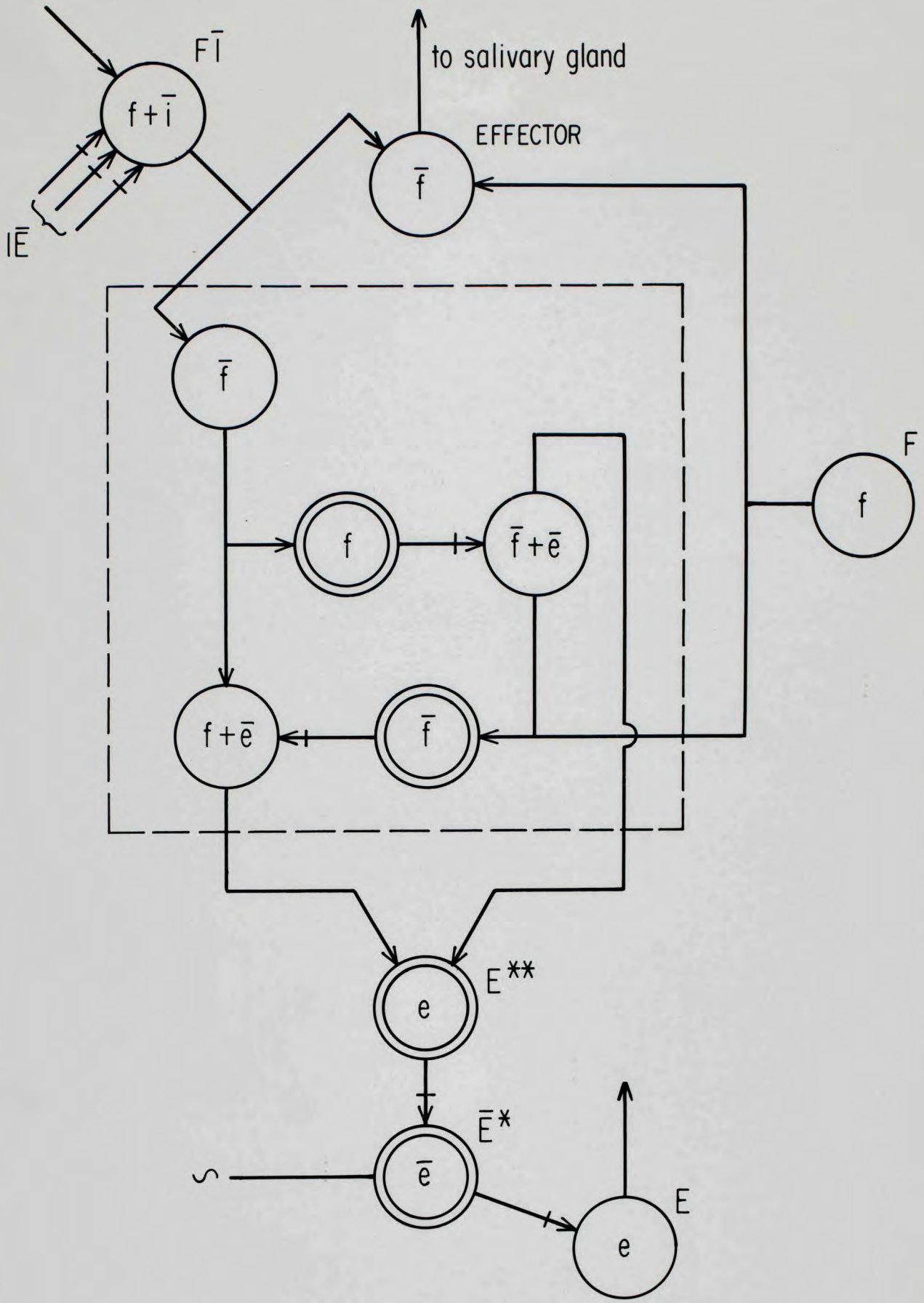
The best we can do in these circumstances is to assume that there is a class of compounds which plays a key role in differentiation that takes

