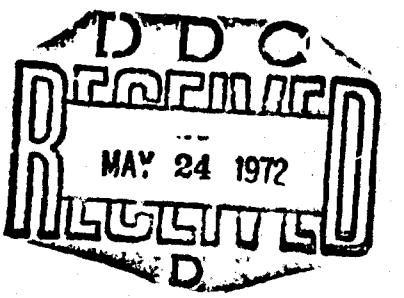
AFCRL-72-0164
3 MARCH 1972
SPECIAL REPORTS, NO. 133

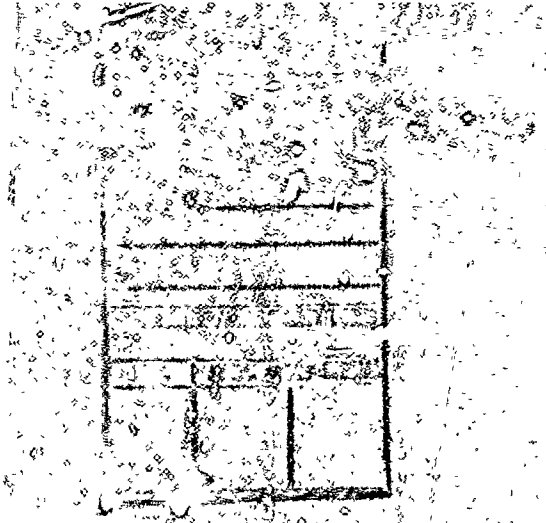
AIR FORCE CAMBRIDGE RESEARCH LABORATORIES
L. G. HANSCOM FIELD, BEDFORD, MASSACHUSETTS

Brain Function and Adaptive Systems - A Heterostatic Theory

A. HARRY KLOPF



AIR FORCE SYSTEMS COMMAND
United States Air Force



Unclassified
Security Classification

DOCUMENT CONTROL DATA - R&D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author)	2a. REPORT SECURITY CLASSIFICATION
Air Force Cambridge Research Laboratories (LR) L. G. Hanscom Field Bedford, Massachusetts 01730	Unclassified 2b. GROUP

3. REPORT TITLE

BRAIN FUNCTION AND ADAPTIVE SYSTEMS – A HETEROSTATIC THEORY

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

Scientific. Interim.

5. AUTHOR(S) (First name, middle initial, last name)

A. Harry Klopf

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
3 March 1972	77	128

8a. CONTRACT OR GRANT NO.	9a. ORIGINATOR'S REPORT NUMBER(S)
b. PROJECT, TASK, WORK UNIT NOS. 5632-02-01	AFCRL-72-0164
c. DOD ELEMENT 61102F	9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report) Special Reports, No. 133
d. DOD SUBELEMENT 681305	

10. DISTRIBUTION STATEMENT

Approved for public release; distribution unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
TECH, OTHER	Air Force Cambridge Research Laboratories (LR) L. G. Hanscom Field Bedford, Massachusetts 01730

13. ABSTRACT

A new theory of intelligent adaptive systems is proposed. The theory provides a single unifying framework within which the neurophysiological, psychological, and sociological properties of living adaptive systems can be understood. Furthermore, the theory offers a new basis for the synthesis of machines possessing adaptive intelligence.

The proposed theory is of a heterostatic type. That is to say, it is a theory which assumes that living adaptive systems seek, as their primary goal, a maximal condition (heterostasis), rather than assuming that the primary goal is a steady-state condition (homeostasis). It is further assumed that the heterostatic nature of animals, including man, derives from the heterostatic nature of neurons. The postulate that the neuron is a heterostat (that is, a maximizer) is a generalization of a more specific postulate, namely, that the neuron is a hedonist. This latter postulate is interpreted strictly in terms of physical variables, yielding the heterostatic neuronal model that is the basis for the detailed development of the theory.

The theory is shown to be consistent with and capable of suggesting underlying mechanisms for neurophysiological and psychological phenomena. The implications of the theory for the mind-body problem and for the global organization of the brain are considered. Some sociological aspects of the theory are examined. Finally, the relationship of the theory to cybernetic research on neural nets and heuristic programs is explored.

DD FORM 1 NOV 65 1473

Unclassified
Security Classification

Unclassified

Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Adaptive networks						
Cybernetics						
Artificial intelligence						
Neurophysiology						
Heterostasis						
Neural nets						
Psychology						
Brain theory						
Hedonism						
Mind-body problem						
Homeostasis						
Soci gy						
Neuronal model						
Heuristic programs						

Unclassified

Security Classification

AFCRL-72-0164
3 MARCH 1972
SPECIAL REPORTS, NO. 133

DATA SCIENCES LABORATORY PROJECT 5632

AIR FORCE CAMBRIDGE RESEARCH LABORATORIES

L. G. HANSCOM FIELD, BEDFORD, MASSACHUSETTS

Brain Function and Adaptive Systems - A Heterostatic Theory

A. HARRY KLOPF

The research reported herein was accomplished while the author held a National Research Council Postdoctoral Resident Research Associateship supported by the Air Force Systems Command, United States Air Force.

AIR FORCE SYSTEMS COMMAND
United States Air Force

Abstract

A new theory of intelligent adaptive systems is proposed. The theory provides a single unifying framework within which the neurophysiological, psychological, and sociological properties of living adaptive systems can be understood. Furthermore, the theory offers a new basis for the synthesis of machines possessing adaptive intelligence.

The proposed theory is of a heterostatic type. That is to say, it is a theory which assumes that living adaptive systems seek, as their primary goal, a maximal condition (heterostasis), rather than assuming that the primary goal is a steady state condition (homeostasis). It is further assumed that the heterostatic nature of animals, including man, derives from the heterostatic nature of neurons. The postulate that the neuron is a heterostat (that is, a maximizer) is a generalization of a more specific postulate, namely, that the neuron is a hedonist. This latter postulate is interpreted strictly in terms of physical variables, yielding the heterostatic neuronal model that is the basis for the detailed development of the theory.

The theory is shown to be consistent with and capable of suggesting underlying mechanisms for neurophysiological and psychological phenomena. The implications of the theory for the mind-body problem and for the global organization of the brain are considered. Some sociological aspects of the theory are examined. Finally, the relationship of the theory to cybernetic research on neural nets and heuristic programs is explored.

Contents

Illustrations

Tables

Brain Function and Adaptive Systems — a Heterostatic Theory

I. INTRODUCTION AND SUMMARY

A new theory of intelligent adaptive systems is proposed in this paper. It appears to be the first theory capable of providing a single unifying framework within which the neurophysiological, psychological, and sociological properties of living adaptive systems can be understood. Furthermore, the theory provides a new basis for the synthesis of machines possessing adaptive intelligence.

In this introductory section, the fundamental ideas underlying the theory are discussed and the principal conclusions are reviewed. The remaining sections of the paper are devoted to a more formal development of the theory.

The theory originated with the belief that the similarities between social systems and nervous systems might be much more important than their differences. Both social and nervous systems may be viewed as networks, each constructed basically out of a single type of element. In the case of nervous systems, the neuron constitutes the network element or building block; in the case of social systems, the network elements are people. In both systems, the elements receive inputs from many other elements and likewise send their outputs to many others. Thus, both systems possess connectivity patterns exhibiting substantial convergence and divergence. Beyond these simple structural parallels are more important similarities relating to information processing characteristics. Both social and

(Received for publication 3 March 1972)

nervous systems are adaptive; they acquire new forms of behavior as a function of experience. In both, memory and learning are distributed; information is not stored in a highly localized fashion. Both types of networks employ a redundant structure permitting them to function reliably despite the fact that individual elements are continually dying or malfunctioning. A measure of the plasticity of the systems is suggested by the fact that both can undergo extensive damage resulting in the permanent loss of large numbers of network elements (for example, severe head wounds in the case of nervous systems or large-scale bombing attacks during war in the case of social systems) and yet, after a period of time, lost functions can often be recovered by the surviving elements. Such an array of network properties is indeed remarkable. At least it is fair to say that most people find these properties impressive in the case of nervous systems. Social systems, on the other hand, exhibiting the same properties inspire less awe, because there is the feeling that such networks are at least partially understood.

Suppose it is true that we understand social systems, at least to some extent. Why not apply that understanding of human networks to an analysis of neural networks? While recognizing that the proposed analogy was a tenuous one, the decision was made to examine it further. What kinds of mechanisms could account for the adaptive powers of these systems? Where are the mechanisms to be located within the networks? Concerning the latter question, a plausible answer is that the adaptive mechanisms are completely localized within the individual elements, rather than that adaptation is an emergent global phenomenon appearing only in large assemblies of interacting elements. A corrollary to this belief is the notion that the pattern of interconnections within a network is derived from each element's individual actions, each element forming connections based on local circumstances. In the case of a social system, the network elements (people) can be observed as they continually evolve new patterns of communication. It seems likely that neurons might carry out a similar process in the brain, producing new connectivity patterns as a function of experience. Supporting this idea is the observation that the coding capacity of the genetic apparatus appears to be inadequate for the provision of a highly detailed innate specification of the interconnections among a thousand billion neurons.

Complete localization of the adaptive mechanism within the individual element makes sense from another point of view. The similarity between the evolution of the human brain and the evolution of a modern society is interesting. The primitive brain in the one case and simple villages in the other both seem to have simply grown (evolved) by layers without the older and more interior structures ever having to undergo sudden and massive reorganizations. The neuron and the man both appear to have the inherent capacity to make adjustments when their surroundings change, thus permitting an evolutionary process which, to some extent at least, simply adds additional network elements.

If the neuron and the man are indeed self-contained adaptive units, are their adaptive mechanisms complex or simple? It seems plausible that they could be highly complex in the case of a human being functioning as an adaptive element in a social system. However, is a neuron likely to embody highly complex adaptive machinery? Probably not. If it doesn't, then the existence of nervous systems functioning as powerful adaptive networks suggests that elements possessing simple adaptive mechanisms can ultimately yield networks exhibiting highly complex adaptive behavior. Which brings us to an important question: If the mechanism is simple in the case of neurons and brains, might it also in essence be simple in the case of people and societies? Pushing the speculation even further, might the adaptive mechanisms, in fact, be essentially one and the same in both systems, since the overall properties of both networks are so similar? It seemed to be worth testing as an assumption.

At this point, the case can be restated very simply. Two systems will be said to be equivalent ($\equiv$) if their adaptive mechanisms and the information processing characteristics growing out of these mechanisms are essentially similar. It is suggested that from this systems theoretic point of view, the following equivalence may be valid:

$$\text{NERVOUS SYSTEMS} \equiv \text{SOCIAL SYSTEMS} . \tag{1}$$

It is further conjectured that Equivalence (1) is a consequence of the fundamental equivalence of the following two systems:

$$\text{NEURON} \equiv \text{MAN} . \tag{2}$$

Equivalence (2) seems to offer no insights whatever into neuronal function until a philosophical notion is introduced into the analysis. Aristotle (384-322 B.C.) observed: "Happiness being found to be something final and self-sufficient, is the End at which all actions aim." Thus, a philosophical theory supplies a third equivalence:

$$\text{MAN} \equiv \text{HEDONIST} . \tag{3}$$

From Equivalences (2) and (3), a fourth equivalence is easily obtained and it provides the cornerstone of the theory to be proposed:

$$\text{NEURON} \equiv \text{HEDONIST} . \tag{4}$$

Unlike Equivalence (2), Equivalence (4) readily lends itself to an interpretation. Hedonism implies pleasurable and painful states, and there is a straightforward way of classifying neuronal states into two categories, these being the states of depolarization and hyperpolarization. Given the evident excitatory nature of pleasure and the inhibitory nature of pain, the following equivalences suggest themselves for the neuron:

DEPOLARIZATION ≡ PLEASURE (5)

HYPERPOLARIZATION ≡ PAIN . (6)

One implication of these equivalences is that a neuron will seek to obtain excitation and to avoid inhibition. Does such a statement have a simple mechanistic interpretation? It does if one assumes, as many brain researchers currently do (with experimental evidence now accumulating in favor of the assumption), that the effectiveness of a synapse in causing a neuron to fire is a variable quantity, altered as a function of experience. Variable synaptic transmittances are then assumed to be the repository of learning and memory. Granting this assumption, a simple neuronal adaptive mechanism can be proposed by utilizing Skinner's (1938) framework of operant conditioning, this time in conjunction with the neuron instead of the whole animal. The idea is this. After a neuron fires, it waits for a few hundred milliseconds or more to see how it will be affected by the action it has taken. If it experiences further depolarization within at most a second or so, it increases the effectiveness of the excitatory synapses that led to its firing in the first place, thereby increasing the probability that it will fire the next time that some fraction of these synapses is active. If, however, the action of firing is followed within at most a second or so with the experience of hyperpolarization, the neuron then increases the effectiveness of those inhibitory synapses that were active when it fired. In this way, the probability of responding again to the input configuration has been diminished. Thus, the neuron views excitation as positively reinforcing and inhibition as negatively reinforcing. A highly effective excitatory synapse, when active, "informs" the neuron that it should fire because, by doing so, it is likely to receive additional excitation. A highly effective inhibitory synapse, when active, "informs" the neuron that it had better not fire because, to do so, is likely to bring on additional inhibition. The effectiveness of a synapse, therefore, encodes a causal relation, providing predictive information concerning the consequences for the neuron if it fires when the synapse is active. It can be seen that the adaptive mechanism, over a period of time, will cause the neuron to behave so as to tend to maximize the amount of excitation and minimize the amount of inhibition being received.

The proposed neuronal model can be described in psychological terms. For a neuron, temporal and spatial configurations of active synapses represent conditioned stimuli (CS), firing represents a conditioned response (CR), and the excitation or inhibition that arrives for a limited period of time after firing constitutes the unconditioned stimulus (US). (If this is so, how does a neuron distinguish between an input configuration that represents a CS and one that represents a US? The answer may be that it doesn't. The neuronal US may simultaneously represent a CS with respect to signals that will arrive still later. At all times, neuronal inputs may be playing dual roles, representing conditioned stimuli with respect to near-future inputs and unconditioned stimuli with respect to recent-past inputs. This would permit an associative chaining of sequences of events.)

Given the less-than-rigorous nature of the reasoning employed up to this point, the implications of Equivalence (4) would not have been pursued as far as they have, had it not been for the support the idea receives from the experimental literature of neurophysiology and psychology. The observations obtained from neuronal and cortical polarization experiments and from the study of the mirror focus appear to be explained by the theory. Also, the theory provides explanations for the experimental results obtained in psychological studies of conditioning and related phenomena. For these reasons, it was decided that a more rigorous development of the theory should be undertaken. This development begins in Section 2. However, in order to provide the reader with some perspective beforehand, the main conclusions derived from the theory are now summarized:

1. The primary goal of animals, including man, is the achievement of a maximal condition, not the achievement of a steady-state condition. Animals are not homeostats, they are heterostats (a heterostat is defined to be a system that seeks a maximal condition — the condition itself will be termed heterostasis). The variable to be maximized is that of the amount of neuronal polarization being experienced. The amount of polarization is defined to be equal to the amount of depolarization (pleasure) minus the amount of hyperpolarization (pain). The heterostatic nature of animals derives from the heterostatic nature of neurons. In psychological terms, neurons are hedonists and thus the living systems they control are hedonists.

2. Nervous systems are so structured that homeostasis is a necessary (but not sufficient) condition for the maintenance of heterostasis. This explains why organisms vigorously pursue homeostasis even while it is not their primary goal. That homeostasis is a subgoal suggests that survival may not be as central a concern of living systems as has previously been assumed.

3. Experimental results obtained in neuronal and cortical polarization studies and the results obtained in the study of epileptic foci can be explained in terms of whether a depolarizing or hyperpolarizing bias is imposed on the

neurons involved. A depolarizing bias is positively reinforcing and causes the effectiveness of recently active excitatory synapses to increase. A hyperpolarizing bias is negatively reinforcing and causes the effectiveness of recently active inhibitory synapses to increase. Epilepsy is a progressive disease because established epileptic foci deliver polarizing biases to normal neural tissue. This tissue, in turn, undergoes adaptation in response to the imposed bias and becomes hyper- or hypoactive.

4. Habituation, dishabituation, classical conditioning, operant conditioning, and extinction are phenomena that can be understood in terms of the neuronal adaptive response to depolarizing or hyperpolarizing reinforcement. Also of importance in understanding the mechanism underlying habituation are feed-forward and recurrent inhibition as well as the existence of long-duration IPSP neurons. Habituation and extinction are fundamentally the same phenomena.

5. The solution to the mind-body problem is an identity theory. A neuron undergoing depolarization is elementary pleasure; a neuron undergoing hyperpolarization is elementary pain. The subjective event of the experience of pleasure or pain is identical to the objective event of neurons undergoing depolarization or hyperpolarization, respectively. Pleasure and pain provide the single bidirectional dimension necessary to analyse mental phenomena. Pleasure and pain are much more fundamental and broader concepts than previously assumed. Complex forms of pleasure and pain (that is, the full range of possible mental states) are one and the same as complex spatial configurations of depolarizing and hyperpolarizing neurons. Each configuration corresponds to a particular mental state. The neurons which constitute the dynamic configuration corresponding to the "mind" are those of the midbrain and thalamic reticular formation (MTRF). Our "mind" is aware of other neural structures only to the extent that their outputs impinge on the MTRF. Pleasurable mental states result when depolarizing neurons are more prevalent in the MTRF. Painful mental states result when hyperpolarizing neurons are more prevalent. The exact mental state depends upon the exact configuration of depolarizing and hyperpolarizing neurons within the MTRF. The mental states corresponding to feeling tone, sensation, and ideation result when the principal inputs to the MTRF are supplied by the limbic system and hypothalamus, sensory cortex, or nonsensory-nonmotor cortex, respectively.

6. The global organization of the brain can be understood in terms of three major subsystems: the midbrain and thalamic reticular formation (MTRF), the limbic system and hypothalamus (LSH), and the neocortex. The MTRF is the command and control center and the seat of conscious awareness. The MTRF seeks to obtain excitation and to avoid inhibition. Its sources for

both types of signal are the sensory and pain fibers, the LSH, and the neocortex. Sensory fibers are excitatory with respect to the MTRF; pain fibers are inhibitory with respect to the MTRF. The LSH, utilizing the innate mechanisms of its reinforcement centers, delivers generalized excitation or inhibition to the MTRF, depending on whether the MTRF's decisions are leading toward or away from homeostasis. Also, the innate mechanisms corresponding to drive centers permit the LSH to make some decisions concerning more specific actions to be taken to maintain homeostasis. The neocortex provides the input/output functions and memory for the MTRF. It preprocesses incoming information, provides storage, and elaborates on the MTRF's motor commands. Curiosity is a manifestation of the fundamental drive for depolarizing stimuli. (Novel stimuli must be sought because the habituation mechanism renders repetitive stimuli ineffective.) The attention mechanism and the facilitation of limited cortical areas by the MTRF are synonymous. Facilitation delivered by the MTRF to the cortex is positively reinforcing and we therefore remember that to which we attend, while not remembering that to which we do not attend. Recall occurs when the MTRF is driven by its variety of inputs into a state approximating one it has been in before. It then sends a signal configuration to the cortex like that which it sent out before. The result is that the MTRF facilitates cortical memories laid down during the previous experience. The MTRF becomes aware of these memories when the outputs of the facilitated cortical neurons reach the MTRF. Memory is that information acquired with cortical reinforcement supplied by the MTRF attention mechanism; learning is that information acquired with reinforcement supplied by sensory, pain fiber, or LSH activity.

7. Not only neurons, assemblies of neurons, and whole nervous systems are heterostats; the same is true for families, neighborhoods, cities, regions, and nations. The reason is always the same. In each case, the more fundamental elements out of which the given system is constructed are heterostats and therefore the nature of the total system is that of a heterostat.

8. Every aspect of neuronal function relevant to information processing is realizable in hardware form. The heterostat can therefore serve as a fundamental building block for the construction of intelligent adaptive systems. The development of such systems in the future may be limited by progress in (a) the miniaturization of circuit elements, (b) the reduction of power requirements and (c) the development of appropriate assembly techniques.

2. HETEROSTASIS

The analysis and synthesis of intelligent adaptive systems is the objective of the theory presented in this report. The fundamental postulate of the theory is discussed in this section. Sections 3, 4, and 5 test and further develop the theory, utilizing data from neurophysiology, psychology, and sociology. Cybernetic issues, in general, are considered in Section 6.

An adaptive system is one which modifies its internal structure as a function of experience such that, over a period of time, the system's behavior tends to become more appropriate relative to some criterion. Within the class of adaptive systems, of particular interest here will be the human brain and the nonliving "learning automata" that have been studied in recent years.