place during the embryonal development and in the ensuing persistence of the chemical specificity of the neurons which have attained their final chemical specificity during embryonal development and that each of these key compounds is maintained at a high level of concentration after differentiation has taken place. We cannot say, however, whether the key compound is a protein molecule, an RNA molecule or some other kind of molecule, nor can we exclude the possibility that the key compound might be the specific protein molecule itself. On this basis we may then say that one cell may induce the differentiation of another cell with which it is in physical contact if, for one reason or another, the membrane of another cell which has not reached its final chemical specificity and with which it is in physical contact. If something happens that renders the membranes of both cells permeable to all relevant key compounds and provided these cells remain permeable for a sufficiently long period of time, a few minutes perhaps, so that there is enough time to permit the diffusion of the key compounds from the fully differentiated cell into the cell that has as yet remained plastic.

The description of the biochemical processes involved in transprinting will in these circumstances have to remain rather fragmentary, yet we propose to formulate further below, with some precision, the circumstances in which transprinting may take place and shall postulate with equal precision the chemical specifities that are conferred on the memory neuron as a result of transprinting.

Figure 1: Excitatory neurons are represented by circles and inhibitory neurons are represented by double circles. Excitatory synapses are represented by simple arrows, except if they belong to neurons which are capable of transprinting, in which case they are represented by double arrows. Inhibitory synapses are represented by arrows with a crossbar. The transprintable neuron E is represented by a dotted circle.


F$\bar{I}$
f+$\bar{i}$
I$\bar{E}$
to salivary gland
EFFECTOR
$\bar{f}$
$\bar{f}$
f
$\bar{f}$+$\bar{e}$
F
f
f+$\bar{e}$
$\bar{f}$
E**
e
$\bar{E}$*
$\bar{e}$
E
e


F$\bar{I}$
f+$\bar{i}$
I$\bar{E}$
to salivary gland
EFFECTOR
$\bar{f}$
$\bar{f}$
f
$\bar{f}$+$\bar{e}$
F
f
f+$\bar{e}$
$\bar{f}$
E**
e
$\bar{E}$*
$\bar{e}$
E
e

02307 NAS KB 50 Gal 1
10 on 12 QA-8a 5-5-64 Grewe 4
June PNAS Physiology: Szilard

*ON MEMORY AND RECALL- I**

By Leo Szilard

THE SALK INSTITUTE FOR BIOLOGICAL STUDIES, LA JOLLA, CALIFORNIA

Communicated April 27, 1964

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

The Efficacy of a Synapse Bridging Two Neurons.—Our neural network models involve excitatory neurons and inhibitory neurons (of the kind which exert a postsynaptic inhibitory effect).

Let us consider an excitatory neuron which contacts through a synapse another neuron. If such an excitatory neuron sends a volley of nerve impulses to this synapse, then a certain quantity of an excitatory "transmitter substance" is released in the vicinity of the presynaptic membrane which diffuses across a gap—the synaptic cleft—into the postsynaptic neuron and raises the level of excitation of that neuron by a certain amount. We shall designate this excitatory transmitter substance as "acetylcholine." The "acetylcholine" which diffuses into the postsynaptic neuron is destroyed, in the vicinity of the postsynaptic membrane, by an enzyme which we shall designate as "choline esterase."

The rate at which "acetylcholine" is released in the vicinity of the presynaptic membrane is a function of the frequency of the nerve impulses which reach the synapse, and we shall designate this rate as the "signal intensity." For the sake of simplicity, we shall assume that the signal intensity is for all synapses the same function of the frequency of the nerve impulses which are fed into the synapse.

The rate at which "acetylcholine" is destroyed in the postsynaptic neuron is proportional to the product of the concentration of "acetylcholine" and the concentration of the enzyme "choline esterase" in the vicinity of the postsynaptic membrane. Therefore, if, at a given point in time, nerve impulses of a certain frequency begin to arrive at a synapse, the "acetylcholine" concentration will begin to rise and will asymptotically approach a limit in the vicinity of the postsynaptic membrane, which is proportional to the signal intensity and inversely proportional to the concentration of "choline esterase," prevailing in the vicinity of the postsynaptic membrane. The "acetylcholine" concentration which is asymptotically approached at the postsynaptic membrane constitutes the "excitatory input," which is received from the synapse by the postsynaptic neuron. On this basis we may then say that, for any given signal intensity, the excitatory input received from a given synapse by the postsynaptic neuron is inversely proportional to the "choline esterase" concentration prevailing in the vicinity of the postsynaptic membrane of that synapse.

We assume that the enzyme "choline esterase" is inactivated at the postsynaptic membrane in different synapses at different rates and that this rate of inactivation is determined by the chemical specificities of the two neurons which are bridged by the synapse. We shall assume, for the sake of simplicity, however, *that the enzyme "choline esterase" is produced at the same rate in all excitatory neurons.*

We designate as the efficacy of the synapse the excitatory input which a postsynaptic neuron receives from that synapse, *per unit of signal intensity.* On the basis of the above assumptions we may then say that the efficacy of the synapse is inversely proportional to the rate at which "choline esterase" is inactivated at the postsynaptic membrane which in turn is determined by the chemical specificities of the two neurons which are bridged by the synapse.

The Rate of Inactivation of "Choline Esterase."—We assume that neurons which differ from each other in their response-specificity contain a different set of certain specific proteins in their cell membrane. We shall refer hereafter to these proteins as the "specific membrane proteins."

We postulate that to each specific membrane protein, there exists a complementary specific membrane protein and that a specific membrane protein molecule can combine with its complementary counterpart, just as an antibody molecule can combine with a molecule of its antigen. Accordingly, two complementary specific membrane proteins may behave as if they were, so to speak, each other's antibodies, as well as each other's antigens.

When an antibody molecule combines with an antigen molecule, it undergoes an allosteric transition, and an antibody molecule, when it is thus "dimerized," can bind complement. We assume that quite similarly a molecule of a specific membrane protein, when it combines with its complementary counterpart, undergoes an allosteric transition, and, when it is thus "dimerized," it can bind—and not only bind but also inactivate—the enzyme "choline esterase."

The gap (synaptic cleft) between the presynaptic membrane and the postsynaptic membrane is estimated to be about 200 Å wide. According to the notions here adopted, there must be, however, a number of places within the active zones of the two synaptic membranes at which this gap is narrowed down, so that the presynaptic membrane and the postsynaptic membrane are in physical contact. We assume that at such a point of contact, a molecule of a specific membrane protein, located in the postsynaptic membrane, can "dimerize" across the synaptic gap, with its complementary counterpart, located in the presynaptic membrane. ~~*The number of such "dimers," contained within the active zone of the synaptic membranes, would then determine the rate at which the enzyme "choline esterase" is inactivated at the postsynaptic membrane.*~~

Let us now consider two neurons, A and B, which are bridged by a synapse. Neuron A is characterized by a set (*a*) of specific membrane proteins, which are present in its cell membrane, and neuron B is characterized by another set, (*b*). *We shall designate as the "overlap number" of these two neurons the number of specific membrane proteins contained within the set* (*a*) *which have their complementary counterpart contained within the set* (*b*) [*or vice versa*].

From this overlap number we may compute the efficacy of the synapse which bridges these two neurons. In order to simplify this computation we shall assume that the area of the active zone of the synaptic membrane is the same for all synapses, and also assume that the concentration of each specific membrane protein in the cell membrane is the same for any given neuron. On the basis of these simplifying assumptions, we may then say that the number of "dimers" contained within the active zone of the membrane of a synapse, which bridges neuron A neuron B, is determined either by the ratio of the overlap number to the total number of specific membrane proteins of neuron A, or by the ratio of the overlap number to the total number of specific membrane proteins of neuron B—*whichever ratio is smaller.* We shall designate *the smaller* one of these two ratios as the "overlap fraction" of the neurons A and B. Accordingly, we may then say that *the efficacy of a synapse bridging two neurons is proportional to the overlap fraction of the two neurons. This is the first fundamental postulate of our model.*

We assume that the same holds true also for the synapses of our inhibitory neurons, except that in this case the "transmitter substance" which diffuses across the synaptic gap into the postsynaptic neuron lowers, rather than raises, the level of excitation of the postsynaptic neuron.