One general theory concerned with adaptive systems already exists. The theory is based on a concept proposed in 1859 by Claude Bernard. The concept is now referred to as homeostasis, a term introduced by Walter Cannon (1929). Homeostasis refers to the condition of a system in which a set of "essential variables" have assumed steady-state values compatible with the system's continued ability to function. Essential variables are defined by Ashby (1960, p. 42) to be those "which are closely related to survival and which are closely linked dynamically." A theory that evolved from the concept of homeostasis suggests that this condition is the goal of all animal behavior. Ashby (1960, p. 62) has concluded, for example, that homeostasis is the goal of "a great deal, if not all, of the normal human adult's behavior." Young (1966, p. 5) has stated a similar conclusion: "Brains are the computers of homeostats and the essence of homeostats is that they maintain a steady state. Put in another way, the most important fact about living things is that they remain alive."

It will be proposed in this paper that homeostasis is not the primary goal of living systems; rather, it is a secondary or subgoal. It is suggested that the primary goal is a condition which, following the example of Cannon (1929), will be termed heterostasis. An organism will be said to be in a condition of heterostasis with respect to a specific internal variable when that variable has been maximized. Heterostasis, as the term itself suggests, is not associated with a steady-state condition. In general, the internal variable to be maximized will have an upper limit that changes as environmental constraints change. Therefore, even if heterostasis is continuously maintained by an organism, a steady-state condition with respect to the internal variable will not result. Furthermore, maintenance of the condition of heterostasis is not necessary for survival, in contrast to the maintenance of homeostasis which is an essential condition for life. With respect to heterostasis, it will in fact be seen that living systems infrequently or never achieve the condition. For this reason, it is appropriate to define a heterostatic

system as one that seeks (moves toward) a maximal condition while not necessarily ever achieving it.

To some, it may appear paradoxical that heterostasis is hypothesized to be the primary goal of living systems, while it is further suggested that this condition is not essential to the maintenance of life. Homeostasis, on the other hand, is here hypothesized to be a secondary goal and yet it is essential to life. This apparent paradox will dissolve if one is careful not to assume that the maintenance of life is necessarily the primary goal of living systems. It will be seen later in the development of the theory that homeostasis is a necessary but not sufficient condition for the achievement of heterostasis. This explains why living systems, in seeking the primary goal of heterostasis, devote much time and energy to the maintenance of homeostasis.

Some further definitions are now possible. A system that utilizes feedback information in order to seek or maintain a condition of homeostasis has been termed a homeostat. A system that utilizes feedback information in order to seek or maintain a condition of heterostasis will be termed a heterostat. A system that seeks both goals will be classified in accordance with that goal which is primary.

We will begin the development of the theory by examining brain function for evidence of heterostasis. This examination will reveal that a wide variety of the neurophysiological, psychological, and sociological phenomena that have been cataloged for man and the lower animals can be understood in terms of a single fundamental postulate: neurons are adaptive heterostats. This postulate, in turn, suggests a reason for the generally disappointing performance of the learning automata that have come out of past cybernetic research: these systems have lacked adequate heterostatic adaptive mechanisms. The evidence relating to these propositions is reviewed in the next four sections.

3. NEUROPHYSIOLOGY

A neuronal model will be proposed based on the following assumptions: Neurons seek to maximize the amount of depolarization and to minimize the amount of hyperpolarization they are experiencing. A neuron accomplishes this by modifying the effectiveness of its synapses after impulse generation. If impulse generation is followed by further depolarization, the effectiveness of recently active excitatory synapses is increased. If hyperpolarization follows impulse generation, the effectiveness of recently active inhibitory synapses is increased. These assumptions will be more precisely defined below in Section 3.1. "Recently active" synapses are those that contributed the excitation or inhibition which was effective during the generation of the impulses. "To seek a goal" and "to converge,

by means of feedback, toward a particular system state" will be equivalent expressions in this paper. It can be seen that neural inputs in the proposed model serve a dual role: as stimuli relating to the production of action potentials and as positive and negative feedback signals relating to the regulation of synaptic transmittances. It will be shown that this model possesses the two basic properties required of a circuit element from which learning automata are to be constructed. Furthermore, the model offers a simple and consistent basis for understanding neuronal adaptive behavior.

3.1 The Neuronal Model

3.1.1 NEURONAL HETEROSTASIS

It will be helpful to define the model in mathematical terms. To this end, a neuron will be viewed as receiving n numbered synaptic inputs, each of which delivers a series of action potentials. The frequency of arrival of impulses at the i^{th} synapse will be represented as the input intensity, $f_i(t)$. This measure of frequency reflects the activity of the synaptic knob at time, t. Also associated with each synapse will be a weight, w_i, representing the synaptic transmittance or effectiveness (that is, the amplification factor) applied to the frequency, f_i. Weights are constrained to be positive in the case of excitatory inputs and negative in the case of inhibitory inputs. The weights are postulated to be the repository of learning and memory and therefore they vary with time according to the experience of the organism. Also, a weight value in this model reflects an input's effectiveness as a function of the location of the synapse on the soma or within the dendritic field.

The computation that a neuron performs in "deciding" whether to fire consists of a spatial and temporal summation of the weighted inputs followed by a thresholding operation. That is to say, the neuron generates an action potential only if the following relationship holds:

$$\sum_{i=1}^{n} w_i(t)\, f_i(t) \geq \theta(t_o) \tag{7}$$

where

n = the number of synaptic inputs,

$w_i(t)$ = the synaptic transmittance associated with the i^{th} input; positive and negative weights correspond to excitation and inhibition, respectively,

$f_i(t)$ = a frequency measure of the input intensity at the i^{th} synapse,

$\theta(t_o)$ = the neural threshold,

t = time,

t_o = the time elapsed since the generation of the last action potential.

In Inequality (7), the spatial variations in the effectiveness of the inputs are reflected in the weights. Temporal variations are incorporated into the measure of frequency employed in arriving at the value of $f_i(t)$. θ must be a function of t_o in order to account for the variation in the threshold during the absolute and relative refractory periods.

The variable that is maximized when a system achieves the condition of heterostasis is denoted by μ and termed the heterostatic variable. Neuronal heterostasis is defined as the condition in which the amount of polarization being experienced by a neuron is maximal, where depolarization and hyperpolarization represent the positive and negative components, respectively, of polarization. More precisely, a neuron is said to be in a condition of heterostasis for the time interval ranging from t to $t+\tau$ if its synaptic transmittances are such as to maximize the quantity, $\mu_1^{t,\ t+\tau}$:

$$\mu_1^{t,\ t+\tau} = D_{t,\ t+\tau} - H_{t,\ t+\tau} \tag{8}$$

$$= \int_t^{t+\tau} [v_p(t) - v_r]dt - \int_t^{t+\tau} -[v_n(t) - v_r]dt \tag{9}$$

$$= \int_t^{t+\tau} [v(t) - v_r]dt \tag{10}$$

where

$D_{t,\ t+\tau}$ = the amount of depolarization experienced during the interval, t to $t+\tau$,

$H_{t,\ t+\tau}$ = the amount of hyperpolarization experienced during the interval, t to $t+\tau$,

$v(t)$ = the potential difference across the neuronal membrane,

v_r = the resting potential of the neuron,

$v_p(t)$ = $v(t)$ if $v(t) \geq v_r$, otherwise $v_p(t) = v_r$,

$v_n(t)$ = $v(t)$ if $v(t) \leq v_r$, otherwise $v_n(t) = v_r$.

In Eq. (8), μ_1 is subscripted in order to distinguish this approach to the definition of neural heterostasis from another presented below.

Constraints imposed by the neuron's adaptive process (not yet described) make the maximization of μ_1 an unattainable goal, in general. The definition is offered, however, because it represents a useful way of specifying the condition towards which a neuron moves (within a stable environment) and in this sense does represent the neuronal goal. An alternative definition of neuronal heterostasis, suggesting a condition that is more nearly attainable, is that which defines the goal to be the maximization of μ_2:

$$\mu_2 = E[\sum_{i=1}^{n} w_i(t)\, f_i(t)] = E[\alpha(t) - \beta(t)] \qquad (11)$$

where

$$\alpha(t) = \sum_{i=1}^{m} w_i(t)\, f_i(t), \qquad (12)$$

$$\beta(t) = -\sum_{i=m+1}^{n} w_i(t)\, f_i(t), \qquad (13)$$

$[1 \leq i \leq m]$ = the range of the numbered excitatory synapses,

$[(m+1) \leq i \leq n]$ = the range of the numbered inhibitory synapses.

E[x] refers to the expected or average value of x, in this case computed for the set of all input configurations that the neuron may receive. If μ_2 is to be maximized, the weights, $w_i(t)$, must be adjusted such that the average difference between the amount of excitation, $\alpha(t)$, and the amount of inhibition, $\beta(t)$, that is received is maximized. To the extent that $\alpha(t)$ and $\beta(t)$ are independent variables, the following equation holds:

$$\max(\mu_2) = \max\left\{E[\alpha(t) - \beta(t)]\right\} = \max\left\{E[\alpha(t)]\right\} - \min\left\{E[\beta(t)]\right\}. \qquad (14)$$

Equation (14) is presented in order to point out that, to some extent, neurons may appear to be maximizing the amount of excitation and minimizing the amount of inhibition received, even though they are actually seeking to maximize the difference between these two quantities.

3.1.2 ADAPTATION

Having suggested that neurons seek to obtain depolarization and to avoid hyperpolarization, the question that must now be considered is whether there is an adaptive process that will produce such behavior. The answer is yes and, in fact, the required mechanism is a very simple one.

Let us consider two possible ways in which μ might be made to increase. A trivial way is to set the inhibitory weights equal to zero and the excitatory weights at their upper limits (which are assumed to be finite). Alternatively, with the proper adjustment of the weights, the firing of the neuron can be made to occur at such times as to result in the frequency of arrival of depolarizing input configurations being maximized and the frequency of arrival of hyperpolarizing input configurations being minimized. Only the latter possibility can provide a basis for learning (and survival). Therefore, the adaptive process to be proposed employs the second approach and excludes the first.

When a neuron fires, the output signal may be viewed as being fed back to the input side of the neuron via three channels (see Figure 1). One channel involves only the remainder of the nervous system and constitutes neural feedback. Another channel provides information from the remainder of the organism's internal environment. The third channel is the external environmental feedback loop. The number of individual feedback loops available to a neuron can be very large.

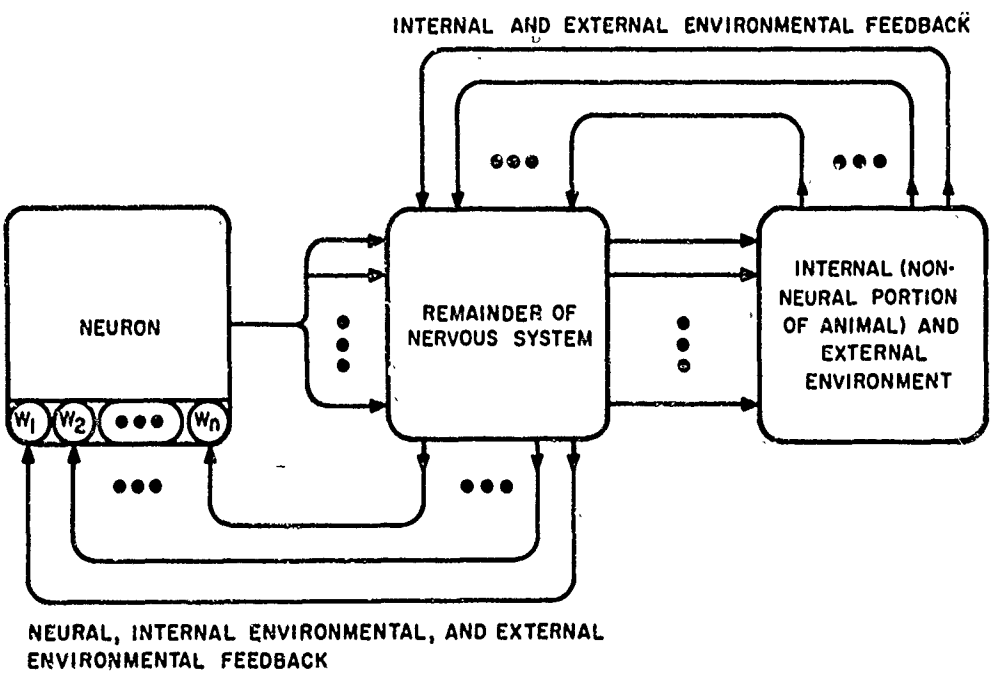


Figure 1. Relationship of a Neuron to the Remainder of the Nervous System and to the Internal and External Environments. w_1, w_2, ..., w_n = variable synaptic transmittances

A Purkinje cell, for example, receives about 100,000 excitatory synaptic inputs from the system of parallel fibers (Eccles, 1969a, p. 75). Large cortical pyramidal cells may receive more than 10,000 excitatory inputs (Eccles, 1966, p. 333). It is therefore possible for a neuron to receive a tremendous variety of input combinations, indicating that a neuron's view of the world is a highly varied and complex one.

It will be assumed that neurons utilize the feedback loops in "deciding" to which input configurations they will respond. Specifically, it is proposed that neurons will be more likely to respond again to stimulus configurations that led to firing if the neuronal response, when fed back, results in further depolarization. However, if the feedback arrives in the form of inhibition, the neuron eventually ceases to respond to the input configuration preceding the inhibition. The following postulated mechanism will produce these neural behavioral changes by altering synaptic transmittances according to three rules:

1. When a neuron fires, all of its excitatory and inhibitory synapses that were active during the summation of potentials leading to the response are eligible to undergo changes in their transmittances.

2. The transmittance of an eligible excitatory synapse increases if the generation of the action potential is followed by further depolarization, that is,

$$\left| \sum_{i=1}^{m} w_i(t)\, f_i(t) \right| > \left| \sum_{i=m+1}^{n} w_i(t)\, f_i(t) \right| \tag{15}$$

for a limited time interval after the response.

3. The transmittance of an eligible inhibitory synapse increases, that is, it becomes more negative, if the generation of an action potential is followed by hyperpolarization, that is,

$$\left| \sum_{i=1}^{m} w_i(t)\, f_i(t) \right| < \left| \sum_{i=m+1}^{n} w_i(t)\, f_i(t) \right| \tag{16}$$

for a limited time interval after the response.

This process whereby synaptic transmittances are altered so as to increase μ will be termed heterostatic adaptation.

It is of interest to define this adaptive process in terms of the psychological variables of operant conditioning, both for the purpose of further clarifying the mechanism and because of the relevance of a psychologically oriented definition to later considerations. In the interest of simplifying the discussion, incoming neuronal stimuli will be viewed as arriving at discrete time intervals. We can then discuss a stimulus, s_t, consisting of some configuration of excitatory and

inhibitory inputs, followed by the stimulus, s_{t+1}, etc. Heterostatic adaptation may then be viewed as follows. A neuronal input configuration, s_t, which causes the neuron to fire, will be termed a neural conditioned stimulus (NCS). The resulting action potential represents the neural conditioned response (NCR). The stimuli, s_{t+1}, thru s_{t+j}, received by the neuron for a limited time interval after the NCS, correspond to the neural unconditioned stimulus (NUS). The NUS is positively reinforcing if s_{t+1} thru s_{t+j} generally result in depolarization, and negatively reinforcing if these inputs generally result in hyperpolarization.

Having broadly outlined the process of heterostatic adaptation, speculation can now be offered concerning some of the detailed aspects of the mechanism. For example, extensive data that will be further reviewed in Section 3 indicates that, in Pavlovian conditioning, the most effective US is one that arrives approximately 400 ms after the CS (see review by Russell, 1966, p. 136). A US that arrives sooner or later than 400 ms after the CS has a reduced reinforcing effect. The reinforcing effect decreases to zero with (a) simultaneous occurrence of the CS and the US or (b) with a CS-US interval that exceeds a few seconds. It is hypothesized that these temporal relationships between the CS and US reflect similar temporal relationships between the NCS and the NUS.

It seems reasonable to expect that synapses mediating an NUS cannot undergo transmittance changes in response to that NUS, even if these synapses also mediated the NCS. Synapses that mediated only the NCS will undergo transmittance changes if the NUS (utilizing other synapses) follows within no more than a few seconds. With these assumptions, it can be seen that a synapse must be inactive for about 400 ms in order to be maximally reinforced. This model does not rule out the possibility of an input configuration to a neuron serving as an NUS relative to a preceding input configuration and serving as an NCS relative to a future input configuration. The proposed mechanism that prevents a synapse from being modified if it mediates both the NCS and NUS will be termed zerosetting.

The motivation for introducing zerosetting into the model is to prevent an input configuration from being reinforced simply by repetition. If reinforcement could occur due to repetition alone, behavioral patterns could then be established without regard for their consequences. If such an undesirable feature is to be avoided, it is concluded that neuronal reinforcement must result from the interaction of distinct subsets of synapses. It may turn out that zerosetting is accomplished by having specialized synapses that mediate only the NCS or the NUS, depending perhaps on whether the synapse is located on a dendrite or on the soma. If synapses are specialized to this extent, then the transmittances of NUS-mediating synapses might not have to be modifiable.

It is hypothesized that the magnitude of an adaptive weight change, Δw_i, is not only a function of the delay between the NCS and the various portions of the NUS but is also a function of the magnitude of the polarization occurring throughout the

effective reinforcement interval. Thus, mildly depolarizing or hyperpolarizing reinforcing signals accomplish lessor alterations in the synaptic transmittances than do stronger signals. This relationship makes it possible to explain the greater extent of learning and the more vivid memories that result when an animal is highly aroused. More will be said about this in Section 4.

It should be noted that since the NUS is distributed over a time interval measuring up to perhaps 4 seconds in duration (see review by Russell, 1966, p. 136), it will sometimes consist of a mix of depolarizing and hyperpolarizing input configurations. Whether an NUS positively or negatively reinforces, therefore, will depend on the relative amounts of each type of signal that are present and also on their arrival times during the 4-second interval.

A neuron must compute two functions; these will be designated as the I/O function and the adaptation or α-function. The former refers to the computation of the axonal output and the latter to the computation of the synaptic transmittance changes.

The extent of polarization of a neuron will be denoted by Σ. At the resting potential Σ will be defined to be equal to zero. Positive and negative values of Σ will correspond to depolarized and hyperpolarized neuronal states, respectively.

In the heterostatic model of the neuron, Σ plays a dual role. In the computation of the I/O function, the value of Σ relative to the neuronal threshold determines whether the neuron will fire. In the case of the α-function, the value of Σ determines whether the neuron is positively or negatively reinforced. Also, the amount of the change for a synaptic transmittance has been hypothesized above to be approximately proportional to the absolute value of Σ, that is,

$$\Delta w_i \cong k \left| \Sigma \right| \tag{17}$$

where k = constant.

The dual role of Σ is noted at this point because it suggests a characteristic to be observed in nervous systems. Namely, the computations of the I/O and the α-function should, at times, interfere with one another since both functions utilize the same parameter. When the reinforcement centers (to be discussed in Section 4) deliver substantial excitatory or inhibitory reinforcement to a neuron, this will alter Σ in a way that may be appropriate relative to the modification of the transmittances but may be inappropriate relative to the current I/O computations. This interference effect can indeed be observed with people. The experience of substantial positive or negative reinforcement can momentarily alter a person's performance significantly. For example, when people experience great disappointments, their immediate reaction may be quite inappropriate. Our hypothesis suggests that the negatively reinforcing signals have elevated neural thresholds by decreasing Σ. The experience of substantial positive reinforcement can render a person incapable

of complex information processing for a period of time. The lower neuronal thresholds are evidenced by hyperexcitability which is sometimes incompatible with activities requiring a high degree of concentration. In addition to observing these interference effects, we should also be able to observe their reduction over a period of time. This should occur because the effects, themselves, will be subject to reinforcement.

3.1.3 PROPERTIES OF THE MODEL

If a system learns through experience it does so by (1) acquiring a library of causal relationships and (2) acting on the basis of this stored information. It can be seen that the heterostatic neuron has these capabilities. Furthermore, the codes involved are very simple ones; the model assumes a binary code for data transmission and modifiable weights for storage. A high value for an excitatory weight indicates the desirability of a neuronal response when the associated synapse is active, since a response to that input in the past has led to further depolarization. A high (absolute) value for an inhibitory weight indicates the desirability of the neuron not responding when the associated synapse is active, since a response to that input in the past has led to hyperpolarization. High transmittances of excitatory synapses represent a record of those inputs that led to further depolarization. High transmittances of inhibitory synapses represent a record of those inputs that led to hyperpolarization. This record of causal relations is directly translated into a neuronal response each time an input configuration presents itself. Thus, the postulated model is seen to possess precisely the characteristics required for learning.

3.1.4 OTHER NEURONAL MODELS

Two general categories of connectionistic adaptive mechanisms have been considered by researchers in their efforts to account for neuronal plasticity. The earlier of these proposals suggested, as Eccles (1964) has noted, that "activation of synapses increases their efficacy by some enduring change in their fine structure (Tanzi 1893; Ramon y Cajal 1911; Konorski 1948, 1950; Hebb 1949; Tonnees 1949; Young 1951; Eccles 1953; Jung 1953; McIntyre 1953; Thorpe 1956)." Hebb (1949), in the development of his theory of cell assemblies, extensively explored the consequences of a variation of this assumption. Hebb assumed that a synapse which repeatedly contributed to the firing of a neuron became increasingly effective in causing that neuron to fire.