The Transprinting of Neurons.—We divide neurons of the central nervous system into two broad classes: the "congenitally determined" neurons and the "memory" neurons. We designate neurons which attain their full chemical specificity of their cell membrane during the development of the individual (mostly during embryonal life and at the latest during the early postnatal period) as "congenitally determined neurons." If all the neurons of the central nervous system were of this sort, then the individual would not be able to learn and his behavior would be wholly governed by inborn reflexes. According to the notions here adopted, an adult can learn, and recall what he has learned, because his central nervous system contains memory neurons and each of these can, once in his lifetime, acquire an additional set of specific membrane proteins, through a process which we designate as "transprinting."

2C1(82

02307 NAS KB 50 Gal 2
10 on 12 QA-8a 5-5-64 Grewe 5
June PNAS Physiologg: Szilard

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 postsynaptic membranes become permeable for specific membrane proteins. We further assume that if, on such an occasion, a specific membrane protein penetrates into the memory neuron from a presynaptic neuron, then 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.

We postulate that there is a class of congenitally determined neurons which are capable of participating in the transprinting of a memory neuron. We assume that if such a congenitally determined neuron fires, then those parts of its cell membrane (covering the boutons of the branch fibers of its axon) which form the active zones of the presynaptic membranes become permeable for its specific membrane proteins.

Let us now consider what happens if a congenitally determined neuron of this sort, which contacts a memory neuron through a synapse, fires "simultaneously," with this memory neuron, so that for a period of time both the presynaptic and postsynaptic membranes are permeable for the specific membrane proteins. If this takes place, then the specific membrane proteins of the presynaptic neuron will penetrate into the memory neuron and induce in the memory neuron their complementary counterparts. If several such presynaptic neurons fire simultaneously with the memory neuron, *then the memory neuron will on such an occasion acquire the sets of specific membrane proteins which are complementary to the sets of all of these presynaptic neurons. This is the process of transprinting. Its occurrence as an "all or none" process constitutes our second fundamental postulate.*

We shall refer to memory neurons before they are transprinted as transprintable neurons and thereafter we shall refer to them as transprinted neurons. Like congenitally determined neurons, transprinted neurons may also participate in the transprinting of a transprintable neuron.

If a neuron participates in the transprinting of a transprintable neuron, then we may expect this neuron and the transprinted neuron to have *a large overlap* fraction and, accordingly, *we may expect synapses bridging these two neurons to have a high efficacy.*

The Conditioned Response.—In order to illustrate how transprinting may take place, we shall use as an example the classical (Pavolovian) conditioning of the salivary reflex of the dog. We shall indicate on this occasion, however, only rather sketchily what takes place during conditioning.

When food is introduced into the mouth of a dog, the dog responds with salivation. This is the inborn, or unconditioned, response. Let us now expose the dog to a compound stimulus which has an auditory as well as a visual component, and let us—before the compound stimulus is turned off—place food into the mouth of the dog. If, after several such "conditioning exposures," the dog is then presented for the first time with the compound stimulus, unreinforced on this occasion by the introduction of food into its mouth, the dog may be expected to salivate. This is the conditioned response.

We assume that there is a neuron F in the central nervous system, characterized by the set (f), which preferentially responds to the stimulus of food in the mouth. *Moreover, we shall assume in particular that the signal to which the neuron F responds is the onset of this stimulus.* As shown in Figure 1, the neuron F is connected through a synapse to an effector neuron, which innervates the salivary gland. This effector neuron is characterized by the set $(\bar{f})$, where $(\bar{f})$ denotes a set of specific membrane proteins which is complementary to the set (f). Because the overlap fraction of the neuron A and the effector neuron is *one*, the synapses which bridge these two neurons have a high efficacy. Therefore, placing food into the dog's mouth may be expected to cause the dog to salivate.

In order to account for the conditioned response, we postulate the existence of a number of groups of transprintable neurons E. Each of these groups may consist of several hundred neurons E, all of which have the following in common: The neuron F contacts through a synapse each of the neurons E and in turn each neuron E contacts, through a synapse, an interneuron $F\bar{I}$ [characterized by the set $(f) + (\bar{i})$] which in turn contacts, through a synapse, the effector neuron.

Until something happens which is "significant" from the point of view of the salivary reflex, all the transprintable neurons E are repressed, because they are inhibited by signals which are continuously being sent out by the inhibitory neuron $\bar{E}^*$ [characterized by the set $(\bar{e})$]. This inhibition is assumed to be strong enough to prevent a transprintable neuron E from firing, even if it should receive a substantial aggregate "excitatory input," because the overlap fraction of the inhibitory neuron $\bar{E}^*$ and of the transprintable neuron E is *one*. It should be noted, however, that after the neuron E is transprinted and acquires a set of specific membrane proteins which is composed of a large number of such proteins, then its overlap fraction with the inhibitory neuron $\bar{E}^*$ is reduced by a substantial factor and the efficacy of the synapse bridging the two neurons is also reduced by the same factor. Accordingly, such a "transprinted" neuron E may be caused to fire in spite of receiving inhibitory signals from the neuron $\bar{E}^*$.

The transprintable neurons E get *derepressed* if the inhibitory neuron $\bar{E}^*$ is inhibited by signals emanating from a neural network designated as the "derepressor." This will happen if the derepressor sends out signals which are sufficiently strong to excite the inhibitory interneuron E^{**}, which in turn will inhibit the inhibitory neuron $\bar{E}^*$. The derepressor network may receive an input signal from the neuron F and it may also receive an input signal, via the interneuron $F\bar{I}$, from neurons E. These two input signals counteract each other 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.* In our second paper we shall describe a very simple "neural network" which would function in this fashion.

As will be seen later, the derepressor network may be expected to send out strong signals if food is introduced into the mouth of an unconditioned dog and to send out strong signals also when a dog, whose salivary reflex has been fully conditioned to a certain stimulus, is presented with that stimulus, without having, on this occasion, food placed into its mouth. The depressor network will not send out signals, however, if the fully conditioned dog is presented with the correct stimulus and food is placed into its mouth. Accordingly, no additional neurons would be transprinted as the result of such "routine exposures."

It is probably generally true that a sensory experience is recorded only if there is "significance" attached to that experience. In our model of the conditioned salivary reflex there is significance attached to the stimulus "food in the mouth" for an unconditioned (or not fully conditioned) dog, while for a fully conditioned dog, there is significance attached to the conditioned stimulus, but only if that stimulus is *not* accompanied by the signal "food in the mouth."

We shall try to indicate next in what manner a conditioned salivary response may be established to a compound stimulus which has a visual and an auditory component. To this end we assume that in the central nervous system there is a neuron $A\bar{E}$ which responds preferentially to the auditory component of the compound stimulus, and another neuron $V\bar{E}$ which responds preferentially to the visual component. These two neurons are characterized by the sets $(a) + (\bar{e})$ and $(v) + (\bar{e})$, respectively. We assume that the number of different specific membrane proteins contained in the neurons $A\bar{E}$ and $V\bar{E}$, which we designate by $n(A\bar{E})$ and $n(V\bar{E})$, respectively, are large compared to the number of different specific membrane proteins contained in the transprintable neuron E, which we designate by $n(E)$. Accordingly, we have: $n(A\bar{E}) > n(E)$, and $n(V\bar{E}) > n(E)$.

We assume that, out of a group of several hundred neurons E, a certain fraction is contacted through a synapse by the neuron $A\bar{E}$, a certain fraction is contacted through a synapse by the neuron $V\bar{E}$, and a certain fraction is contacted by both the neuron $A\bar{E}$, as well as the neuron $V\bar{E}$. Because the neuron $A\bar{E}$ as well as $V\bar{E}$ have an appreciable—even though small—overlap fraction with the transprintable neurons E, we may assume that if either of these two neurons fires, or if both of them fire, at a time when the transprintable neurons E are derepressed, then one or more transprintable neurons will fire also and will on that occasion be transprinted by the neurons $A\bar{E}$ or $V\bar{E}$ or both. If at the same time the neuron F fires also,

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. 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 a certain lymphatic cells of the rabbit.(1)