An alternative assumption and the one adopted in the present theory is that synaptic transmittance changes are contingent upon reinforcement signals received by a neuron after it fires. Cybernetic investigations by Minsky (1954) and Farley and Clark (1954) were the first tests of this assumption in synthesized networks. The potential of the model to explain the experimental results of conditioning

studies was recognized as one of its attractive features. Neuronal and cortical polarization studies relevant to this type of model will be discussed in Section 3.2 below. Relevant cybernetic research will be reviewed in Section 6.

3.2 Experimental Evidence

Neurophysiological data relating to the adaptive behavior of single or limited numbers of neurons will be reviewed to determine if the postulate of neuronal heterostasis is consistent with experimental results. Most of the evidence relating to large assemblies of neurons and whole nervous systems will be reviewed in Section 4. Much of the data considered here has been reviewed by Meldrum (1966) or in a collection of papers edited by Horn and Hinde (1970).

The heterostatic model of the neuron offers the advantage of immediately suggesting a well-defined experimental test. What is required is the application of a depolarizing or hyperpolarizing bias to a neuron, while noting what long-term changes, if any, occur with respect to the neuron's average frequency of firing. Fortunately, such experiments have been performed and are reviewed next.

3.2.1 NEURONAL AND CORTICAL POLARIZATION

A type of experiment that has been highly successful in producing effects resembling learning is that involving cortical polarization by means of small direct currents. Surface-positive polarization has a depolarizing effect on cortical neurons; surface-negative polarization has a hyperpolarizing effect (Bindman, Lippold, and Redfearn, 1964). According to the postulated neuronal model, if a neuron fires and depolarization follows, then the recently active excitatory synapses will increase in effectiveness. If hyperpolarization follows, the effectiveness of recently active inhibitory synapses will increase. We would, therefore, predict that surface-positive polarization of cortex should be positively reinforcing and surface-negative polarization should be negatively reinforcing. This prediction has been confirmed.

Rusinov (1953) discovered that surface-positive polarization of the rabbit's motor cortex resulted in limb responses for previously ineffective visual and auditory stimuli and, furthermore, the responses could be elicited for up to 30 minutes after cortical polarization ceased. Any stimulus that had been ineffective prior to polarization remained so afterward if it was not presented and responded to during polarization. The neuronal model suggests that the continued response beyond the period of polarization resulted from increases in the transmittances of those excitatory synapses that were involved in producing the limb response during polarization. The polarization provided continuous positive reinforcement for these synapses. Rusinov's observations have since been confirmed by Morrell (1961). The question remains as to why the apparent learning

disappeared completely after 30 minutes. One possible explanation is that the experimental procedure did not produce a sufficient increase in the excitatory transmittances to prevent interference due to subsequent learning and changing patterns of neural activity. A more permanent form of learning is exhibited in the experiments described next.

Bindman, Lippold, and Redfearn (1964) have provided data with respect to highly localized neural behavioral changes involving at most about five neurons. In some of their experiments, instead of using surface polarization techniques, localized polarization was accomplished by passing a suitable current between a recording electrode (which was doubling as a stimulator) and an indifferent electrode. In the rat cortex, tip-positive polarization was found to increase the average frequency of firing of neurons in the immediate vicinity of the electrode. It should be noted that the tip-positive polarization alone was insufficient to cause the neuron to fire. The increased frequency resulting from tip-positive polarization was sustained by the neuron for up to 5 hours after stimulation ceased, if the duration of stimulation was not less than 5 to 10 minutes. The heterostatic model predicts this result since the recording electrode was supplying a depolarizing bias that functioned as positive reinforcement each time the neuron fired. Excitatory synapses were, therefore, undergoing increases in their transmittances during the period of stimulation. The model further predicts that the average frequency of firing should undergo a long-term decrease if tip-negative polarization is applied with the recording electrode. This should occur because, according to the model, the resulting hyperpolarization causes the effectiveness of active inhibitory synapses to increase each time a neuron fires. Precisely this kind of symmetry to the tip-positive case was observed. Tip-negative polarization for 5 to 10 minutes yielded a decreased average firing frequency that was maintained by the neuron after the polarizing electrode was turned off.

Similar results, again consistent with the model, were obtained using surface-positive and negative polarization. Bindman, Lippold, and Redfearn (1964) state in their summary: "Surface-positive current enhances neuronal firing and increases the size of evoked potentials; surface-negative current has the opposite effect." Furthermore, "If current flow is prolonged for 5 to 10 minutes, a persistent after-effect in the same direction is produced, lasting often without decrement, for at least 1 to 5 hours."

The changes in synaptic transmittances that were obtained in the above experiments by means of surface-positive or tip-positive polarization may also be obtained by means of stimulation of subcortical facilitatory pathways (Bindman, Lippold and Redfearn, 1964, p. 381; Burns, 1968, p. 153, Fig. 73). Stimulation of a subcortical pathway for a period of time produced long-lasting increased activity. This result is consistent with the neural model since the electrode stimulation forces the pathway involved to deliver a depolarizing bias, thereby causing some of the

excitatory synaptic transmittances of recipient neurons to increase. This is an important observation because it suggests that physiological and artificially induced depolarization have equivalent reinforcing effects on synaptic transmittances.

It is important to note that the cortical surface potential changes that were artificially induced in the above-described experiments have been observed to occur naturally during conditioning studies (Rowland and Goldstone, 1963; Walter et al., 1964). This is further evidence that cortical polarization simulates naturally occurring conditions of learning within the cortex.

Bureš and Burešová (1970) have reviewed classical conditioning studies (Bureš and Burešová, 1965, 1967; Gerbrandt et al., 1968; Gerbrandt, Bureš and Burešová, 1970) conducted "on nearly 250 neurons in the cerebral cortex, hippocampus, thalamus, and reticular formation of curarized unanesthetized rats. The classical conditioning paradigm was used throughout, the CS (acoustic or tactile stimulus) preceding and partly overlapping with the UCS (anodal or cathodal current applied through the recording extracellular microelectrode)." Bureš and Burešová (1970) also report on single-cell conditioning studies utilizing the same US but with a CS consisting of "electrical stimulation of neurons in the close vicinity of the nerve cell." In this case, the CS was considered to be effective only when it resulted in impulses being delivered to the recorded cell. In both kinds of study, about half of the neurons that exhibited a plastic reaction behaved in accordance with the heterostatic model and the experimental results reviewed above. The other half behaved in an opposite manner: a depolarizing US was negatively reinforcing and a hyperpolarizing US was positively reinforcing. These results suggest that either (a) more than one type of neuronal adaptive mechanism is to be found in the nervous system or (b) that the nonphysiological source of polarization has introduced uncontrolled variables producing the appearance of an inverse neuron. When polarization was provided by the more physiological means of subcortical stimulation (studies noted above) or in the production of a mirror focus (reviewed below), no inverse neurons were observed. Thus, the present theory will be developed, assuming only the existence of the neuron modeled in Section 3.1. If the inverse neuron does in fact occur also, it will be necessary to revise the theory to include two types of heterostatic neuron: one that seeks depolarization and another that seeks hyperpolarization.

3.2.2. THE MIRROR FOCUS

Along with neuronal and cortical polarization, the mirror focus appears to be one of the neurophysiological phenomena most capable of providing insights into the cellular mechanisms of learning (see review by Morrell, 1969). The procedure for inducing a mirror focus is a straightforward one. A primary epileptic focus, obtained by the localized freezing of cortical tissue, produces a secondary or "mirror" focus at a corresponding point in the opposite hemisphere. Initially, the

secondary focus exhibits abnormal activity only in concert with the primary focus, the facilitation reaching the mirror focus via the corpus callosum. The postulated neuronal model suggests that the depolarizing bias supplied by the primary focus should cause excitatory transmittances to increase in the mirror focus. Therefore, eventually, the mirror focus should exhibit an abnormally high excitability independent of the primary focus. This is indeed what happens. After approximately one week, in the case of the rabbit, the primary focus may be removed and the mirror focus will continue independently. If the primary focus is removed before the end of the first week, an independent mirror focus is never established.

Due to the postulated zerosetting mechanism in the neuronal model, the development of an independent mirror focus should be slowed or stopped if all except callosal inputs are eliminated. This prediction has been confirmed (Morrell, 1969) by experiments in which the tissue of the secondary focus is deprived of all subcortical and long intrahemispheric inputs by undercutting. The result of this partial isolation is the appearance of a dependent mirror focus only; an independent focus is never achieved. Consistent with the neuronal model, the interaction of thalamocortical and callosal inputs is necessary for a pathologic increase in the excitatory transmittances of the mirror focus. Each set of inputs is supplying a positively reinforcing (depolarizing) bias for the other set. If one set is eliminated, the other cannot supply its own reinforcement simply through repetitive firing.

Morrell (1969, p. 357) notes that a primary cortical focus may induce secondary epileptogenesis, not only in the opposite hemisphere, but also in subcortical structures such as thalamic (Wada and Cornelius, 1960) and limbic system nuclei (Guerrero-Figueroa et al., 1964). Also, the primary focus may assume a subcortical location (Morrell et al., 1965; Proctor et al., 1966). These observations are consistent with the assumption of a heterostatic neuron. Whenever a primary focus is established, it may be expected that it will induce abnormal neural assemblies in any area to which it projects a sufficient positive or negative reinforcing bias. In fact, it will be hypothesized that this is the mechanism by which epileptic seizures become progressively more extensive if unchecked by treatment. Specifically, the sequence of events in epilepsy is suggested to be the following:

1. A pathologic process causes an assembly of neurons to become hyperactive. Ward, Jasper, and Pope (1969, p. 10) note that "seizure discharges may arise from such divergent reactions to injury as those associated with intracerebral traumatic lesions, chronic sepsis, infraction, neoplasis, and spontaneous, degenerative neuronal disease of many kinds."

2. The hyperactive assembly, once formed, impresses a depolarizing, hyperpolarizing, or mixed bias, depending upon the assembly's neural composition, on those neurons to which it sends its output.

3. Neurons receiving the bias undergo heterostatic adaptation for a period of time, thus becoming hyper- or hypoactive assemblies themselves.

4. Formation of new assemblies continues until a stage is reached where either (a) the bias generated by the most recently formed assemblies is so weak or is distributed in such a diffuse manner as to be incapable of inducing the formation of further assemblies, or (b) the collection of assemblies forms a closed loop.

Viewing epilepsy as a progressive disease (Morrell and Baker, 1961) is supported not only by work on the mirror focus but also more recently by the work of Goddard (1967). Goddard's observations provide further evidence that a depolarizing bias may induce the formation of an epileptic focus in a number of regions of the nervous system. Goddard implanted electrodes at various locations in the rat's brain and found that stimulation of only 50 μA for 1 minute per day, while posing no serious behavioral problems for the rat during the first few days, eventually produced convulsions. It appears that a permanent epileptic focus was established, as evidenced by the fact that repetition of the stimulation after it had been withheld for 3 months yielded a convulsion once again. It is of interest to note that epileptic foci were established by Goddard in the amygdala after 15 days of stimulation, in the caudate putamen after 74 days, but not in the reticular formation after 200 days of stimulation. The question of why epiletogenicity varies as a function of the brain region will be taken up in Section 4.

Some evidence relating to the spinal cord is relevant here. The spinal cord has been closely examined for synaptic changes associated with learning (Eccles, 1964, 1966), but these studies have not, in general, yielded evidence of long-term synaptic modifications. One exception concerns a phenomenon described in a review by Gurowitz (1969, p. 31). After a unilateral cerebellar lesion, a rat experiences a postural asymetry in the hind limbs. The asymetry disappears if a midthoracic section is performed immediately following the lesion. However, if this section is delayed sufficiently, the asymetry becomes permanent. Gurowitz states that the asymetry is considered to be the result of asymetrical facilitatory stimulation via the descending spinal tracts." These observations suggest that the spinal cord is capable of learning when subjected to the influence of a depolarizing bias. This result, in combination with the evidence reviewed above, implies that most, if not all neural tissue, is capable of undergoing heterostatic adaptation.

3.2.3 MODIFIABILITY OF SYNAPTIC TRANSMITTANCES

While the evidence reviewed above indicates that synapses do undergo changes in their transmittances, or that some functional equivalent of this occurs, the cellular mechanism by which this is accomplished has, in the past, remained obscure. Whether new synapses are grown, old synapses are lost, or existing

ones are modified has been a matter for speculation. Variations in the amount of transmitter released or the extent of postsynaptic receptor sites are among the possible means that have been considered for achieving plasticity. Dendritic spines have appeared to be promising candidates for research on this question. Also, molecular mechanisms have been researched as a possible substrate for learning and memory. At the present time, there is evidence (Deutsch, 1971) favoring the conclusion that "as a result of learning, the postsynaptic endings at a specific set of synapses become more sensitive to transmitter." The proposed heterostatic model is therefore consistent with the current evidence. On the subject of synaptic variability we will limit ourselves in Section 3.3 below to some speculation concerning which synapses are modifiable and whether the transmittances may undergo either increases or decreases or both.

3.3 Variations of the Heterostatic Model

In the neuronal model that has been described, adaptation is accomplished with synaptic transmittances that undergo only increases in their absolute values. We will now consider some variations of the model that are obtained when other kinds of constraints are imposed on the variability of the transmittances. All of the models to be considered are equivalent in that they implement the basic heuristic already described which tends to maximize the amount of polarization experienced by the neuron.

Some of the constraints to be considered are those in which (a) synaptic transmittances increase only, (b) synaptic transmittances decrease only, (c) only excitatory synaptic transmittances vary, (d) only inhibitory synaptic transmittances vary, (e) only the transmittances associated with excitatory neurons vary, or (f) only the transmittances associated with inhibitory neurons vary. These possibilities, including some in combination, are shown in Table 1, which defines the variations of the heterostatic model to be commented on below.

3.3.1 TYPE 1

This is the model which has been detailed above in Sections 3.1 and 3.2. One attractive feature of models of this kind, in which transmittances increase only, is that some kind of growth process could provide the basis for the transmittance changes. Furthermore, a model employing unidirectional transmittance changes might offer an explanation for the apparent loss of neuronal plasticity with increasing age. Bidirectional changes might be expected to provide a higher degree of plasticity for a nervous system, independent of how much experience had already been "read in." With a bidirectional system, the advantage of decreased rigidity might have an associated disadvantage: the increased ease of new learning might have to be purchased with an increased tendency to lose (erase) previously acquired

Table 1. Variations of the Heterostatic Model of the Neuron

Type of Adaptive Mechanism	Type of Reinforcement* (P = Positive) (N = Negative)	Direction of change of absolute value of transmittance (I = increase, D = decrease; where no symbol appears, synaptic transmittance is not modifiable).			
		Excitatory Neurons		Inhibitory Neurons	
		Eligible Excitatory Synapses	Eligible Inhibitory Synapses	Eligible Excitatory Synapses	Eligible Inhibitory Synapses
1	P	I		I	
	N		I		I
2	P		D		D
	N	D		D	
3	P	I		I	
	N	D		D	
4	P		D		D
	N		I		I
5	P	I			
	N		I		
6	P			I	
	N				I
7	P		D		
	N	D			
8	P				D
	N			D	
9	P	I			
	N	D			
10	P			I	
	N			D	
11	P		D		
	N		I		
12	P				D
	N				I

*This refers to the type of signals that impinge upon the neuron within the first few seconds after it fires. Depolarizing signal configurations constitute positive reinforcement; hyperpolarizing signal configurations constitute negative reinforcement.

information; that is, forgetting due to interference might be more of a problem with bidirectional transmittance changes.

3.3.2 TYPE 2

This model may be viewed as the mirror-image of type 1 adaptation. Since transmittances only decrease, this adaptive mechanism could be realized by some kind of regulated decay process. Here, the transmittances would initially have to assume high values, relatively speaking, whereas with type 1 adaptation, the transmittances would initially assume low values.

3.3.3 TYPES 3 AND 4

These models differ from types 1 and 2 in that transmittance modifications of only the excitatory (type 3) or the inhibitory (type 4) synapses are permitted. In order to maintain the power of the adaptive process (that is, in order to still be able to positively and negatively reinforce), bidirectional transmittance changes must be utilized instead of the unidirectional changes employed with types 1 and 2.

It has been observed in some portions of the brain that excitatory synapses tend to be located on the dendrites and, especially, on the dendritic spines (Eccles, 1964, p. 253), whereas the inhibitory synapses tend to be located on the soma or on the dendritic stump (Andersen, Eccles, and Lyning, 1963; Blackstad and Flood, 1963; Anderson, Eccles, and Voorhoeve, 1963). If the mechanism whereby transmittances are modified is located only within the dendrites—for example, if it is associated with the dendritic spines—then type 3 adaptation may be a relevant model because it leaves inhibitory synaptic transmittances fixed. As noted above, one of the prices to be paid for having a model that permits only the excitatory or the inhibitory transmittances to vary, but not both, is that the modifiable transmittances must then be capable of changing in either direction if the power of the system is not to be drastically reduced. Therefore, bidirectional changes might mean that the part of the neuron corresponding to the transmittance-modifying mechanism would have to be more complex than in the case of type 1 or type 2 adaptation.

Type 4 adaptation might be the appropriate model of the neuronal adaptive process if it should turn out that only axosomatic synaptic transmittances are modifiable.

3.3.4 TYPES 5 THROUGH 12

These represent variations of the first four types. In these eight models, we examine several ways of obtaining adaptive processes that possess essentially all of the power of the first four types, but which require that fewer of the synapses be modifiable. The power of these eight adaptive mechanisms is reduced from that of the first four types only to the extent that fewer modifiable synapses are

available to store the information of memory and learning. This is a minor consideration. The important point, as noted above, is that the fundamental nature of the adaptive mechanism is unchanged with any of the 12 models shown in Table 1. Types 5 through 12 differ from types 1 through 4 in that the modifiable transmittances are restricted to either the excitatory or inhibitory neurons, depending on the model. The models in which only the inhibitory neurons are adaptive might be relevant to the cerebellum where it has been observed that inhibition dominates information processing (Eccles, Ito, and Szentagothai, 1967, p. 311).

In the sections that follow, type 1 adaptation will be assumed for the purpose of discussion. It is important to note, however, that none of the arguments presented would be substantially altered if one of the other models were adopted.

4. PSYCHOLOGY

In this section, the assumption of a heterostatic neuron provides the basis for speculation concerning the neurophysiological mechanisms that underly psychological phenomena. The question of the mind-body relationship is taken up and consideration is given to the global organization of the brain.

4.1 Learning

4.1.1 HABITUATION

Nervous systems become habituated to redundant stimuli while continuing to respond to novel and significant inputs. It is proposed that two types of localized neural circuitry are important in understanding this behavior. These two circuit types are feed-forward and recurrent inhibition, both of which, Eccles (1969a, pp. 43-44) has noted, are found throughout the central nervous system. Feed-forward inhibition makes it probable that when a neuron fires, some inhibitory synapses will have been active. This is important because it insures that recently active inhibitory synapses will generally be available to undergo increases in their transmittances should hyperpolarization follow impulse generation. Recurrent inhibition guarantees that hyperpolarization will follow impulse generation (and thus will cause the inhibitory synapses associated with the original stimulus to become increasingly effective) unless sufficient facilitation arrives to counter the inhibitory feedback. Evidence of the increasing effectiveness of inhibitory synapses during habituation has been obtained by Holmgren and Frenk (1961) in their study of the pleural ganglion of a snail. It can be seen that the extensive occurrence of feed-forward and recurrent inhibition provides neurons with the means for what is, in effect, self-habituation. Only the intervention of positive reinforcement (excitation) sufficient to counteract the recurrent inhibition can save a stimulus from the neural judgment of irrelevance.

That the frontal cortex can markedly accelerate habituation is supported by much research (see review by Griffin, 1970). It is therefore hypothesized that the process of encephalization has resulted in the frontal cortex becoming a primary source of inhibition to be utilized during the habituation process. It is conjectured that frontal lobe inhibition can be generated in much the same manner as recurrent inhibition, that is, as an inhibitory feedback response to neuronal firing. This type of inhibition may be termed extended recurrent inhibition to distinguish it from that type discussed by Eccles (1969a, pp. 43-44) which will be termed direct recurrent inhibition.

A primary source of the counteracting facilitation, when habituation is avoided, is hypothesized to be the limbic system and the hypothalamus (henceforth collectively designated as the LSH). It is postulated that the LSH, using genetically specified mechanisms, identifies stimuli relevant to the animal's needs (these are the unconditioned stimuli) and responds by delivering facilitation to many parts of the brain. Actually, this postulated LSH action is expected to apply only to appetitive unconditioned stimuli. Defensive unconditioned stimuli will be discussed later, along with a more detailed discussion of the LSH's reinforcement centers.