Szilard
2307-163.50
Mrs June

D4 ϕ 66

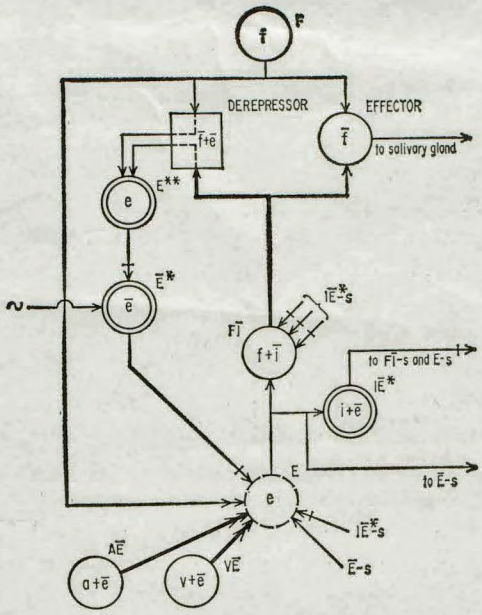


FIG. 1.—Excitatory neurons are represented by circles, and inhibitory neurons are represented by double circles. Excitatory synapses are represented by simple arrows, except if they belong to neurons which are capable of transprinting, in which case they are represented by double arrows. Inhibitory synapses are represented by arrows with a crossbar. The transprintable neuron E is represented by a dotted circle.

then these same neurons E will be transprinted also by the neuron F. The firing of neuron F alone would, however, not cause the neurons E to fire, even when the neurons E happen to be derepressed, because the neurons F and the transprintable neurons E have zero overlap.

If the unconditioned dog is for the first time exposed to the compound stimulus, and at the same time food is introduced into its mouth, the derepressor network will send out a strong signal and one, or several, of the transprintable neurons E will be caused to fire. These neurons E will then be transprinted, on this occasion, with the sets $(\bar{f})$, $(\bar{a})$, and $(\bar{v})$. If this dog is exposed, for the second time, to the compound stimulus and at the same time food is introduced into its mouth, then the signal sent out by the derepressor network will be somewhat weaker than the first time. This is so because the neurons E which have been transprinted at the time of the first conditioning exposure, and which will be excited at the time of the second exposure, have a large overlap fraction with the interneuron $F\bar{I}$ and will, therefore, send a signal to the derepressor network, which counteracts the signal received by this network from the neuron F. As the conditioning process is continued and the dog is repeatedly subjected to such "conditioning" exposures, the neurons E which are transprinted with the sets $(\bar{f})$, $(\bar{a})$, and $(\bar{v})$ will increase in number. Finally, the derepressor network will no longer send out a signal when the dog is exposed to the compound stimulus and at the same time "food" is introduced into its mouth. Such a dog is then fully conditioned and continuing the conditioning exposures would not make the conditioning any deeper.

Let us now expose such a fully conditioned dog to the compound stimulus, unreinforced on this occasion by the introduction of food into its mouth. The neurons E which have been transprinted during the previous conditioning exposures with the set $(\bar{f})$ as well as the sets $(\bar{a})$ and $(\bar{v})$ will now be caused to fire. Because of the substantial overlap of the transprinted neurons E, which contain the set $(\bar{f})$, with the interneuron $F\bar{I}$, the firing of the neurons E will lead to the firing of the interneuron $F\bar{I}$ and this in turn will lead to the firing of the effector neuron. Accordingly, on the occasion of this exposure of the dog to the compound stimulus, the dog will salivate. This is the conditioned response.

Incidentally, on this occasion, when the interneuron $F\bar{I}$ fires, it will cause the derepressor network to send out a strong signal because this network does not on this occasion receive a signal from the neuron F. Accordingly, on this occasion, one or more neurons E will get transprinted with the sets $(\bar{a})$ or $(\bar{v})$ or both, but none of them will be transprinted with the set $(\bar{f})$. Therefore, if the dog is repeatedly exposed to the compound stimulus in such a fashion, i.e., without reinforcement, then the number of neurons E which are transprinted with the sets $(\bar{a})$ or $(\bar{v})$ or both, but not with the set $(\bar{f})$, will increase on each such occasion. The overlap fraction of these transprinted neurons E with the interneurons $F\bar{I}$ is zero and therefore the excitation of these transprinted neurons E would not contribute to the excitation of the effector neuron. Their activation would, however, contribute to the excitation of the inhibitory neurons $I\bar{E}$, with which the neurons E have an appreciable overlap. This is the reason why the accumulation of neurons E which are transprinted with $(\bar{a})$ or $(\bar{v})$ or both, but not with $(\bar{f})$, will extinguish the previously established conditioned response to the compound stimulus.