It is now possible to understand what happens when a neutral stimulus (a potential conditioned stimulus) is processed by the RF, for example. First, we note that the RF is always aroused by a novel stimulus. This indicates that excitatory synapses of excitatory RF neurons dominate over inhibitory synapses initially. However, after the RF receives the stimulus, if the LSH is not aroused within a few seconds by a stimulus that it judges to be significant (the arrival of an unconditioned stimulus) and, therefore, the LSH does not supply facilitation to the RF, frontal lobe inhibition, arising out of the RF's initial response, insures that active RF neurons will be negatively reinforced; that is, hyperpolarization will follow impulse generation and recently active inhibitory synapses will undergo increases in their transmittances. With repetitive presentations of the stimulus, these inhibitory transmittances continue to increase until the neurons no longer respond. In a similar fashion, habituation may occur elsewhere, such as in the cortex, where direct or extended recurrent inhibition or both may be active.

4.1.2 DISHABITUATION

This phenomenon results when (a) the habituated stimulus is withheld for a period of time, (b) a novel stimulus is presented prior to the habituated stimulus, or (c) the habituated stimulus is paired with another stimulus. In all of these cases, dishabituation is hypothesized to be due to the nervous system assuming a sufficiently different state of activity by the time the habituated stimulus is presented again so that a new set of synapses is involved. Newly involved inhibitory synapses then undergo increases in their transmittances and habituation is established once more. In this way, inhibition relative to a particular stimulus becomes

more extensive and it therefore becomes progressively more difficult to dishabituate a stimulus following repeated habituation. The suggestion of Spencer et al. (1969 a, b) that a dishabituating stimulus has a generalized facilitatory effect further explains how increased inhibitory transmittances can be overridden during dishabituation.

Habituation is often highly specific to the stimulus presented. Slight changes in the stimulus cause the animal to respond again. This is reasonable because the increased inhibitory transmittances mediating habituation will be of such a magnitude as to only minimally inhibit the neurons involved. Minimal inhibition occurs because, as postulated in Section 3, a neuron must fire if synaptic modifications are to follow. For this reason, as soon as the inhibitory transmittances increase just enough to terminate the neuronal response, the transmittances cannot increase any further. In this situation, it can be seen that inhibition will have only a very small edge over excitation. Therefore, when even a small number of the increased inhibitory transmittances are not utilized, such as in the case of a slightly different stimulus, the animal once again responds. Also, the fact that habituation utilizes minimal inhibition is undoubtedly an important factor in explaining why dishabituation can occur so easily.

It is interesting to note that the inhibitory neurons of the cerebrum, when they are induced to fire, continue to do so for 200 to 500 ms (Eccles, 1966, pp. 332-333). This period resembles the optimal CS-US interval of conditioning studies, indicating that the cerebral inhibitory neuron is well suited for its role as a generator of negative reinforcement. On the other hand, Eccles notes that inhibitory neurons in the spinal cord are not capable of this kind of long-term firing; instead, a single volley lasts for no more than 20 ms. Also, the amplitude of the spinal IPSP is, at most, one-tenth of that for cerebral inhibitory neurons. Since reinforcing signals do not appear to have their maximal effect on a neuron until approximately 400 ms have elapsed after impulse generation, one would expect habituation to proceed slowly, if at all, in the spinal cord. This is consistent with what is observed. For example, cervical transection of the spinal cord of a kitten, performed to eliminate the influence of higher centers, yields an animal that requires 2 weeks of training before habituation of the withdrawal reflex will occur (Meldrum, 1966, p. 103). In the intact animal, habituation to a stimulus is accomplished within a few trials.

Since the evidence suggests that there are two basic types of inhibitory neurons, it would be of interest to know exactly how they are distributed throughout the CNS. This question assumes new significance because the present theory suggests that only one of the two types is well suited for use in adaptive circuits. Eccles (1969a, pp. 106-110) has reviewed evidence suggesting that inhibitory neurons having a short-duration IPSP are restricted to the spinal cord, while inhibitory neurons

at suprasegmental levels are of the long-duration IPSP type. This distribution has been deduced on the basis of observations of the depressant action of strychnine, which affects only inhibitory neurons having short-duration IPSP's. The observed distribution is interesting because it shows that the midbrain is the lowest level at which we find inhibitory neurons that are well suited for the purpose of generating negative reinforcement. Penfield has reviewed evidence suggesting that the midbrain and thalamic regions of the brainstem are of critical importance for the maintenance of conscious awareness. Penfield (1969, p. 800) notes that "compression or injury in this region . . . extinguishes awareness and is followed by amnesia." Thus, interestingly enough, adaptive circuits appropriate to such advanced learning phenomena as rapid habituation appear at just that level in the CNS which may also serve as the seat of consciousness.

Habituation below the midbrain may be accomplished by an entirely different mechanism than that employed at and above the midbrain. This suggestion is supported by a study of the bulbar reticular formation (Segundo et al., 1967) in which the investigators concluded that habituation occurred because "interneuronal pools have junctions whose activation is followed by prolonged subnormality [of the presynaptic terminal]."

4.1.3 CLASSICAL PAVLOVIAN CONDITIONING

In this discussion of Pavlovian conditioning and that which follows on operant conditioning, only appetitive reflexes will be considered. Defensive reflexes will be considered later, after the role of pain in nervous systems has been analyzed.

The proposed neuronal model predicts that no conditioning will occur with simultaneous presentation of the CS and US and indeed this is observed to be the case. As has been noted, the optimal CS-US interval is approximately 400 ms, with the US becoming completely ineffective if the interval exceeds a few seconds (see review by Russell, 1966, p. 136).

In the following discussion, the sets of neurons that fire in response to the presentation of a CS or US will be denoted by $\{CS\}$ and $\{US\}$, respectively. The neuronal model predicts that when these sets are sequentially active within a few seconds of one another, those neurons belonging to the intersection, $\{CS\} \cap \{US\}$, will be positively reinforced. That is to say, excitatory synapses of the neurons belonging to the intersection will undergo increases in their transmittances if these synapses were activated by the CS but not activated (and thus zeroset) by the US. Furthermore, the LSII's response to the US is a source of positive reinforcement for neurons in $\{CS\}$. The effect of both kinds of reinforcement during conditioning will be to cause the sets $\{CS\}$ and $\{CS\} \cap \{US\}$ to increase in size and average frequency of firing. As these increases occur for the set, $\{CS\} \cap \{US\}$, one would expect an increasing resemblance between the CR and the UR, as indeed happens.

It is now known that the CR is a fractional component of the UR (Konorski, 1948); the CR and UR do not become identical as was once thought. The fractional nature of the CR is to be expected since the intersection, $\{CS\} \cap \{US\}$, will in general contain only a fraction of the neurons in $\{US\}$.

4.1.4 OPERANT CONDITIONING

In the following discussion, CS, CR, US, and UR refer to instrumental stimuli and responses. The sets, $\{CS\}$ and $\{US\}$, will be defined as in the case of classical conditioning.

Whenever a CS-CR sequence occurs, the US that follows causes the LSH (according to our earlier hypothesis) to supply positive reinforcement to $\{CS\}$, thereby causing the active excitatory synapses of $\{CS\}$ to undergo increases in their transmittances. As these transmittances continue to increase with further conditioning, the CR must follow the CS with increasing reliability. For the case where the CR is not preceded by the CS, habituation eventually eliminates the response since the US will not be present to arouse the LSH and the LSH, therefore, will not supply positively reinforcing facilitation.

In the case of habituation, it was noted that because only minimal inhibition developed, the habituated stimulus did not generalize. Slight changes in the stimulus produced a response from the animal again. In the case of conditioning with an appetitive US, excitatory transmittances are increasing and there is nothing to prevent them from continuing to increase, with repetition of the CS-CR sequence, far beyond the minimal level required to produce the conditioned response. We would therefore predict that conditioned stimuli should generalize, since the large transmittance values will make it unnecessary that every excitatory synapse involved in the response to the original stimulus be active in order to obtain the same kind of response to a modified stimulus. The predicted generalization does, of course, occur; a variety of stimuli similar to the CS will evoke the CR (see review by Russell, 1966, p. 135).

4.1.5 EXTINCTION

When a classical or operant CR ceases to be reinforced, extinction of the response eventually occurs. It is proposed that the mechanism involved is one and the same as that for habituation. An identity of the underlying mechanisms of habituation and extinction has been considered by others (Sharpless, 1964; Thompson and Spencer, 1966). Furthermore, Kling and Stevenson (1970, p. 46) note that "The theories of extinction cited most often have postulated some process of inhibition." They cite Pavlov's (1927) and Hull's (1943) theories as examples.

The fact that extinction proceeds more slowly and erratically than habituation can be understood if one considers that the CS not only becomes associated with the CR but may be expected to also become associated with the US. That is to say, the CS becomes conditioned to activate not only the CR but also to some extent the set $\{US\}$, and the CR becomes partially self-reinforcing. Thus, when the US is no

longer presented, the set US may still be partially active, following the occurrence of the CS-CR sequence. It is hypothesized that this activity of the set, $\{US\}$, causes the extinction of the CS-CR association to proceed more slowly and erratically than the normal process of habituation.

This ability of a CS to become associated with a US, as well as with a CR, thereby providing the CS with a dual role, is perhaps what permits long behavioral chains to develop, such as in the running of a maze. In such a chain, leading ultimately to a "true" US, an intermediate stimulus comes to serve as a CS with respect to the next response and as a US with respect to the preceding response.

The ability of an extinguished response to reappear after only a very small amount of renewed conditioning is reasonable, due to the fact that only minimal inhibition would be expected to develop during the extinction process. As in the case of habituation, minimal inhibition occurs because the inhibitory transmittances stop increasing immediately upon cessation of the response. Thus, only small increases in excitatory transmittances will be necessary for the previously extinguished response to appear again.

4.2 The Mind-Body Problem

The heterostatic model of the neuron was tested in Section 3 with neurophysiological data relevant to the problems of learning and memory. The model was found to be consistent with the experimental results. In Section 4.1 we continued the testing of the model, this time in the context of psychological studies of habituation, classical and operant conditioning, and extinction. Again, the neuronal model has been found to be consistent with the available data and has offered insights into underlying mechanisms. Having thereby provided experimental support for the model at both the neurophysiological and psychological levels, we now turn to the problem of considering, in a more complete fashion, what the assumption of a heterostatic neuron implies for the global organization of the brain. As a first step we will take up the issue of the mind-body problem.

4.2.1 ANIMALS ARE HETEROSTATS

For a given system, if all of the constituent elements seek a common goal, it is plausible that the behavior of the total system will exhibit the same goal. On this basis, we would expect to find that nervous systems seek to obtain depolarizing experiences and to avoid hyperpolarizing experiences. More precisely, it is hypothesized that nervous systems tend to maximize the average amount of neuronal polarization experienced. Thus, it is postulated that animals are heterostats.

4.2.2 AN IDENTITY THEORY

Since human nervous systems and, by implication, the nervous systems of other animals seek to obtain pleasure and to avoid pain, it is hypothesized that

pleasure and pain are the psychological counterparts of the physicalistic concepts of depolarization and hyperpolarization, respectively. This solution to the mind-body problem is proposed as an identity theory (Feigl, 1958, 1960; Pepper, 1960) of the Φ-Ψ (physiological-psychological) relationship.

The suggestion is that pleasure and pain as experienced by individual neurons undergoing depolarization or hyperpolarization, respectively, constitute the two fundamental "psychic atoms" out of which all of our more complex mental states are constructed. ("Psychic" and "mental" are synonymous in this discussion.) Each neuron may be thought of as generating a psychic field during polarization, this field corresponding to some as yet unidentified aspects of the polarization process. We may consider that there are d-fields and h-fields, resulting from depolarization and hyperpolarization, respectively.

It is suggested that psychic fields of individual neurons combine to form more complex psychic fields and thus arises the entity termed "mind." If this proposal seems farfetched, consider that if the concept of mind is ever to be related to physical processes, it would seem that the processes will have to be located in individual neurons and then "integrated" to form the more complex entity we term conscious awareness. If we wish to explain the phenomenon of consciousness without integrating primitive psychic fields to obtain more complex fields, our only alternative would seem to be that of associating our consciousness with processes in a single neuron. This hypothesis seems less plausible than that requiring the integration of multiple primitive fields.

4.2.3 LOCALIZATION OF CONSCIOUSNESS

The reticular formation, at the medullary, midbrain, and thalamic levels, has been hypothesized to serve as the command and control center for the brain (Kilmer, McCulloch, and Blum, 1967). It has been suggested that the RF commits us at any given moment to one of approximately two dozen possible gross modes of behavior such as running, fighting, sleeping, eating, or mating. As noted earlier, Penfield (1969) has reviewed evidence that points toward the localization of conscious processes within the mesencephalon and diencephalon. We also reviewed evidence earlier that showed that long-duration IPSP neurons that appear to be well suited as adaptive circuit elements occur at the level of the midbrain and above but not below. On the basis of all of these considerations, it is hypothesized that the midbrain and thalamic RF (MTRF) is the seat of the mental experiences of which we are aware.

MTRF neurons appear to simultaneously possess (a) a strategic location providing access to information from most other areas of the brain, (b) substantial complexity of organization, and (c) control of an advanced communication apparatus. Apparently, for these reasons, MTRF neurons can collectively become aware of

themselves (or at least, of the body they control) and they can make themselves heard via a spoken or written language. The MTRF speaks for us while the other brain structures remain mute, they having to be content with a lesser understanding of affairs going on about them and with only an indirect influence on the communication process. The other brain structures also experience pleasure and pain but it is hypothesized that the physical extent of psychic fields is such that fields associated with other brain structures do not significantly interact with the psychic field of the MTRF. Thus, we are no more aware of the psychic field associated with our cortex, for example, than we are of the psychic field associated with someone else's cortex.

4.2.4 QUANTITATIVE ASPECTS OF PSYCHIC PHENOMENA

The intensity with which we experience pleasure or pain is hypothesized to be approximately proportional to the absolute difference between the number of MTRF neurons undergoing depolarization, denoted by Q(D), and the number undergoing hyperpolarization, denoted by Q(H). Therefore, positive and negative differences for the quantity

$$\Delta(t) = Q_t(d) - Q_t(H) \tag{18}$$

correspond to pleasurable and painful mental states, respectively. The terms, pleasure and pain, are used in a most general sense here, denoting all psychic states as either agreeable or disagreeable.

The kind of pleasure or pain experienced, for example, love or hatred, joy or grief, is hypothesized to depend on the particular configuration of depolarizing and hyperpolarizing events occurring within the MTRF at a given moment. To make this notion explicit, we may define an MTRF state vector, $\overline{\text{MTRF}}$, which will have one component for each neuron in the midbrain and thalamic reticular formation. Each component will always assume one of three values representing the momentary state of the corresponding neuron. Thus, the MTRF state vector

$$\overline{\text{MTRF}}(t) = (R,\ D,\ D,\ H,\ \ldots) \tag{19}$$

indicates that at time, t, neuron 1 is in the resting state (R), neurons 2 and 3 are undergoing depolarization (D), and neuron 4 is undergoing hyperpolarization (H). The dimension of the state vector is estimated to be that of a 10^9-tuple since the total number of neurons for man is estimated to be 10^{12} and the RF of higher vertebrates has been estimated to constitute about 1/1000 (Kilmer, McCulloch, and Blum, 1967, p. 1) of the CNS.

In accordance with assumptions noted above, it is expected that the set of all possible psychic states maps into the set of all possible MTRF state vectors. The delta value, as defined by Eq. (18), measures the degree of pleasure or pain experienced. The state vector, $\overline{MTRF}$, specifies the exact mental state of the organism, that is, the kind of pleasure or pain being experienced.

Since there are an extremely large number (3^{10^9}) of possible MTRF state vectors, this implies a similar number of possible mental states. For a particular brain, however, one might expect that the specific pattern of neural interconnections and a specific environment would act to constrain the MTRF such that it could, in fact, only assume a limited number of basic configurations with variations occurring within each major category. This appears to be the case with the human brain for which the range of psychic states has been encompassed within the major categories of feeling tone, sensation, and ideation. It is hypothesized that these three broad classifications correspond to MTRF states where the LSH, sensory cortex, or nonsensory-nonmotor cortex, respectively, supply the principal input driving the MTRF. It is expected that the highest delta values are associated with LSH-dominant inputs to the MTRF. When one considers the higher information content that would seem to be required with sensation and ideation, it appears reasonable that these mental experiences might be less compatible with high delta values. All of this seems consistent with Pratt's (1937) observation that "Feeling tone is notably different from sensation and ideation. It is less objective; it is relatively independent of external objects and peripheral stimulation."

We are suggesting, therefore, that such mental phenomena as emotions, ideas, and sensations do not differ within or among themselves as much as might be assumed. They are all formed out of the same ingredients, namely, configurations within the MTRF of the neuronal events of depolarization and hyperpolarization. The particular configuration at any given moment is hypothesized to depend to a large extent on which input lines to the MTRF are active. When inputs from the LSH are the primary ones driving the MTRF, a large portion of the MTRF neurons are simultaneously either excited or inhibited, depending on whether the LSH is delivering positive or negative reinforcement. In these circumstances, we refer to the experience as one of emotion. When MTRF inputs from the sensory cortex are primary, we report the experience of sensations. When nonsensory-nonmotor cortex supplies the dominant inputs to the MTRF, we report the experience of memories or ideas. It is anticipated that the mix of excited and inhibited neurons within the MTRF is more nearly balanced in the case of sensations, memories, or ideas than when LSH inputs are "washing over" this structure.

It should be mentioned that the delta value probably represents an idealized measure. It is expected that the interaction of the psychic fields of individual neurons will be influenced by the geometry of the MTRF tissue, and therefore this

geometry will have to be considered in establishing the exact level of pleasure or pain experienced.

4.2.5 THE Φ OF Φ-Ψ

If our speculation thus far on the mind-body problem bears some useful resemblance to the true state of affairs, what then might be the observable physical variables that are identical with the psychic variables? We have proposed that every psychic experience is identical with (some aspects of) a three-dimensional configuration of events, where the events consist of depolarizations and hyperpolarizations of neurons. More precisely, psychic experience consists of a continuous temporal sequence of such configurations. We are therefore proposing to construct complex mental experiences out of the simple mental experiences of the individual neurons. [While we have suggested that the neuron experiences pleasure when undergoing depolarization and pain when undergoing hyperpolarization, it would perhaps be more accurate to say that the neuron (some aspects of it, actually) is pleasure or pain when undergoing polarization.] We propose to integrate large numbers of these two simple psychic building blocks of pleasure and pain into the whole range of complex psychological experiences such as vision, taste, anger, and self-awareness. The question to be dealt with now is this: exactly which physical aspects of the neuronal polarization process are identical with mental experience? Speculation along two lines seems promising.

First, it would appear that we need to look for a field phenomenon. Otherwise, one has to integrate over something as discrete as individual neurons and come up with something possessing the apparent continuity of the mind. As Sherrington (1941) observed, "Matter and energy seem granular in structure and so does 'life', but not so mind." If we do require a field, a natural candidate for the role is of course the electromagnetic field. Can it be that elementary pleasure and pain are one and the same, respectively, as the collapsing and expanding electromagnetic fields that arise during the neuronal polarization process? Is the mind identical with the electromagnetic field generated collectively by the MTRF neurons?

There is another possible approach to the problem of identifying the Φ of the Φ-Ψ relationship. We have indicated that we feel a need to derive the seeming continuity of the mind from some field phenomenon associated with the brain. There is an alternative to the choice of the electromagnetic field as the required continuous substrate of mental experience. Perhaps the answer is to be obtained, as many others have recognized, from a consideration of the wave properties of matter itself. The physicist has found it necessary to ascribe both the properties of particles and of waves to matter in order to explain his experimental data. Could it be that the mind-body problem places us at exactly the same point? The brain appears to be of discrete composition, being made up of such entities as

atoms, molecules, and neurons. The mind, on the other hand, appears to be of a continuous nature, suggesting a wave phenomenon. It seems plausible that mind and (some aspects of) brain are identically the same, they being the wave and particle aspects, respectively, of a single mind-brain (wave-particle) entity.

4.2.6 PROPERTIES OF THE PSYCHIC FIELD

If the solution of the mind-body problem requires a field phenomenon (what we have termed a psychic field) and if this field is in fact either an electromagnetic field or a field derived from the wave properties of matter, then we may try to anticipate some of this field's properties. For example, when a d-field occurs, it must be followed by repolarization during the recovery process of the neuron. When an h-field occurs, it must be followed by partial depolarization during the recovery after inhibition ceases. Now, Eq. (18), which defines the delta value, implies that simultaneous d-fields and h-fields tend to cancel one another. If so, do the respective recovery processes then cancel the initial depolarization or hyperpolarization of a neuron? For example, is the production of a d-field (pleasure) then followed by an equal amount of pain corresponding to the repolarization? That this kind of cancellation of psychic field types occurs seems unlikely because, from introspection, we know that we can experience pleasure or pain continuously over an extended period of time. Introspective evidence therefore suggests that the delta value does not always hover near zero. If such an introspective observation is valid, how do we account for the wide deviations of the delta value from zero? A possible explanation is that the intensity of the pleasure or pain is proportional to the rate of change of Φ. If this were true, it would eliminate the symmetry between the depolarization or hyperpolarization and the respective recovery processes, since the recovery occurs at a lower rate than the initial polarization. With the elimination of the symmetry, the cancellation of the fields would not be expected to occur.