— One more thing needs to be said at this point: It seems to be a fact that if we establish a conditioned ~~slivary~~ response in the dog to a compound stimulus, which has an auditory as well as a visual component, and if we subsequently extinguish the response, say to the visual component, we thereby automatically extinguish the conditioned response to the auditory component also. It can be shown that, in order to account for this fact, we must assume that the central nervous system contains, in addition to a number of neurons E which are characterized by the set (e), about an equal number of neurons $\bar{E}$ which are characterized by the complementary set $(\bar{e})$, and that the neurons $\bar{E}$, characterized by one of these two sets, must contact through synapses the neuron E characterized by the complemntary set (and vice versa). Presumably, this would mean that quite generally neurons characterized by ~~complementary~~ sets of specific membrane proteins must be present in about equal numbers in the central nervous system of the individual.

The Second Model.—We may escape this complication (if a complication it is), by assuming that *every specific membrane protein is complementary to itself.* According to this second model, any set of specific membrane proteins is then identical with the complementary set of specific membrane proteins. Thus we may write (e) in place of $(\bar{e})$ and (f) in place of $(\bar{f})$, etc. Accordingly, in our second model, *the overlap number of two sets is then defined as the number of specific membrane proteins which the two sets have in common, and when transprinting takes place, the transprinted neuron incorporates the sets of specific membrane proteins of the transprinting neurons.* Whatever functions neural networks, ~~of the kind~~ represented in Figure 1, would be capable of fulfilling on the basis of our first model, ~~they~~ would fulfill on the basis of our second model also, and the remainder of our discussion will be couched in terms of this second model rather than the first one.

The Orderliness of the Inborn Neural Code.—According to the notions here adopted, we assume that two neurons in the central nervous system, which preferentially respond to two different sensory stimuli that "resemble" each other, must have a large overlap number. We assume that in the code of the congenitally determined neurons there is an orderly transition to smaller and smaller overlap numbers, as we go from one neuron to other neurons which differ from it more and more in their response-specificity. If it were otherwise, our model could not account for the phenomenon of "the generalization of stimuli" in the conditioned salivary reflex of the dog, first described by Pavlov.

Postcript.—If our two fundamental postulates are correct, then it ought to be possible to devise a neural network which would fully account for the phenomena exhibited by the conditioned responses of the autonomic nervous sytem. (The network described by Figure 1 represents ~~only~~ a first attempt in this direction.) If one wanted to see, however, whether higher mental functions could be explained on the basis of our two fundamental postulates, then one would first have to invent adequate neural networks. Thus, if one wanted to see whether one could explain on this basis the mental functions which man is capable of performing, but the primates are not, one would perhaps have to invent the very same networks which are contained in the brain of man, but not in the brain of the primates. Clearly, this would be no mean tkas.

The "mental capacity" of suitable neural network models, operating on the basis of our two fundamental postulates, might be very high. For instance the recording of information such as may be contained in a simple sentence would have to tie down only one "transprintable" neuron. Thus, if one were to expose an individual to a simple sentence every 4 seconds, 24 hours a day, and if, on each such occasion, one would tie down one transprintable neuron, then one would tie down just about 10^9 neurons over a period of 100 years. This is one tenth of the number of neurons believed to be contained in the human brain.

* This work was supported by a research grant, administered by The University of Chicago, of the General Medical Division of the National Institutes of Health.

† No neuron may, however, incorporate into its cell membrane the complementary counterpart of a specific membrane protein which its cell membrane already contains.

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

H.S. Anker, Nature, 188 938, 1960.
Leo Szilard, Proc. Nat. Acad. Sc. 46, pp. 293-302, 1960.