Another property of psychic fields that we might anticipate is that the field intensity should diminish as a function of the distance from the source. This is, of course, the case with electromagnetic fields. Such would seem to be a necessary property of Φ if we are to explain the fact that the mind apparently does not reflect the state of all nearby brain structures but only that of the MTRF.

4.2.7 DEFINING CONSCIOUSNESS

Consciousness is a term that has been used up to this point without having been explicitly defined. Let us now remedy this situation. A conscious system is defined to be one that fulfills two conditions: it contains a model of itself, its environment, and the relationship of the two; furthermore, the information contained within the model must be encoded in terms of Φ, the physical variables that are identical with the psychic variables we introspectively know. Should the

system's information be encoded in terms of some other set of physical phenomena, the system could behave as we do but not necessarily feel as we do. This is a way of observing that systems may be isomorphic with respect to their behavior and still differ with respect to their psychic existence. This brings up the question of whether there are, within the universe, relationships of a Φ-Ψ nature besides the one we are immediately concerned with here. For example, if electromagnetic fields are the physical substrate of mental existence as we know it, might another kind of mental existence be known by a system in which gravitational fields provided the physical substrate? Or, if it should turn out that the mind corresponds, in some respect, to the wave aspect of the wave-particle nature of matter, then panpsychism would appear to become a tenable philosophy and we would have to consider the possibility of differing Φ-Ψ relationships for all systems.

Consciousness involves an awareness of self and is therefore an emerging property of brains, a property that is acquired as experience and learning provide increasing information with which to define the self, the environment, and their relationship. In accordance with this definition, new-born infants are conscious only to a slight degree, if at all. (We can speak of degrees of consciousness just as we can speak of degrees of clarity of the definition of the self and the environment, etc.)

There has been a tendency in the past to confuse responsiveness in living systems with consciousness. Here, we are narrowing the definition of consciousness with respect to some of the previous usages. Consciousness, in the past, has also been associated with the idea of simply having a psychic existence, that is, having such experiences as those of pleasure and pain in their elementary or complex forms. The present speculation on the mind-body problem, however, suggests that pleasure and pain may be as pervasive within the universe as electromagnetic fields or matter itself. If so, the definition of consciousness must be narrowed if the term is to remain useful. The definition proposed above is an attempt to accomplish this by requiring a certain kind of organization, information content, and encoding before a system is judged to be conscious.

4.3 Global Organization of the Brain

4.3.1 THE RETICULAR FORMATION

The global organization of the brain will be analyzed in terms of three major subsystems: the reticular formation (RF), the limbic system and hypothalamus (LSH), and the neocortex (see Figure 2). It was suggested above that animals are heterostats. By way of explaining this statement, we can now add that animals seek to obtain maximal polarization for the neurons of the MTRF. The remaining neural tissue is also seeking maximal polarization, but only the MTRF possesses

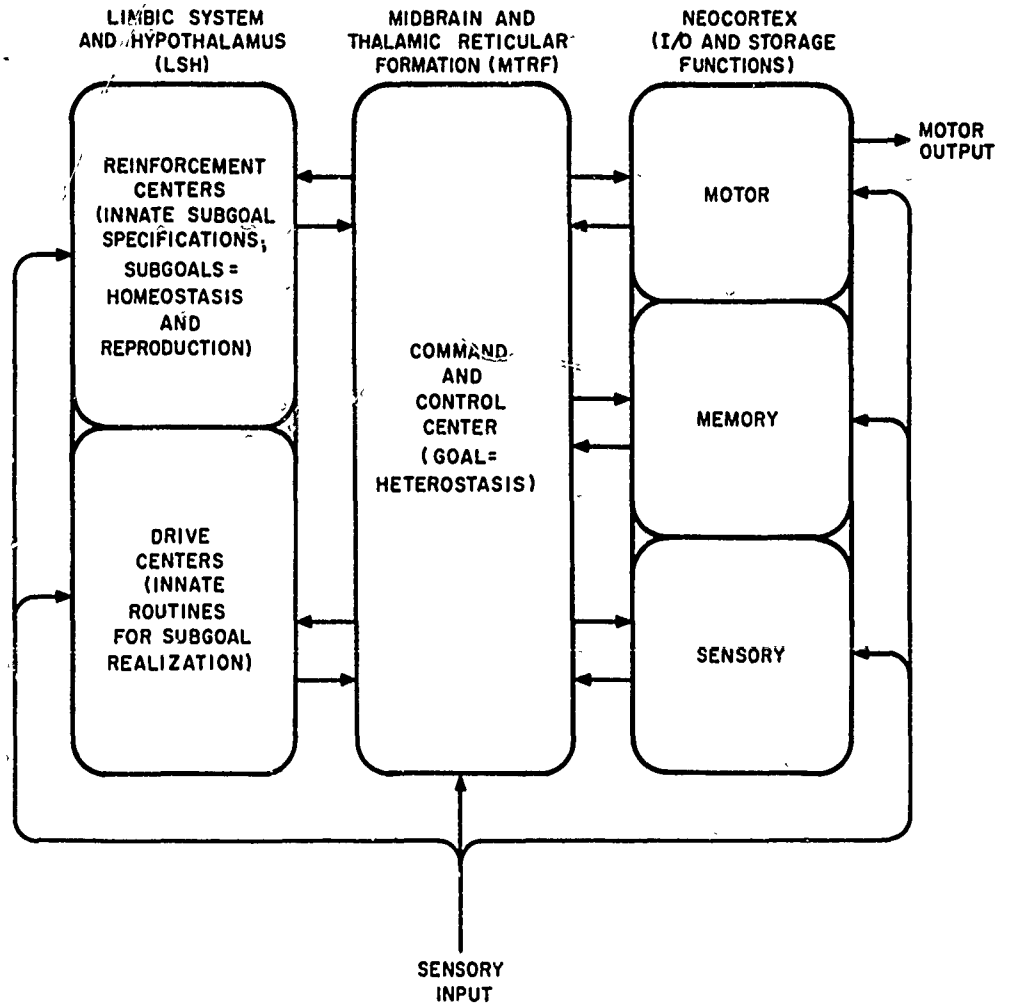


Figure 2. Global Organization of the Brain Illustrated in Terms of the Major Subsystems. (Direct connections between the LSH and the neocortex are not shown)

the strategic anatomical position and connections necessary for the primary decision-making role relative to the animal's total behavior. Other neural tissue may lobby but the MTRF alone ultimately makes the decisions, and the nature of the heterostatic mechanism is such that the MTRF can decide only in favor of itself.

Having suggested that the MTRF seeks to obtain depolarization and to avoid hyperpolarization, what then are the MTRF's sources of excitation and inhibition? The sensory input channels, excluding the pain fibers, are a principal source of excitation. On the other hand, it is hypothesized that the fibers mediating pain are

inhibitory with respect to the MTRF. Viewing both the MTRF and the individual neuron as heterostatic systems, it is thus proposed that sensory inputs excluding pain are to the MTRF what excitation is to the neuron. Similarly, pain fiber inputs are to the MTRF what inhibition is to the neuron.

If the MTRF is a heterostat and sensory inputs, in general, are excitatory, it would be expected that animals would seek novel stimuli since such behavior would yield positive reinforcement. (Novel stimuli are required in order to circumvent the habituation mechanism.) Of course animals do exhibit such a drive and we term it curiosity. In the context of the present theory, therefore, curiosity is seen to be simply one route to sources of excitation for MTRF neurons. On the other hand, neuronal assemblies in the MTRF that produce behavior leading to painful stimuli will eventually become inactive and thus painful stimuli will be avoided. Inactivation of these assemblies will occur because associated inhibitory transmittances will increase with each experience of pain. The increased transmittances result from hyperpolarization of the active neurons following their firing.

4.3.2 THE LIMBIC SYSTEM AND HYPOTHALAMUS

The sensory tracts and pain fibers are not the only sources of excitation and inhibition for the MTRF. Another very important source of both types of signals is the LSH. It is hypothesized that the principal function of the LSH is that of an interface and transducer between the MTRF, on the one side, and the internal and external environments of the organism on the other side. The particular variables to be transduced are well known. The evolutionary process has taught the LSH to pay close attention to the essential variables of homeostasis. Specifically, it is postulated that the LSH, using genetically specified mechanisms, delivers (a) generalized excitation to the MTRF (and other structures) when essential variables are moving toward homeostasis, (b) no signal when homeostasis is obtained, and (c) generalized inhibition to the MTRF (and other structures) when essential variables are deviating from homeostasis. This hypothesis suggests that it is the achievement of or deviation from homeostasis and not the maintenance of the condition that is innately reinforcing. It must be remembered that the condition of homeostasis refers to a range of values for the essential variables. Survival is possible so long as the variables all fall within prescribed limits. Thus, when one speaks of "moving toward homeostasis" or the "achievement of homeostasis", this usually refers to movement toward an optimal point within the prescribed limits. It can be seen that if the MTRF is excited or inhibited by the LSH, depending on the direction of change of an animal's condition relative to homeostasis, then the MTRF will, as a consequence, seek to maintain homeostasis in its pursuit of heterostasis. Homeostasis is thus seen to be a necessary but not sufficient condition for the achievement of heterostasis.

The transducer function just outlined for the LSH is hypothesized to be the mechanism responsible for producing emotions. It is suggested that the delivery by the LSH of generalized excitation or inhibition to the MTRF causes this center of consciousness to report pleasurable or painful emotional states, respectively. The specific state of the MTRF resulting from a particular configuration of excitation and inhibition impinging on it determines exactly what emotional experience is reported. The intensity of the stimulation delivered to the MTRF determines the intensity of the emotion and also determines the magnitude of the resulting transmittance changes. Larger transmittance changes account for faster learning and more vivid memories in "emotionally charged" situations.

Another LSH transduction of essential variables, also accomplished with innate mechanisms, involves the generation by the LSH of specific commands appropriate to the maintenance of homeostasis. These commands result in visceral, vascular, glandular, and motor responses by the organism. It is hypothesized that what we term motivation or drive is the result of this type of LSH transducer action. The generalized motor arousal indicating a motivated state is hypothesized to be produced by a specific facilitatory command originating within the LSH and which is sent either directly or via the MTRF to the motor cortex and perhaps other cortical areas. This facilitatory signal appears when there is movement away from homeostasis. The facilitation may be focused on the particular motor activities required to correct the type of deviation from homeostasis that is occurring.

The two basic LSH transducer functions that have just been outlined are mediated by what may be termed reinforcement centers and drive centers. The existence of both types of centers can be demonstrated by stimulation at appropriate points within the LSH. The reinforcement centers are sometimes referred to as pleasure and pain centers. Pain centers were first observed by Delgado, Roberts, and Miller (1954); pleasure centers were first observed by Olds and Milner (1954). Drive centers, when stimulated, produce behavior related specifically to anger, fear, or the satisfaction of hunger, thirst, or sexual needs.

It should be noted that the LSH is concerned not only with the preservation of the organism but also with the preservation of the species. It is therefore assumed that the essential variables of reproduction are transduced in much the same manner as those associated with homeostasis.

In recent years, much knowledge has been accumulated concerning the reinforcement centers. Stein (1964) has noted that "Largely as a result of the efforts of Olds (1962), it now seems highly probable (although direct evidence is lacking) that the hypothalamic medial forebrain bundle and its connections play a central role in the mediation of reward and that the periventricular system of the diencephalon and midbrain has a critical part in the mediation of punishment." Stein and Wise (1971) have summarized the results of research as it relates to the

positive reinforcement system: "Physiological (Olds and Milner, 1954; Olds, 1962; Miller, 1957), histochemical (Fuxe, 1965; Hillarp, Fuxe, and Dahlstrom, 1966), and psychopharmacological (Stein and Seifter, 1961; Poschel and Ninteman, 1963; Stein, 1964; Wise and Stein, 1969; Stein and Seifter, 1970) work has led to the suggestion (Stein, 1967, 1968) that rewarded or goal-directed behavior is controlled by a specific system of norepinephrine-containing neurons in the brain. The cells of origin of this system are localized in the lower brain stem, and the axons ascend via the medial forebrain bundle to form noradrenergic synapses in the hypothalamus, limbic system, and frontal cortex. Electrical stimulation of the medial forebrain bundle serves as a powerful reward and also elicits species-typical consummatory responses, such as feeding and copulation, which produce pleasure and permit the satisfaction of basic needs (Olds and Milner, 1954; Olds, 1962; Miller, 1957). Electrolytic lesions of the medial forebrain bundle, or pharmacological blockade of its noradrenergic function, cause severe deficits in goal-directed behavior and the loss of consummatory reactions (Teitelbaum and Epstein, 1962; Hoebel, 1971). There is some evidence that these findings in animals may be extrapolated to man (Heath and Mickle, 1960; Sem-Jacobsen and Torkildsen, 1960)." The portion of the lower brain stem in which the positive reinforcement neurons originate is that which Nauta (1960) has termed the limbic midbrain area (Stein, 1967). We shall enlarge our definition of the LSH so as to include this structure.

An important question to consider is whether the brain can, in essence, be viewed as consisting of two functionally distinct but highly interactive networks. The reinforcement network, which would set the emotional tone, might consist of the LSH reinforcement centers, the MTRF diffuse arrousal system, and axodendritic cortical synapses. The drive network, which would compute the motor outputs, might consist of the LSH drive centers, the MTRF specific projection system, and axosomatic cortical synapses. The specified locations of cortical synapses are arbitrarily chosen and intended only as examples of the kinds of hypotheses to be considered.

The behavior of the LSH becomes more complex with time due to learning. The LSH is assumed to undergo heterostatic adaptation just as the MTRF and other neural tissue does. Both the reinforcement and the drive centers can therefore be conditioned to respond to previously neutral stimuli. A similar suggestion has been made by Stein (1964, p. 117) regarding the reinforcement centers. Stein notes that the hypothesized capability of the reinforcement centers to undergo conditioning to previously neutral stimuli "provides for activation of reinforcement systems of the brain before the occurrence of the reinforcing stimulus and is therefore a mechanism for anticipation or expectation." Stein also suggests that (1) a reciprocally inhibitory relationship exists between the positive and negative reinforcement centers, and (2) the two types of centers are continuously and jointly active in

determining operant behavior. This picture of conditionable, reciprocally inhibitory, and continuously and jointly active reinforcement centers appears to be completely consistent with the present heterostatic theory.

It is now possible to understand how a nervous system acquires conditioned defensive reflexes, as well as the appetitive reflexes already discussed. It must be remembered that the MTRF and other structures may receive either excitation or inhibition from the LSH, depending on whether the animal is moving toward or away from homeostasis. Also, the MTRF receives excitation from the sensory tracts and inhibition from the pain fibers. The operational characteristics of the positive and negative reinforcement centers must be kept in mind, along with the neuronal process of heterostatic adaptation. When all of these factors are considered, the acquisition of defensive conditioned reflexes can be understood in terms of an analysis like that detailed for the appetitive reflexes in Section 4.1.

4.3.3 THE NEOCORTEX

It is hypothesized that the relationship of the neocortex to the MTRF is as follows. When the MTRF is aroused by a stimulus, it delivers signals to the cortex, thereby facilitating cortical neurons that are also receiving the stimulus directly via the sensory tracts. The intersection of direct sensory and MTRF signals within the cortex will cause some cortical neurons to fire and thus deliver signals to the MTRF. In this way, the MTRF has interrogated the cortex and received a response. This response received by the MTRF is hypothesized to be that which we subjectively interpret as sensory input or memory, the particular interpretation depending on factors to be discussed shortly.

The other simultaneous function of the MTRF signals when they reach the cortex is that of serving as positive reinforcement. It can be seen that the continual intersection of sensory and MTRF signals on cortical neurons provides the required conditions for heterostatic adaptation. Thus, when the MTRF interrogates the cortex, it is also reinforcing it. This explains the fact that what we attend to is what we remember, since it is also what is reinforced. That which is not attended to is therefore not reinforced and not remembered. While cortical neurons might initially deliver a report to the MTRF only when they are driven both by the direct sensory tract signals and the MTRF signals, eventually, after sufficient transmittance increases occur with reinforcement, either set of signals alone will be capable of eliciting the report. For example, if one is unexpectedly confronted with a familiar face in a crowd, the cortex is capable of reporting this fact to the MTRF even though the MTRF was not specifically interrogating the cortical neurons relevant to that identification. (Some generalized facilitation of the visual cortex may have to be taking place for this pattern recognition function to occur, however.) Alternatively, a familiar face may be recalled with the delivery of appropriate

signals from the MTRF to the cortex even though the person is not present and therefore the appropriate sensory tract signals are not available to drive the cortex. Thus, it can be seen that the cortical function of memory can be activated either by direct sensory stimulation of the cortex or by means of interrogation by the MTRF.

Attention is hypothesized to be the process whereby the MTRF selectively facilitates cortical neurons and thereby obtains specific reports from the cortex. The selective facilitation can activate sensory or memory systems. These reports, when they reach the MTRF, are interpreted as sensory inputs or memories, a distinction that is not always correctly made. For example, in the case of dreams and hallucinations, some of the information derived from memories is apparently interpreted as sensory inputs. Furthermore, Penfield (1969, pp. 796-797) has observed that when externally applied cortical stimulation activates a memory trace of a person undergoing neurosurgery, the person reports a subjective state like that associated with the original experience. The events seem to be happening again. The subjective state is not that of recall. This effect may occur because Penfield's electrode is supplying energy that would normally be supplied by impulses that reach the cortex directly via the sensory tracts. Thus, the MTRF, having received a report generated by an external energy source, may interpret the signals it is receiving as sensory input, instead of as memory. For the person undergoing the surgery, activation of the memory trace results in his experiencing two "streams of consciousness": the one associated with the sensory inputs provided by the environment of the operating room and the other resulting from the external activation of the memory trace.

The act of recall is hypothesized to occur as follows. First, the MTRF returns to a sufficiently close approximation to the state it was in when the information to be recalled was originally processed. This results in the signals sent to the cortex from the MTRF being sufficiently similar to what they were on the previous occasion to cause the cortex to send back something resembling the original report. (Cortical transmittance increases make it unnecessary for the direct sensory inputs to be present this time. The signals from the MTRF are now sufficient by themselves to drive the relevant neurons.) In this way the MTRF can become aware of, once again, what it was aware of before. It might appear that a regression will occur here because we can now ask "How do we retrieve the appropriate MTRF state in order to then elicit the desired report from the cortex?" The regression, in fact, does not occur because the required MTRF state is already obtained. The fact that one is trying to recall someone's name, for example, means that information about the person is already represented within the MTRF. Thus, the present MTRF state will either be the appropriate one to elicit the needed report from the cortex or the actual elicited cortical reports will drive the

MTRF into the appropriate state. Should this sequence of events not occur, we observe a failure to recall the information. Such failures are to be expected occasionally because of the heuristic nature of the recall mechanism.

The analysis of recall just offered suggests that the retrieval of a memory only occurs in situations where the MTRF has already been driven into an appropriate state by the various signals that impinge upon it. The MTRF does not perform a search function. The inputs to the MTRF, in effect, supply the address for the next memory that will be called up. The context within which a memory is laid down therefore becomes the address. If a sufficient approximation to that context occurs again, as a result of the recall of another memory or due to sensory inputs, then the MTRF must necessarily call up the memory laid down when the context originally occurred. Thus, memory may be viewed as a mapping of MTRF states onto the cortical "file." The brain may be said to employ a search routine as a part of its memory mechanism only to the extent that a recalled memory, in driving the MTRF into a new state, can lead to the recall of another and sometimes more relevant memory.

A simple example may be helpful. Placing one's hand on a very hot stove may establish a memory of the event. The next time one approaches the stove, visual inputs (and perhaps other sensory cues) will drive the MTRF into a state similar to the state it was in when the hand was burned. The cortically stored memory of the previous injury will therefore be activated by MTRF signals (which will be similar to those sent to the cortex during the original experience) or by the direct sensory signals to the cortex, or both, and a cortical report will therefore reach the MTRF in time to inhibit the hand that might otherwise reach out again. The inhibitory nature of the cortical report (memory) is a consequence of the fact that pain ensued when the report was delivered to the MTRF originally, and thus reinforced inhibitory transmittances will dominate when the signals reach the MTRF again. In those cases where the inhibitory transmittances don't yet dominate, the painful act will be repeated a sufficient number of times to bring the inhibitory condition about.

In the past, the distinction between that stored information which constituted memory and that which constituted learning has been unclear or has not been made at all. The following definitions represent an attempt to distinguish the two. Memory is information that is acquired with cortical reinforcement supplied by the MTRF attention mechanism. Learning is information that is acquired with reinforcement supplied by sensory, pain fiber, or LSII activity. Memories serve the purpose, when transmitted to the MTRF, of modifying this structure's state. Memories must be stored on the input side of the MTRF, but learning can be stored on both the input and output sides. When stored on the output side, learning serves the purpose of amplifying MTRF motor commands.

The dynamic processes involved in memory and learning are hypothesized to be essentially the same. Only the terminology differs. When MTRF-generated facilitation impinges upon sensory or memory cortex, we term it attention. When the facilitatory signals are directed toward motor cortex, we term it a voluntary act. In the case of memory, the MTRF signals sent to the cortex may be considered to represent a question; in the case of learning, they represent a command. If a question or command is inappropriate, it will result in pain (inhibition of the MTRF) that will reduce the facilitatory signals being delivered to the cortex, as well as causing an increase in inhibitory transmittances associated with the generation of the question or command. Since facilitation will not be forthcoming from the MTRF, recurrent inhibition will cause recently active cortical neurons to undergo increases in their inhibitory transmittances, insuring that the recent response will be less likely to occur in the future. When the appropriate response occurs, the MTRF facilitation remains and may be enhanced, recently active excitatory transmittances increase, and thus the response is positively reinforced.

In the case of learned motor skills, when cortical excitatory transmittances become large enough, the direct signals that the cortex receives from the sensory tract will, by themselves, be sufficient, or nearly so, to generate the appropriate responses. At this point in the learning it becomes necessary for the MTRF to provide only a more generalized kind of facilitation to the cortex in order to maintain the appropriate activity. No longer having to issue the highly specific commands that were required during learning, much of the MTRF thereby becomes available to process other information, and the learned motor activity moves to a preconscious level. What has happened is that the highly specific facilitation supplied by the MTRF has been replaced, after learning, with highly specific facilitation (made so by selective transmittance increases) from the direct sensory fibers to the cortex. This explains why our awareness of activities decreases with practice of them, it becoming necessary eventually for the MTRF to issue only the generalized command (a form of facilitation) following which the learned response is executed.

In this discussion of the neocortex, it ought to be noted that this structure contains innate as well as acquired information-processing mechanisms. In the case of the cat's visual cortex, for example, Hubel and Wiesel (1965) have shown that at least up to the level of the striate cortex, the functional organization is genetically specified. Thus, when a new-born kitten first interrogates the visual cortex, it is receiving more than "raw" sensory inputs. Presumably, innate preprocessors of this kind also occur in other cortical areas.

One question upon which we have not speculated is that regarding the direct relationship of the LSH to the neocortex. Specifically, does the LSH directly reinforce the cortex in the same way that it has been hypothesized to directly

reinforce the MTRF? Anatomical evidence suggests that there are direct connections, at least in the case of frontal cortex which receives inputs from the medial forebrain bundle (Stein and Wise, 1971). Other questions can be posed. Do some of the various specialized commands issued by the LSH for the maintenance of homeostasis go directly to the cerebral cortex or does the motor arousal associated with a motivated state, for example, result from signals sent indirectly via the MTRF? Does the neocortex supply sensory and memory inputs directly to the LSH? It seems that direct interactions between the MTRF and the neocortex are likely to be important. However, at this point, specific conjectures will not be offered regarding these issues. Such decisions are not essential to the development of other aspects of the present theory.

4.3.4 VECTOR SPECIFICATION OF THE BRAIN STATE

If we wish to characterize the state of a nervous system in a concise manner, given that the neural connectivity pattern has already been specified, we might do so with the specification of two vectors. The MTRF-state vector has already been defined. This vector specifies the dynamic state of the command and control center and thereby displays our state of conscious awareness. We might define a second vector and term it the weight or transmittance vector. This vector will have one component for each synapse in the nervous system, each component specifying the corresponding synaptic transmittance. The vector is therefore a way of representing the state of memory and learning of the organism. If the human nervous system contains 10^{12} neurons and the average neuron receives perhaps 10^4 synapses, then the transmittance vector for man will be a 10^{16}-tuple. (With this number of synapses available, it is interesting to consider that it would be possible for the adaptive process of the brain to modify ten million synaptic transmittances every second for an average human lifetime without ever having to modify the same synapse twice. This measure of the information coding capacity of the human brain provides an upper bound in that every synaptic transmittance is assumed to be modifiable, which may not be the case.) In time, with the accumulation of memories and the occurrence of learning, the transmittance vector comes to represent an increasingly accurate model of man's environment, his relationship to that environment, and of man, himself. Utilizing the information contained in this vector, it eventually becomes possible for the MTRF to make accurate predictions, to discover new relationships among those relationships already known, and to seek answers to questions regarding its own fundamental nature.

4.4 Some Further Observations

4.4.1 MEMORY AND LEARNING

One of the characteristics of brain function that has seemed most remarkable has been the distributed nature of memory and learning within the cortex.

This feature is understandable in light of the global organization proposed above. The MTRF is continuously broadcasting signals to large portions of the cortex. In order to interrogate a memory or initiate a learned response, the MTRF must at least approximately reproduce the signal configuration employed when the memory or learned response was initially established. No aspect of the proposed mechanism makes it necessary for the MTRF to utilize contiguous cortical neurons when adaptation occurs. When a memory is retrieved or a motor act is executed, it would appear to be as easy to signal a widely distributed collection of neurons as it would a localized collection. In fact, with the kind of global organization proposed, a localized memory might be difficult to obtain. Thus the proposed storage and recall mechanism makes plausible the distributed nature of memory and learning.

That the storage may be redundant is also plausible. In effect, the function of the cortex, upon receiving a signal configuration from the MTRF, is to deliver a set of impulses that will either produce an appropriate motor response or drive the MTRF into a new state that represents recalled or sensory information. When a memory is established or learning has occurred, there is no reason to believe that only the required minimum number of cortical neurons will have their synaptic transmittances modified. In fact, it may be that large redundancies are difficult to avoid. Thus, the continued successful function of the brain is plausible in spite of the deaths of thousands of neurons each day, or in spite of the localized destruction of tissue in the case of injury or disease.

It is also possible to understand the recovery of function observed, especially in a younger animal, after extensive damage has been incurred by its nervous system. Due to the localized neural process of heterostatic adaptation, when the structure of the nervous system is altered due to damage, the individual neurons that survive will readjust, seeking to once more maximize the amount of polarization they can obtain from the new signal configurations they will be receiving. This readjustment and the observed recovery of function are hypothesized to be one and the same process. Thus, it is suggested that the key to understanding the plasticity of neural tissue is an appreciation of the fact that each neuron, in a very real sense, is "on its own." Heterostatic adaptation is a localized process where each neuron is wholly responsible for modifying its own behavior. It is important to realize that the behavior of the total organism, no matter how complex and effective, is a by-product of the individual neuron's efforts, not the neuron's ultimate goal.

Concerning the question of the difference between short-term memory (STM) and long-term memory (LTM), it is hypothesized that STM consists of active feedback loops between the MTRF and other structures, principally the cortex. STM is therefore reflected in the MTRF state vector. LTM, on the other hand, is hypothesized to consist of the modified transmittance values of cortical and other neurons and thus is reflected in the transmittance vector.

It should be noted at this point that the feedback relationship that has been proposed between the neocortex and the MTRF may be expected to also exist to some degree between the MTRF and other brain structures. Memory and learning are not the exclusive domain of the neocortex, although this structure is clearly highly advanced in regard to these functions. This issue is raised here because the extent of feedback loops between the MTRF and other structures may correlate with their observed epileptogenicity. It was noted in Section 3 that Goddard (1967), using a stimulation technique that he developed, found that an epileptic focus could be established in the amygdala within 15 days, but none was established in the reticular formation after 200 days of stimulation. Goddard's stimulation of various sites within the brain may have resulted in the activation of MTRF-mediated feedback loops that, in turn, supplied a further polarizing bias to the stimulated area. If this kind of reinforcing feedback is occurring during the development of an epileptic focus, epileptogenicity will, to some extent, be proportional to the extensiveness of feedback loops between the stimulated area and the MTRF.

The extensive feedback loops that occur between the two cortical hemispheres may contribute to the epileptogenicity of the cerebral cortex. That such loops are involved in some cases of epilepsy (but not necessarily in the establishment of epileptic foci) is suggested by the fact that surgical separation of the two hemispheres has produced marked improvement in severely epileptic patients (Sperry, 1966, pp. 298-299; Gazzaniga, 1970).

Are painful memories most easily forgotten? This is reasonable if the LSH reacts to the memory as it did to the original experience. The LSH may be expected to negatively reinforce the MTRF activity involved in the recall and thus this MTRF state is less likely to occur again. Forgetting, in this case, refers to a loss of the ability to retrieve the cortically stored memory. With regard to the general question of whether forgetting is due to interference or decay of memory traces, the neuronal model and global brain organization proposed renders interference highly plausible but offers no basis for the occurrence of memory trace decay. This appears to be consistent with the psychological evidence on the question.

With respect to language-dependent information processing, split brain data (Sperry, 1966; Gazzaniga, 1970) indicate that the integration of information from the left and right sides of the body is accomplished at the cortical level and not at the level of MTRF. When the MTRF is aware of events associated with the left side of the body, as demonstrated through speech or writing, it apparently obtains this awareness by interrogating the left dominant hemisphere that, in turn, receives reports from the right hemisphere. Sectioning of the corpus callosum, the anterior and hippocampal commissures, and the massa intermedia, done for the purpose of controlling severe epileptic seizures, renders a person unable to communicate information via speech or writing if the information is available only to the left side

of the body or to the left half of the visual field. Furthermore, the right side of the body is no longer able to act on the basis of information available only to the left side. All of this evidence points to a very high degree of coupling between the dominant hemisphere and that portion of the RF which is the seat of conscious awareness.

4.4.2 PLEASURE AND PAIN

Hilgard (1969), in a paper entitled "Pain as a Puzzle for Psychology and Physiology", raised four fundamental questions regarding the nature of pain. We will try to answer them now on the basis of the proposed theory.

"Is pain a sensory modality?" It is not, no more than pleasure is. Both the experience of pleasure and that of pain can, at times, be highly correlated with activity within sensory tracts or pain fibers, when these are delivering excitation or inhibition to the MTRF. At other times, however, the source of the excitation of inhibition is the LSH and then the subjective experience is that of diffuse and internally generated sensations that we describe as emotions. The intensity of pleasure or pain has already been hypothesized to be approximately proportional to the delta value [see Eq. (18)]. Also, the specific subjective experience, whether of vision, audition, touch, joy, grief, anger, fear, etc., has been hypothesized to depend on the specific configuration of depolarizing and hyperpolarizing events occurring within the MTRF. It is hypothesized that we place subjective experiences in the same or different categories depending on the extent of similarity between the respective MTRF state vectors. In terms of the hypothesized reinforcement and drive networks, the former network sets the emotional tone while the latter network accomplishes specific computations of motor commands (and also the retrieval of memories to be utilized in computing future motor commands).

Hilgard notes that "any stimulus can qualify to produce pain if it is intense enough." This fact suggests that, with increasing intensity, all stimuli eventually become inhibitory with respect to the MTRF.

The relief from intense pain that is obtained with frontal lobe operations suggests that feedback loops between the MTRF and the frontal lobe have the effect of amplifying inhibition occurring within the MTRF. The relief from anxiety (assumed here to be an inhibition-dominant MTRF state) that is obtained after prefrontal lobotomy further suggests an important relationship between the frontal lobes and MTRF inhibition.

"Are there any satisfactory physiological indicators of pain?" None have been found, as Hilgard notes, and none are likely to be discovered. Within the MTRF, the complex interplay between excitation and inhibition that corresponds to the subjective experiences of pleasure and pain should not be expected to correlate reliably with any physiological indicator.

"Where is the pain that is felt?" It is within the MTRF, along with everything else of which we are said to possess conscious awareness. The projection to a location on the body is accomplished with information that is probably obtained from the regular sensory channels.

It is interesting to note that in the case of phantom-limb pain, the projected location doesn't even exist, although it may have had to at an earlier time since the projection of a sensation to a location on the body may have to be learned. The phenomenon of phantom-limb pain may arise because the amounts of excitation and inhibition being delivered to the MTRF from a given part of the body are normally in a state of balance – a state that apparently can be disturbed when the peripheral nervous system is altered.

"How account for the great individual differences in felt pain?" This is understandable when one realizes that pain is not something as narrow in scope as a sensory modality. Pleasure and pain are the two fundamental currencies with which all of the nervous system's transactions are carried out. The great individual differences in felt pain reflect the genetic and experiental differences among people. It is reasonable that many aspects of nervous system function should influence the subjective experience of pain, because we are really talking about a phenomenon as pervasive as inhibition. The variability among people in their reaction to pain is parallelled by the variability of results obtained by the neurosurgeon when he operates to relieve pain. We can now better understand why this type of surgery does not always produce clearcut, reliable results. As with the sources of inhibition, the sources of pain are multiple and complex.

Having reviewed the questions raised by Hilgard, several other observations may be made before we leave the subject of pleasure and pain. Fear may be considered to be the memory of inflicted pain. The memory can be of genetic origin (in which case we are using the term "memory" in a much wider sense than usual) or it may be acquired. An animal will "freeze" in a situation that produces extreme fear. This may be a consequence of massive inhibition occurring within the MTRF. Also consistent with the proposed theory, pleasurable circumstances reveal an excited MTRF – people feel like "dancing in the street."

On the battlefield (and the playing field) men can be observed with severe injuries and yet they may experience no pain. The inhibition being delivered to the MTRF via the pain fibers, as great as it is, is overwhelmed by the massive excitation simultaneously delivered by the LSH in the heat of the battle. This is a circumstance where the two terms in Eq. (18), Q(D) and Q(H), are both very large. A positive delta value results because the fight mechanism of the LSH is assumed to supply large amounts of excitation. (This assumption that the innate LSH fight mechanism is a source of excitation and therefore of pleasure for the MTRF may also in part explain man's tendency to undertake wars.) If the flight mechanism

takes over and delivers inhibition to the MTRF, causing Q(H) to increase, or the contest ends causing Q(D) to decrease, then the delta value becomes negative and pain is experienced. Roughly speaking, it can be seen that Eq. (18) is to the whole animal what Inequality (7) is to the neuron.

A final observation concerning pain is that we don't habituate to it. This is consistent with the fact that no mechanism has been observed that relates to the inhibitory neurons in the way that the feed-forward and recurrent inhibition underlying habituation relate to excitatory neurons. It can be seen that both the ability to habituate to pleasurable stimuli and the inability to habituate to painful stimuli have survival value.

4.4.3 NATURE OF MAN

The present theory suggests a very simple and very old answer to a basic philosophical question. What is the fundamental nature of man? Man is a hedonist (a special form of the heterostat). Probably most, if not all, other forms of animal life are hedonists also. This can now be seen to be a consequence of the fact that the neurons of which nervous systems are constituted are hedonists.

The conclusion that man is a hedonist has been rejected by many in the past for three scientific reasons: First, definitions of pleasure and pain in other than behavioral terms were lacking. Second, no underlying neurophysiological mechanism was apparent that would support such a thesis. Third, man's behavior appears to be truly altruistic at times (discounting the occasions when apparent altruism is only a ploy). The proposed neuronal model suggests answers to the first two objections. To meet the third objection, that relating to man's altruism, a new hypothesis must be introduced.

It is proposed that the LSH's capacity for distinguishing between self and other is limited. Specifically, it is proposed that man's LSH is a primitive brain that (in conjunction with the MTRF) simply regards men as "others" if the fight-fright-flight drive centers are active, but in all of the remaining situations, the LSH fails to distinguish between other and self. This conjecture, which will be termed the LSH hypothesis, is supported by evidence drawn from human behavior. For example, man's aversion to killing his fellow man decreases as the relevant sensory stimuli diminish. The strong inhibition is present only when the LSH, driven by appropriate sensory inputs, can (mistakenly) interpret the situation as self-injurious. A bomber pilot can knowingly produce immense suffering and death from the cockpit, whereas he might well be rendered psychotic if he had to accomplish the same amount of destruction of life while he was face-to-face with his victims. Remove the relevant sensory input and the inhibition against inflicting pain or death upon another human being diminishes. (Usually, it does not reduce to zero because the cortical functions of memory and learning will provide approximations to the relevant sensory inputs when they are no longer actually present.

In addition, most cultures have, to some degree, conditioned their members not to harm their fellow man.)

Further evidence of the inability of the LSH to make the self-other distinction is provided by an observation that Lorenz (1970) has discussed. Man finds it increasingly difficult to kill animals as their external resemblance to him increases. In order of increasing difficulty, Lorenz mentions fish, frogs ("it is the human-like arms and legs of these creatures that render their slaughtering so odious, though from the point of view of neural and 'mental' development they are vastly inferior to the higher fish"), dogs, and monkeys.

That the LSH tends, except under fight-fright-flight circumstances, to incorporate the other into the self is also evidenced by the fact that the amount of seemingly altruistic behavior exhibited by an individual is a function of the extent to which he is acquainted with the other person or animal. Lorenz (1970) has noted that researchers experience increased difficulty in sacrificing a laboratory animal if the animal has been known to them for a relatively long period of time. The difficulties increase further if the animal has required considerable care.

The present context is perhaps an appropriate one in which to consider the psychological concepts of love and hatred. Hatred would appear to be a psychological state in which the fight-fright-flight drive centers are active and thus the LSH is capable of making the self-other distinction. Love, on the other hand, is hypothesized to be a psychological state in which the LSH does not make the self-other distinction. If this is true, then "Love thy neighbor as theyself" assumes the nature, in part, of a definition.

The innate decision-making processes of the LSH are supplemented in man by much learned behavior. The question arises as to whether the greater ability of man's highly educated cortex to assist the MTRF in making distinctions between self and other has increased man's aggressiveness as compared to that of the other primates. It would seem that learning can either support or oppose the innate processes of the LSH. If, for example, a person is taught prejudicial attitudes toward another group of people, the person will perceive less of a resemblance between himself and the members of the other group. The increased ease in making the self-other distinction increases the likelihood that the person will behave detrimentally toward those against whom he has been prejudiced. On the other hand, an educational process that reveals the similarities between all men will support and strengthen behavior that is a consequence of the LSH's identification of self and other.

4.4.4 EPISTEMOLOGY

The neuron is a cell in pursuit of excitation. At the level of the whole man, this neuronal goal reveals itself, in part, as the pursuit of knowledge or, more

generally, the pursuit of stimulation. Of crucial importance in this process is the habituation mechanism. Without it, redundant stimuli could satisfy the neuron. Novel stimuli would not be sought, knowledge would not be acquired, and the survival of the species would be unlikely.

In the context of the present theory, knowledge is seen to be a collection of causal relations encoded in the form of synaptic transmittance values. Transmittances are measures of the liklihood that excitation or inhibition will follow if the neuron responds when the associated synapses are active. Thus, at the level of the neuron, information is encoded in the simplest of forms. Impulses reach the neuron as weighted advice to fire or not to fire – nothing more. It is only from the viewpoint of the whole brain that strings of excitatory and inhibitory impulses take on more complex meanings.

4.4.5 GENERALITY

Questions as to the generality of the proposed theory have not been considered up to this point. While answers cannot be supplied, some of the questions should be noted. Are only some nervous systems structured as heterostats? Only some neurons? Are nonneural cells heterostats? If so, is this view a route to a better understanding of cells in general? Related questions suggest themselves. Are plants and plant cells heterostats? Does the cancer cell result from excessive positive feedback or inadequate negative feedback? What is the detailed nature of the heterostatic mechanism in neurons? Does the entire neuron act as a single heterostat as we have assumed here, or do limited areas of the soma and dendritic field act independently of one another from the standpoint of adaptation? For example, could it be that recently active synapses are reinforced only by other active synapses in the immediate area rather than by any synapses anywhere on the neuron? Does the genetic apparatus encode modified transmittance values and then insure that these values are maintained? If so, might the recovery of function observed in retrograde amnesia result from a DNA-directed repair process in which transmittances are reset to the values they had assumed before injury? Might the sequence of transmittance vectors that originally occurred be approximated during the recovery process (a mnemonic version of ontogeny recapitulating phylogeny), thereby explaining the order in which memories return?

5. SOCIOLOGY

If neurons are heterostats and, furthermore, if organized collections of neurons (nervous systems) are heterostats, then it seems reasonable to hypothesize that organized collections of nervous systems (societies) will also be heterostats. The observed behavior of societies would seem to support such a hypothesis.

Actually, the list can be broadened. It is anticipated that the class of heterostatic systems includes neurons, assemblies of neurons, whole nervous systems, families, neighborhoods, cities, regions, and nations.

A system is a heterostat if it is organized in such a way that it seeks to maximize a specific internal variable. In the case of social systems, that variable is hypothesized to be the total amount of pleasure being experienced by the system's members (pain, in this context, constitutes negative pleasure). Actually, to suggest that a nervous system or a social system functions as a single, integrated heterostat is to simplify the matter. It will usually not be the case that an entire system is organized in this way. Normally, some part of the system occupies a position of control and it, functioning as a heterostat, imposes constraints on the remaining subsystems. These subsystems, in turn, will also function as heterostats, but in doing so, they will be subject to the imposed constraints from the controlling subsystem. Thus, in general, it is hypothesized that nervous systems and social systems are organized as collections of interacting heterostatic subsystems. Furthermore, a subsystem itself, rather than being a single integrated heterostat, may consist of a collection of heterostatic "subsubsystems" where one of them is controlling and the remainder are subject to imposed constraints.

To consider an example, it was hypothesized earlier that the MTRF is the controlling subsystem of the brain and that, furthermore, it is a heterostat. That is to say, it seeks to maximize the amount of polarization being experienced by MTRF neurons; it does not seek to maximize the amount of polarization being experienced by all of the neurons within the brain. Thus, the brain is a heterostatic system in the sense that its controlling subsystem is a heterostat. In general, we will define a heterostatic system as one in which either (a) the heterostatic variable, μ, is defined with respect to all of the subsystems taken collectively, that is, the system functions as a single, integrated heterostat, or (b) μ is defined with respect to the controlling subsystem only, as in the case of the MTRF of the brain. The first type of system will be termed an integrated heterostat; the second type will be termed a limited heterostat. An integrated heterostat is organized such that whenever the μ of each of the subsystems is maximized, so is the μ of the system taken as a whole. In the case of social systems, this is perhaps the necessary and sufficient condition for the achievement of a perfect democracy. Alternatively, a society that assumes the form of a dictatorship provides an example of a limited heterostat.

The brain appears to be organized as a limited heterostat. Does this evolutionary outcome imply that a limited heterostat represents a more stable form of organization than the integrated heterostat? Such may be the implication at the neural level, but an extrapolation to the societal level may not be valid. However, it is a question that must be considered if one desires to compare, for example, the expected life spans of democracies and dictatorships.

It can be seen that the present theory offers the possibility of a unified systems approach to the study of adaptive networks, whether they be of the neural, social, or man-made variety. It should now be possible to construct a single heterostatic model of an adaptive network such that the model will simultaneously possess relevance for the neurophysiologist, psychologist, sociologist, and cybernetician. This model will consist, essentially, of a set of nested heterostatic subsystems. Each subsystem will receive two types of inputs, one type to be sought, the other to be avoided. The inputs will be variously described as excitatory or inhibitory, appetitive or aversive, pleasurable or painful, and perhaps positive or negative, depending on the particular application. A generalized version of the neuronal process of heterostatic adaptation will provide the maximization mechanism for each subsystem. More will be said about this model in Section 6.

5.1 Comparing Social Systems With Nervous Systems

If, indeed, brains and societies are fundamentally one and the same kind of system, then a comparison of their behavioral characteristics should be instructive. For a start, we might consider that brains exhibit three properties that have been of substantial interest: brain functions are distributed, redundant, and (sometimes) recoverable after extensive damage. It can be seen that societal functions possess these three properties also. The knowledge and skills contained within a society are, in general, not highly localized. Furthermore, just as it is normal for thousands of neurons to die each day while the nervous system, utilizing its redundant structure, continues to function, so too does the continual loss of members go on in a society without rendering the system unworkable.

With either brains or societies, however, if the damage to the system is extensive, so that redundancy of function cannot adequately compensate, then a process of adaptation will, in some cases, permit the function to be recovered. In the case of the brain, it was hypothesized that this recovery occurs because the surviving neurons are subjected to a new set of input conditions and thus they modify their transmittances so as to once more tend to maximize the amount of polarization they are experiencing. It was suggested that these transmittance changes and the recovery of function are one and the same process. The mechanism would appear to be the same within a society. When a substantial portion of a social system is destroyed due to a natural disaster or due to war, the surviving members can be seen to respond to their new input conditions in such a way as to once more tend to maximize the amount of pleasure they are experiencing. Often, to accomplish this, they must learn to perform the function previously accomplished by that portion of the society that has been destroyed. We have marveled at the brain's ability to recover functions after extensive damage. The same process, occurring within a society, produces less awe because the mechanism is more apparent. We can

generalize here. Probably one of the best ways to gain an appreciation of how neurons interact is to observe how people interact within the various social systems of which they are a part. If one adopts this viewpoint, the properties of brains will become less mysterious.

It may be noted that distributed functions, redundancy, and recoverability of function through adaptation are not always adequate protections either in the case of brains or societies. Certain heterostatic subsystems are of critical importance and their loss results in great and permanent damage to the total organization. The MTRF of the brain and the capital cities of some nations serve as examples of such subsystems.

5.2 Restructuring Social Systems

In the context of the present theory, it can be seen that neurons do not exist for the sake of nervous systems and people do not exist for the sake of social systems. People will do what is best for their society only when it coincides with what is best for themselves. Therefore, if a society wishes to achieve a particular result most efficiently, it must assume a structure that provides for a coincidence of the interests of the individual and the society.

In what ways can present societies be restructured in order to enable them to better meet the needs of their members? Some p[illegible]e answers to this question are suggested by observing ways in which so[illegible]ties and brains differ in their organizational structures.

First of all, governments have a tendency to make their decisions at the top and to execute them at the bottom. Nervous systems, on the other hand, appear to make increasingly detailed decisions all along the line, as one moves from the upper to the lower levels in the control hierarchy. In addition to establishing general goals, the upper control levels in the brain appear to regulate behavior by delivering positive or negative reinforcement based on the results obtained, not based on whether a detailed plan of action passed down from above was followed. Is the evolutionary process suggesting to us that decentralization is the most effective organizational structure for a society? One cannot be sure of such an extrapolation, of course, but we should carefully consider advice that has survived millions of years of selection.

A second observation that may be relevant is that, in reinforcing neuronal behavior, brains appear to employ a mix of positive and negative reinforcement. On the other hand, governments tend to employ only negative reinforcement. The services provided in return for taxes do not constitute positive reinforcement because our receipt of them is not, in general, contingent upon our behavior. Essentially, the only behavior-contingent response is the punishment received when one has broken a law. In contrast, note that many private organizations do employ

a mix of positive and negative reinforcers and the result appears to be more effective organizations. Promotions, pay raises, recognition, prestige, and power are highly effective positive reinforcers. The denial of these rewards, or even more extreme, the loss of one's job, provides effective negative reinforcement. Such a balanced set of reinforcers does not confront the citizen in his relationship to his government. Rather, the name of the game for many simply becomes that of avoiding negative reinforcement. Skinner (1971) has recognized, perhaps better than anyone else, the need for us to rectify this situation by building social institutions that implement a more suitable variety of reinforcers.

Finally, concerning the subject of international cooperation, the present theory suggests what many already know. A nation will not willingly surrender power to an international organization unless a point is reached where it is in the nation's best interest to do so. The nature of heterostats is such that they do not cooperate as parts of larger heterostatic systems unless they have more positive than negative reinforcement to gain by doing so. Such was the case for neurons and thus heterostatic brains emerged. Such was the case for people and thus heterostatic societies emerged. Will such be the case for nations and will, therefore, a single heterostatic earth-society eventually emerge? The hazards of gross ecological imbalance and the prospects for nuclear annihilation are potential negative reinforcers of great power. The benefits of cooperation could be many. It would seem, therefore, that a single heterostatic system consisting of the earth's nations may one day be realized.

6. CYBERNETICS

Cybernetic research into adaptive and intelligent systems has usually assumed one of two forms. The "neural net" approach, characterized by a microscopic and neurophysiological orientation, grew out of the early work of investigators such as Rashevsky (1938) and McCulloch and Pitts (1943). The alternative approach of heuristic programming, with its macroscopic and psychological orientation, developed out of the work of investigators such as Newell, Shaw, and Simon (1957, 1958) and Samuel (1959). Some cyberneticians, such as Weiner (1948) and Ashby (1952), have influenced researchers of both orientations. In this section, we will examine the relationship of the present theory to both approaches and will then consider possible directions for future research.

6.1 Neural Nets

In the construction and simulation of adaptive networks for experimental purposes, usually one of two types of connectionistic neuronal models has been

employed. In one type (which will be designated as Type I in the subsequent discussion), it is assumed that the repetitive activation of a synapse, when this contributes to the firing of the recipient neuron, is the basis for increasing the effectiveness of the synapse. Hebb (1949) adopted this assumption in developing his theory of cell assemblies. Hebb's theory provided the basis for further research by Rochester et al. (1956), Milner (1957), Good (1965), Finley (1967), and Sampson (1969), among others.

Another and more frequently assumed connectionistic neuronal model was first employed in adaptive networks developed by Minsky (1954) and Farley and Clark (1954). With this type of model (which will be designated as Type II in the subsequent discussion) alterations of the synaptic transmittances of a neuron occur as a result of reinforcement signals received by the neuron after it responds. The neuronal model proposed in the present theory belongs to the Type II category. Other variations of the Type II model have been employed in the extensive investigations of Perceptrons initiated by Rosenblatt (1957, 1962) and also in studies by Selfridge (1959), Palmieri and Sanna (1960), Gamba et al. (1961), Widrow (1960, 1962, 1963), Brain (1963), and Kaylor (1964), among others. For the reader who desires more than the very brief historical review presented here, the bibliographic notes of Nilsson (1965) and Minsky and Papert (1969) provide a comprehensive overview of the literature of neural net studies.

The following are some of the differences in point of view and assumptions between the present theory and previous work in neural net research:

1. The problem of understanding adaptive networks has usually been viewed from the top down rather than from the bottom up. That is to say, the tendency has been to think in terms of the system goals and then ask what network structure and mechanisms would permit these goals to be realized. The viewpoint of the heterostatic theory is the opposite. A goal is assumed for the individual network elements and then the behavior of the total system is analyzed in terms of this elemental goal. From a theoretical point of view, the two approaches can be equivalent, of course. However, the perspective that results is in fact very different, depending on whether one focuses on the system goals first and the elemental behavior second or on an elemental goal first and the consequ system behavior second. The heterostatic theory emphasizes the view that the individual network elements are goal-seeking systems.

2. The neuronal model adopted in the present theory assumes an adaptive process dependent on sequential events, as do other models belonging to the Type II category. Previous theories that invoked a Type I neuronal model assumed an adaptive process dependent on simultaneous events.

3. Previous researchers who assumed a Type II model viewed the neuron as the recipient of three or four different signal types, these signals being excitation and inhibition along with positive and/or negative reinforcement inputs. The assumption that the reinforcement signals were of a special nature, different from the excitatory and inhibitory inputs, created a wide division between the "teacher" (the source of the reinforcement signals) and the remainder of the adaptive system and the environment. When the division is eliminated, as in the heterostatic theory, then not only may a special teacher – the LSH, for example – provide reinforcement, but the environment may directly do so also. Sensory inputs may be positively reinforcing (excitation delivered to the MTRF) or negatively reinforcing (pain, that is, inhibition delivered to the MTRF). In addition, the signals transmitted between neural structures may be reinforcing as in the case of selective MTRF facilitation of the cortex resulting in our remembering that to which we attend.

4. The proposed heterostatic theory assumes that nervous systems employ a statistical adaptive process, a process in which the chance occurrence of appropriate or inappropriate behavior is positively or negatively reinforced, respectively. Such a mechanism eventually produces a network that, to some degree, models the organism and the environment. As the model becomes increasingly refined, the statistical aspect of the adaptive process becomes less apparent and behavior appears more directed. In contrast to this type of mechanism, much neural net research has assumed a deterministic adaptive process. This assumption grew out of the discovery of the Perceptron convergence theorem (Rosenblatt, 1960) which provides an algorithmic adaptive procedure. Such a mechanism has many advantages. However, the algorithm applies only to single-layered adaptive networks. Much subsequent research has failed to produce a truly viable deterministic adaptive mechanism for the more general case of the multilayer network. The central problem in the general case is that of establishing what any given network element <u>should</u> be doing when the system behavior is inappropriate. If such could be established, then the above-mentioned single-layer algorithm could be applied to steer an element's behavior in the appropriate direction. The deterministic approach to multilayer adaptation can be seen to be very different from the statistical approach in which a system selectively reinforces whatever behavior occurs, without attempting (at least in the beginning) to extrapolate beyond that behavior.

The discovery of the convergence theorem encouraged a search for a global adaptive mechanism – a mechanism possessing a sufficient overview of a network such that decisions could be made as to what an appropriate response would have been for a given element. It has been found that, in general, such a global mechanism cannot be simple because most of the ouputs of individual

elements in a deep network bear a highly indirect relationship to the final output of the system. Therefore, when a final output is judged to be inappropriate, it is very difficult to translate this into a judgment regarding the appropriateness of the outputs of each of the individual network elements.

It should be noted that the discovery of the convergence theorem did not cause Rosenblatt to seek a global adaptive process. He continued to believe that a local process was most appropriate, an assumption that was adopted in formulating the present theory.

If the heterostatic theory is assumed to be correct, then it can be seen that an unfortunate pairing of concepts has occurred in past neural net research: the correct category of neuronal models (Type II) became strongly associated with the incorrect (deterministic) type of adaptive process; likewise, the incorrect category of neuronal models (Type I) was always associated, because of its nature, with the correct (statistical) type of adaptive process. Seldom was the proper category of neuronal models and adaptive processes studied in conjunction. Actually, this criticism has to be qualified. If it is valid, it can only be so with respect to the problem of modeling and understanding nervous systems. With respect to the problem of understanding adaptive networks in general, it has perhaps been worthwhile to investigate a variety of other combinations of neuronal models and adaptive processes, apart from that combination which living systems apparently employ.

6.2 Heuristic Programs

In recent years, heuristic programming research has yielded increasingly powerful information processing systems. For an introduction to the history of this research, it is recommended that the reader consult the collections of papers edited by Feigenbaum and Feldman (1963) and Minsky (1969). In addition, a recent state-of-the-art survey by Nilsson (1971) includes bibliographical and historical notes.

The relationship of the heterostatic theory to heuristic programming research is not as easily defined as the relationship of the heterostatic theory to neural net research. Heuristic programs are macroscopically oriented and also are often developed using heuristics other than those that offer parallels with living systems. However, the ultimate objective remains that of achieving intelligence, a notion that has usually been defined, at least in part, in terms of the capabilities of the human brain. It is therefore of interest to consider heuristic programs in the context of the present theory.

Two central questions in heuristic programming research are: (1) how should subprograms be interrelated to maximize the effectiveness of the total system and (2) how should a system be structured to achieve generality, so that it can operate in a variety of environments with a variety of goals? These questions appear to

have the following answer in the case of neural networks: Translate all information for transmission into a pair of binary codes and then employ subsystems that seek to maximize the difference between the amounts of each of the two types of codes being received. One code is to be sought by the subsystems and the other to be avoided. In the case of the brain, the code that is sought employs excitatory impulses and that which is avoided employs inhibitory impulses. Perhaps the coordination of subprograms and system generality can be achieved in heuristic programs with a similar approach. To some extent, this is being done in robotics research [for example, see Andreae (1963) and Doran (1968)] where a robot's problems have been formulated in terms of variables similar to those of pleasure and pain (see review by Ernst, 1970).

One other comment relates to both neural net research and heuristic programming studies. The brain appears to be organized as a set of nested heterostats, such that the organizational structure is similar at every level from that of the individual neuron to that of the whole nervous system. As noted in Section 1 and as others have recognized, a hierarchic structure may offer an important advantage: such networks may be able to increase in size and capabilities by the more or less simple addition of further network elements. With a nonhierarchic structure, such a bootstrapping operation may be much more difficult and, if so, such structures should perhaps be avoided.

6.3 A Basic Building Block

One of the objectives of neural net research has been to obtain an element that could serve as a fundamental building block for the construction of a wide variety of adaptive systems. At one point it was thought that the variable-weighted linear threshold element, in conjunction with the algorithmic adaptive procedure provided by the Perceptron convergence theorem, was such an element. The present theory offers a new candidate element: the heterostat. The elementary heterostat (one that cannot be reduced further to component heterostatic subsystems) is a device that appears to meet the requirements. Such a claim is based on the assumption that it is the heterostatic nature of neurons that is responsible for the information processing capabilities of nervous systems.

An elementary heterostat derives its power from the special way in which it utilizes feedback information. In the case of positive feedback (for example, excitation delivered to a neuron), not only does the input enhance the output but also the input enhances the effectiveness of recently active positive feedback loops. In the case of negative feedback (for example, inhibition delivered to a neuron), not only does the input diminish the output but also the input enhances the effectiveness of recently active inhibitory feedback loops. In this way, the heterostat not only accomplishes short-term modifications in its behavior but it also accomplishes

long-term modifications, thus encoding causal relations and providing a basis for intelligence.

6.4 Future Research

Until a mathematical theory is developed for heterostatic systems, understanding will have to be sought through experimental studies of synthesized networks. Initially, in the study of small networks of elementary heterostats, the dynamics of the systems may be more readily analyzed if the networks do not interact with an environment. This simplification will permit attention to be focused on the interactions among network elements. Such isolated systems might provide adequate models for neural and social networks where all relevant inputs originate in other heterostats. The model will not be satisfactory for any system in which some of the inputs originate in the environment within which the system is embedded.

Beyond the study of isolated systems, the more general problem will be that of establishing the relationship between a heterostatic system's structure and the system's effectiveness in eliciting positive stimuli from an environment while avoiding negative stimuli. (In this discussion, "positive" and "negative" will refer to the categories which, in the special case of nervous systems, would be described as pleasurable and painful, appetitive or aversive, depolarizing and hyperpolarizing, or excitatory and inhibitory.) Also, the question of how a heterostatic system and an environment should be structured if it is desired that the network solve a particular problem will be of immediate practical interest. In all of this work, there will be a need to develop methods for classifying the structures of environments and of heterostatic systems, so that a wide variety of possibilities can be studied efficiently.

The sequence of developments during future experimental studies will necessarily parallel, to some extent, the developments that must have occurred during the evolution of nervous systems. Initially, simple networks of elementary heterostats can be studied interacting with simple environments, perhaps paralleling the evolutionary phase during which simple multicellular organisms interacted with the simple environment of the sea. It will be necessary to provide the synthesized networks with sensory transducers that can convert environmental stimuli into positive and negative inputs. Motor transducers will convert positive and negative outputs into actions affecting the environment. As it becomes possible to synthesize larger and more complex systems, a variety of subsystems will "evolve", including specialized command and control centers, reinforcement centers that will impose a global type of pressure on the elements so that they will adapt in particular directions, drive centers that will supply innate knowledge, and memory centers that will permit the more effective utilization of past experience.

> It is difficult if not impossible to draw a line separating the regulatory behavior of lower organisms from the so-called intelligent behavior of higher ones; the one grades insensibly into the other. From the lowest organisms up to man behavior is essentially regulatory in character, and what we call intelligence in higher animals is a direct outgrowth of the same laws that give behavior its regulatory character in the Protozoa.
>
> H. S. Jennings (1906)

> In attempting to understand the elements out of which mental phenomena are compounded, it is of the greatest importance to remember that from the protozoa to man there is nowhere a very wide gap either in structure or in behavior. From this fact it is a highly probable inference that there is also nowhere a very wide mental gap.
>
> B. Russell (1921)

> To declare that, of the component cells that go to make us up, each one is an individual self-centred life is no mere phrase. It is not a mere convenience for descriptive purposes. The cell as a component of the body is not only a visibly demarcated unit but a unit-life centred on itself. It leads its own life... The cell is a unit-life, and our life which in its turn is a unitary life consists utterly of cell-lives.
>
> C. S. Sherrington (1941)

7. CONCLUDING REMARKS

At the fundamental level of the cell, and consequently at higher levels, life is seen to be a process aimed at achieving a maximal condition. With that as its goal, living systems utilize positive and negative feedback loops in such a way as to accomplish long-term adaptation as well as the regulation of immediate behavior.

Both positive and negative feedback are essential to life's processes. However, it is positive feedback that is the dominant force – it provides the "spark of life." It is the effects of positive feedback that account for observations such as Sherrington's (1941):

> 'Life has an itch to live.' This itch is universal with it. Under the microscope it gives us tiny lives hurrying hither, thither, feeding. Driven, each says almost as clearly as if it spoke, by 'urge-to-live.' The one key-phrase to the whole bustling scene seems 'urge-to-live.'
>
> If, from the microscope, we turn to look at the busy street, with its individuals hurrying hither and thither, hastening to earn, entering restaurants, is there no resemblance to that other, the microscopic scene? What phrase again sums it up? 'Urge-to-live.' And here, because we know it in ourselves, we can read securely into that urge, as of it and with it, an element 'zeal-to-live', 'zest-to-live.' The physico-chemical has here its mental adjunct. Life's zest to live as outcome of life's tendency to increase in bulk. Long the stretch of distance between that microscopic populati and this human one. Yet if a form of mind there be in that micro pic population, though as a germ so remote that no word of ours can duly fit it, is it not probable that that mind is 'zest-to-live' in germ?

The evolutionary process has established an equivalence between that which has survival value and that which is a source of pleasure. Thus, living systems, in pursuing heterostasis, participate in three broad categories of actions:

1. self-preserving behavior (maintenance of homeostasis),
2. species-preserving behavior (reproduction),
3. stimulation-preserving behavior (knowledge acquisition and play).

Recognition that these categories represent subgoals, pursued only as means to an end (heterostasis), renders certain aspects of human behavior more understandable. We can see why it is that man goes on making love and war while researching both subjects. Concerning stimulation-preserving behavior, it is seen that it is not the sometimes postulated "need to understand" that drives man to explore his universe; rather, it is the need for stimulation.

In conclusion, it is seen that complex forms of behavior in living systems derive not from complex underlying mechanisms but from the combinatorics of billions of interacting subunits and from the complexity of environments. A simple perspective emerges. Neurons, nervous systems, and nations are heterostats.

Acknowledgments

Discussions with Rocco H. Urbano during the course of this research were most helpful in clarifying the ideas contained herein. Also, thanks are due to Edmond M. Dewan, Arthur P. Ginsburg, Earl E. Gose, and Anthony N. Mucciardi for critical readings of the manuscript.

The research was accomplished while the author held a National Research Council Postdoctoral Resident Research Associateship supported by the Air Force Systems Command, United States Air Force.

References

Andersen, P., Eccles, J. C., and Voorhoeve, P. E. (1963) Inhibitory synapses on somas of purkinje cells in the cerebellum, Nature (London) 194:740-741.

Andersen, P., Eccles, J. C., and Lyning, Y. (1963) Recurrent inhibition in the hippocampus with identification of the inhibitory cell and its synapses, Nature (London) 198:541-542.

Andreae, J. H. (1963) Stella, a scheme for a learning machine, Proc. IFAC.

Aristotle (384-322 B. C.) Nicomachean Ethics, Loeb Classical Library, Cambridge, Mass., Harvard, 1939, p. 31.

Ashby, W. (1952) Design for a Brain, Wiley, New York.

Ashby, W. (1960) Design for a Brain, Wiley, New York.

Bernard, Claude (1859) Lecons sur les proprietes physiologigues et les alterations pathologigues des liquides de l'organisme, Bailliere 1, Paris, France.

Bindman, L. J., Lippold, O. C. J., and Redfearn, J. W. T. (1964) The action of brief polarizing currents on the cerebral cortex of the rat (1) during current flow and (2) in the production of long-lasting aftereffects, J. Physiol. 172:369-382.

Blackstad, T. W., and Flood, P. R. (1963) Ultrastructure and hippocampal axosomatic synapses, Nature (London) 198:542-543.

Brain, A., Forsen, G., Hall, D., and Rosen, C. (1963) A Large, self-contained learning machine, 1963 WESCON Paper 6.1 VII, Part 4.

Bureš, J., and Burešová, O. (1965) Discussion. Plasticity at the single neuron level, Proc. Int. Union Physiol. Sci. 4:359-364, XXIII Int. Congr. Tokyo.

Bureš, J., and Burešová, O. (1967) Spreading depression and corticosubcortical interrelations in the mechanism of conditioned reflexes, Progr. Brain Res. 22:378-387.

Bureš, J., and Burešová, O. (1970) Plasticity in single neurons and neural populations, Short-term Changes in Neural Activity and Behavior, (ed. by G. Horn and R. A. Hinde), New York, Cambridge University Press, pp. 363-403.

References

Burns, B. D. (1968) The Uncertain Nervous System, Edward Arnold Ltd., London.

Cannon, W. B. (1929) Organization for physiological homeostasis, Physiological Reviews 9:399-431.

Delgado, J. M. R., Roberts, W. W., and Miller, N. E. (1954) Learning motivated by electrical stimulation of the brain, Amer. J. Physiol. 179:587.

Deutsch, M., and Krauss, R. M. (1965) Theories in Social Psychology, Basic Books, Inc., New York, pp. 4-5.

Deutsch, J. A. (1971) The Cholinergic synapse and the site of memory, Science 174:788-794.

Doran, J. E. (1968) Experiments with a pleasure-seeking automaton, Machine Intelligence III.

Eccles, J. C. (1953) The Neurophysiological basis of mind, The Principles of Neurophysiology, Clarendon Press, Oxford.

Eccles, J. C. (1964) The Physiology of Synapses, Academic Press, Inc., New York.

Eccles, J. C. (1966) Conscious experience and memory, Brain and Conscious Experience (ed. by J. C. Eccles), Springer-Verlag, New York, pp. 314-344.

Eccles, J. C. (1969a) The Inhibitory Pathways of the Central Nervous System, Liverpool University Press, Liverpool, p. 135.

Eccles, J. C. (1969b) Excitatory and inhibitory mechanisms in brain, Basic Mechanisms of the Epilepsies (ed. by Herbert H. Jasper, Arthur A. Ward, Jr. and Alfred Pope), Little, Brown and Company, Boston, pp. 229-252.

Eccles, J. C., Ito, M., and Szentagothai, J. (1967) The Cerebellum as a Neuronal Machine, Springer-Verlag, New York, p. 335.

Ernst, H. A. (1970) Computer-Controlled Robots, IBM Research Report RC 2781 (No. 13043).

Farley, B. G., and Clark, W. A. (1954) Simulation of self organizing system by a digital computer, IRE Trans. on Information Theory, Vol. PGIT-4, pp. 76-84.

Feigenbaum, E. A., and Feldman, J. (1963) Computers and Thought, McGraw-Hill, New York.

Feigl, H. (1958) The 'mental' and the 'physical', Minnesota Studies in the Philosophy of Science, Vol. II (ed. by Herbert Feigl, Michael Scriven, and Grover Maxwell), University of Minnesota Press, Minneapolis, pp. 370-497.

Feigl, H. (1960) Mind-body, not a pseudoproblem, Dimensions of Mind, (ed. by Sidney Hook), Collier Books, New York, pp. 33-44.

Finley, M. (1967) An Experimental Study of the Formation and Development of Hebbian Cell-assemblies by Means of a Neural Network Simulation, Technical Report, Department of Communication Sciences, the University of Michigan, Ann Arbor, Michigan.

Fuxe, K. (1965) Acta Physiol. Scand. 64 (Suppl. 247):39.

Gamba, A., Bamberini, L., Palmieri, G., and Sanna, R. (1961) Further experiments with PAPA, Nuovo Cimento Suppl. Vol. 20 (No. 2): 221-231.

Gazzaniga, M. S. (1970) The Bisected Brain, Appleton-Century-Crofts, New York.

References

Gerbrandt, L.K., Skrebitsky, V.G., Buresova, O. and Bures, J. (1968) Plastic changes of unit activity induced by tactile stimuli followed by electrical stimulation of single hippocampal and reticular neurons, Neuropsychologia 6:3-10.

Gerbrandt, L., Bures, J., and Buresova, O. (1970) Investigations of Plasticity in Single Units (In the Press).

Goddard, G.V. (1967) Development of epileptic seizures through brain stimulation at low intensity, Nature (London) 214:1020.

Good, I.J. (1965) Speculations concerning the first ultraintelligent machine, Advances in Computers 6:31-88.

Griffin, J.P. (1970) Neurophysiological studies into habituation, Short-term Changes in Neural Activity and Behavior (ed. by G. Horn and R.A. Hinde), Cambridge University Press, New York, pp. 141-179.

Guerrero-Figueroa, R., Barros, A., Heath, R.G., and Gonzalez, G. (1964) Experimental subcortical epileptiform focus, Epilepsia (Amst.) 5:112.

Gurowitz, E.M. (1969) The Molecular Basis of Memory, Prentice-Hall, Englewood Cliffs, New Jersey.

Heath, R.B., and Mickel, W.A. (1960) Evaluation of seven years' experience with depth electrode studies in human patients, Electrical Studies on the Unanesthetized Brain (ed. by E.R. Ramey and D.S. O'Doherty), Harper, New York.

Hebb, D.O. (1949) Organization of Behavior, Wiley, New York.

Hilgard, E.R. (1969) Pain as a puzzle for psychology and physiology, Am. Psychol. 24:103-113.

Hillarp, N.A., Fuxe, K., and Dahlstrom, A. (1966) Pharmacol. Rev. 18:727.

Hoebel, B.G. (1971) Feeding: neural control of intake, Annu. Rev. Physiol. 33:533-568.

Holmgren, B., and Frenk, S. (1961) Inhibitory phenomena and 'habituation' at the neuronal level, Nature (London) 192:1294-1295.

Horn, G., and Hinde, R.A. (1970) Short-term Changes in Neural Activity and Behavior, Cambridge University Press, New York.

Hubel, D.H., and Wiesel, T.N. (1965) Bincoular interaction in striate cortex of kittens reared with artificial squint, J. Neurophysiol. 28:1040-1072.

Hull, C.L. (1943) Principles of Behavior, Appleton-Century-Crofts, New York.

Jennings, H.S. (1906) Behavior of the Lower Organisms, Columbia University Press, New York, p. 335.

Jung, R. (1953) Allgemeine neurophysiologie, Handbuch der Inneren Medizin, Sprinter, Berlin, pp. 1-181.

Kaylor, D.J. (1964) A Mathematical Model of a Two-layer Network of Threshold Elements, Rome Air Development Center Technical Documentary Report, RADC-TDR-63-534.

Kilmer, W.L., McCulloch, W.S., and Blum, H. (1967) Some Mechanisms for a Theory of the Reticular Formation, Final Scientific Report: AF-AFOSR-1023-66.

Konorski, J. (1948) Conditioned Reflexes and Neuron Organization, Cambridge University Press, Cambridge.

References

Konorski, J. (1950) Mechanisms of learning, Symp. Soc. Exp. Biol. 4:409-431.

Kling, J. W., and Stevenson, J. G. (1970) Habituation and extinction, Short-term Changes in Neural Activity and Behavior (ed. by G. Horn, R. A. Hinde), Cambridge University Press, New York, pp. 41-61.

Lorenz, K. (1970) On killing members of one's own species, Bulletin of the Atomic Scientists XXVI, No. 8.

McCulloch, W. S., and Pitts, W. (1943) A logical calculus of the ideas immanent in nervous activity, Bull. Math. Biophys., 5:115-137 [reprinted in McCulloch, W. S. (1965) Embodiments of Mind, M. I. T. Press, Cambridge, Mass., pp. 19-39].

McIntyre, A. K. (1953) Synaptic function and learning, Abst. 19th Int. Physiol. Congr., pp. 107-114.

Meldrum, B. S. (1966) Electrical signals in the brain and the cellular mechanism of learning, Aspects of Learning and Memory (ed. by Derek Richter), Basic Books, Inc., New York, pp. 100-120.

Miller, N. E. (1957) Science 126:1271.

Milner, P. M. (1957) The cell assembly: mark II, Psychological Review 64(4): 242-252.

Minsky, M. (1954) Neural Nets and the Brain-model Problem, doctoral dissertation, Princeton University, Princeton, New Jersey.

Minsky, M. (1969) Semantic Information Processing, M. I. T. Press, Cambridge, Mass.

Minsky, M., and Papert, S. (1969) Perceptrons: An Introduction to Computational Geometry, M. I. T. Press, Cambridge, Mass.

Morrell, F. (1961) Effect of anodal polarization on the firing pattern of single cortical cells, Ann. New York Acad. Sci. 92:860-876.

Morrell, F., and Baker, L. (1961) Effects of drugs on secondary epileptogenic lesions, Neurology (Minneap.) 11:651.

Morrell, F., Proctor, F., and Prince, D. A. (1965) Epileptogenic properties of subcortical freezing, Neurology (Minneap.) 15:744.

Morrell, F. (1969) Physiology and histochemistry of the mirror focus, Basic Mechanisms of the Epilepsies (ed. by Herbert H. Jasper, Arthur A. Ward, Jr., and Alfred Pope), Little, Brown and Company, Boston, pp. 357-374.

Nauta, W. J. H. (1960) Some neural pathways related to the limbic system, Electrical Studies on the Unanesthetized Brain (ed. by E. R. Rainey and D. S. O'Doherty), Hoeber, New York, pp. 1-16.

Newell, A., Shaw, J. C., and Simon, H. (1957) Empirical explorations with the logic theory machine, Proceedings of the Western Joint Computer Conference 15:218-239 [reprinted in Feigenbaum, Edward A. and Feldman, Julian (1963), Computers and Thought, McGraw-Hill, New York, pp. 109-133].

Newell, A., Shaw, J. C., and Simon, H. (1958) Chess playing programs and the problem of complexity, IBM Journal of Research and Development 2:320-335 [reprinted in Feigenbaum, Edward A. and Feldman, Julian (1963), Computers and Thought, McGraw-Hill, New York, pp. 39-70].

Nilsson, Nils J. (1965) Learning Machines, McGraw-Hill, New York.

Nilsson, Nils J. (1971) Problem-solving Methods in Artificial Intelligence, McGraw-Hill, New York.

References

Olds, J., and Milner, P. (1954) Positive reinforcement produced by electrical stimulation of septal area and other regions of rat brain, J. Comb. Physiol. Psychol. 47:419-427.

Olds, J. (1962) Hypothalamic substrates of reward, Physiol. Rev. 42:554.

Palmieri, G., and Sanna, R. (1960) Methods 12 (No. 48).

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

Penfield, W. (1969) Epilepsy, neurophysiology, and some brain mechanisms related to consciousness, Basic Mechanisms of the Epilepsies (ed. by Herbert H. Jasper, Arthur A. Ward, Jr., and Alfred Pope), Little, Brown and Company, Boston, pp. 791-805.

Pepper, S. C. (1960) A Neural-identity theory of mind, Dimensions of Mind (ed. by Sidney Hook), Collier Books, New York, pp. 45-61.

Poschel, B. P. H., and Ninteman, F. W. (1963) Life Sci. 3:782.

Pratt, J. B. (1937) Personal Realism, MacMillian, New York.

Proctor, F., Prince, D., and Morrell, F. (1966) Primary and secondary spike foci following depth lesions, Arch. Neurol (Chicago) 15:151.

Ramon y Cajal, S. (1911) Histologie du Systeme Nerveux de l'Homme et des Vertebres, Vol. 2, Maloine, Paris.

Rashevsky, N. (1938) Mathematical Biophysics, University of Chicago Press, Chicago, Illinois.

Rochester, N., Holland, J., Haibt, L. H., and Duda, W. L. (1956) Tests on a cell assembly theory of the action of the brain, IRE Transactions on Information Theory, IT-2:80-93.

Rosenblatt, F. (1957) The Perceptron: A Perceiving and Recognizing Automaton, Project PARA, Cornell Aeronautical Laboratory Report 85-460-1.

Rosenblatt, F. (1960) On the Convergence of Reinforcement Procedures in Simple Perceptrons, Cornell Aeronautical Laboratory Report VG-1196-G-4, Buffalo, New York.

Rosenblatt, F. (1962) Principles of Neurodynamics, Spartan Books, New York.

Rowland, V., and Goldstone, M. (1963) Electroenceph. clin. Neurophysiol. 15:474.

Rusinov, V. S. (1953) An Electrophysiological analysis of the connecting function in the cerebral cortex in the presence of a dominant area, Abstr. XIX Int. Physiol. Congr., pp. 719-720.

Russell, B. (1921) The Analysis of Mind, MacMillian, New York.

Russell, I. S. (1966) Animal learning and memory, Aspects of Learning and Memory (ed. by Derek Richter), Basic Books, New York, pp. 121-171.

Sampson, J. R. (1969) A Neural Subassembly Model of Human Learning and Memory, Technical Report, Computer and Communication Sciences Department, The University of Michigan, Ann Arbor, Michigan.

Samuel, A. L. (1959) Some studies in machine learning using the game of checkers, IBM Journal of Research and Development 3:211-229 [reprinted in Feigenbaum, Edward A. and Feldman, Julian (1963), Computers and Thought, McGraw-Hill, New York, pp. 71-105].

References

Segundo, J. P., Takenaka, T., and Encabo, H. (1967) Electrophysiology of Bulbar Reticular Neurons, J. Neurophysiol 30:1194-220.

Selfridge, O. G. (1959) Pandemonium: a paradigm for learning, Proc. Symp. on Mechanisation of Thought Processes, Natl. Phys. Lab., Teddington, England. Her Majesty's Stationery Office, London, 2 vols., pp. 511-529.

Sem-Jacobsen, C. W., and Torkildsen, A. (1960) in Electrical Studies on the Unanesthetized Brain (ed. by E. R. Ramey and D. S. O'Doherty), Harper, New York.

Sharpless, S. K. (1964) Reorganization of function in the nervous system – use and disuse, Ann. Rev. Physiol. 26:357-388.

Sherrington, C. S. (1941) Man on his Nature, MacMillian, New York.

Skinner, B. F. (1938) The Behavior of Organisms: An Experimental Analysis, Appleton-Century, New York.

Skinner, B. F. (1971) Beyond Freedom and Dignity, Knopf, New York.

Spencer, W. A., Thompson, R. F., and Neilson, D. R., Jr. (1966a) Alterations in responsiveness of ascending and reflex pathways activated by iterated cutaneous afferent volleys, J. Neurophysiol. 29:240-252.

Spencer, W. A. M., Thompson, R. F., and Neilson, D. R., Jr. (1966b) Decrement of ventral root electrotonus and intracellulary recorded PSP's produced by iterated cutaneous afferent volleys, J. Neurophysiol 29:253-273.

Sperry, R. W. (1966) Brain bisection and consciousness, Brain and Conscious Experience (ed. by J. C. Eccles), Springer-Verlag, New York, pp. 298-308.

Stein, L., and Seifter, J. (1961) Science 134:286.

Stein, L. (1964) Fed. Proc. 23:836.

Stein, L. (1964) Reciprocal action of reward and punishment mechanisms, The Role of Pleasure in Behavior (ed. by Robert G. Heath), Harper and Rowe, New York, pp. 113-139.

Stein, L. (1967) in Antidepressant Drugs (ed. by S. Garattini and N. M. G. Dukes), Excerpta Medica Foundation, Amsterdam.

Stein, L. (1968) Psychopharmacology: A Review of Progress: 1957-1967 (ed. by D. H. Efron), U. S. Government Printing Office, Washington, D. C.

Stein, L., and Seifter, J. (1970) in Amphetamines and Related Compounds (ed. by E. Costa and S. Garanttini), Raven, New York.

Stein, L, and Wise, C. David (1971) Science 171:1032.

Tanzi, E. (1893) I fatti e la induzione nell' odierne istologia del sistema nervoso, Riv. sper. Freniat. 19:149.

Teitelbaum, P., and Epstein, A. N. (1962) Psychol. Rev. 69:74.

Thompson, R. F., and Spencer, W. A. (1966) Habituation: a model phenomenon for the study of neuronal substrates of behavior, Psychol. Rev. 73:16-43.

Thorpe, W. H. (1956) Learning and Instinct in Animals, Methuen, London.

Tonnies, J. F. (1949) Die erregungssteuerung im zentralnervensystem, Arch. Psychiat. Nervenkr. 182:478-535.

Wada, J. A., and Cornelius, L. R. (1960) Functional alteration of deep structures in cats with chronic focal cortical irritative lesions, Arch. Neurol (Chicago) 3:425.

References

Walter, W. G., Cooper, R., Aldridge, V. J., McCallum, W. C., and Winter, A. L. (1964) Nature (London) 203:380.

Ward, Jr., A. A., Jasper, H. H., and Pope, A. (1969) Clinicial and experimental challenges of the epilepsies, Basic Mechanisms of the Epilepsies (ed. by Herbert H. Jasper, Arthur A. Ward, Jr. and Alfred Pope), Little, Brown and Company, Boston, pp. 1-12.

Weiner, N. (1948) Cybernetics, Wiley, New York.

Widrow, B., and Hoff, M. E. (1960) Adaptive Switching Circuits, Stanford Electronics Laboratories Technical Report 1553-1, Stanford University, Stanford, California.

Widrow, B. (1962) Generalization and information storage in networks of adaline neurons, Self-Organizing Systems - 1962 (ed. by Marshall C. Yovits, George T. Jacobi and Gordon D. Goldstein), Spartan Books, Washington, D. C., pp. 435-461.

Widrow, B., Groner, G. F., Hu, M. J. C., Smith, F. W., Specht, D. F., and Talbert, L. R. (1963) Practical applications for adaptive data-processing systems, 1963 WESCON Paper 11.4 VII, Part 7.

Wise, C. D., and Stein, L. (1969) Science 163:299.

Young, J. Z. (1951) Growth and plasticity in the nervous system, Proc. Roy. Soc. B 139:18-37.

Young, J. Z. (1966) The Memory System of the Brain, University of California Press, Berkeley,