[5] Bialasiewich, J., Multi-measurement model of the process of identification of objects with a study based on the approximation and stochastic method. *Archiwum Automatyki i Telemechanki* **10** (1965).
[6] Robbins, H., and Monro, S., A stochastic approximation method. *Ann. Math. Statist.* **22**, 400–407 (1951).
[7] Kiefer, E., and Wolfowitz, T., Stochastic estimation of the maximum of a regression function. *Ann. Math. Statist.* **23**, 462–466 (1952).
[8] Shmetterer, L., Stochastic approximations. *Proc. 4th Berkeley Symp. Math. Statist. I, 1961.*
[9] Loginov, N. V., Methods of stochastic approximation (survey). *Automation and Remote Control* **4**, 706–728 (1966).
[10] Gavurin, M. K., Nonlinear functional equations and continuous analogues of iterative methods, reports of the higher educational institutions. *Izv. Vysshikh Uchebn. Zavedenii Matematika* **5** (6), 18–31 (1958).
[11] Driml, M., and Hans, O., Continuous stochastic approximation. *Trans. 2nd Prague Conf. Information Theory Stat. Decision Functions. Random Processes, Prague, 1959.*
[12] Arrow, K. G., Hurwitz, L., and Uzawa, H., "Studies in Linear and Nonlinear Programming." Stanford Univ. Press, Stanford, California, 1958.
[13] Sebestyen, G., "Decision-making Processes in Pattern Recognition." Macmillan, New York, 1962.
[14] Aizerman, M. A., Automatic control learning systems (in the light of experiments on teaching the systems to pattern recognition), "Automation and Remote Control" (*Proc. 2nd Congr. IFAC, Basle, 1963*). Butterworths, London and Washington, D.C., 1966; Oldenbourg-Munchen, 1964.
[15] Rosenblatt, F., "Principles of Neurodynamics." Spartan Books, Washington, D.C., 1962.
[16] Yakubovich, V. Ya., "Certain General Theoretical Principles of Constructing Learned Cognition Systems, Computer Engineering and Problems in Programming," **4**. Leningrad Publishing House, Leningrad, 1965.
[17] Motzkin, T. S., and Schoenberg, I. J., The relaxation method for linear inequalities. *Can. Jo. Math.* **6**, 393–404 (1954).
[18] Vapnik, V. N., and Chervonenkis, A. Ya., A class of perceptrons. *Automation and Remote Control* **25**, 106–109 (1964).
[19] Flake, R. H., Volterra series representation on time varying non-linear systems, "Automation and Remote Control" (*Proc. 2nd Congr. IFAC, Basle, 1963*). Butterworths, London and Washington, D.C., 1964; Oldenbourg-Muchen, 1964.
[20] Kokotovich, P., and Rutman, R., S., Sensitivity of control systems (survey). *Automation and Remote Control* **25**, 1512–1518 (1964).
[21] Kulikowski, R., "Optimum and Adaptive Processes in Automatic Regulation Systems." Warsaw-Wroclaw, Poland, 1965.
[22] Koopman, B. O., The theory of search. *Operations Res.* **5**, 613–626 (1957).

Recent Work on Theoretical Models of Biological Memory*

Frank Rosenblatt
CORNELL UNIVERSITY
ITHACA, NEW YORK

At the First COINS Symposium, a model for long-term sequential memory in the nervous system was described. Since that time, this model has been improved by the introduction of a simpler and more biologically plausible *C*-system (a network which serves as a sequential clock for maintaining the temporal order of stored events), and simulation studies of the entire system have been completed. At the same time, an improved biochemical model for the postulated synaptic changes has been developed, and will be evaluated in relation to recent experimental evidence on the biochemical basis of memory. In addition to providing an explanatory model, the new theory suggests a number of biological experiments which are currently being carried out.

I was at the Mathematical School, where the Master taught his Pupils after a Method scarce imaginable to us in *Europe*. The Proposition and Demonstration were fairly written on a thin Wafer, with Ink composed of a Cephalick Tincture. This the Student was to swallow upon a fasting Stomach, and for three Days following eat nothing but Bread and Water. As the Wafer digested, the Tincture mounted to his Brain, bearing the Proposition along with it. But the Success hath not hitherto been answerable, partly by some Error in the *Quantum* or Composition, and partly by the Perverseness of Lads; to whom this Bolus is so nauseous, that they generally steal aside, and discharge it upwards before it can operate; neither have they been yet persuaded to use so long in Abstinence as the Prescription requires.

Jonathan Swift, *Gulliver's Travels*

I. Introduction

Some three years ago, I was privileged to present a theory on the storage of memory in neural networks at the First Symposium on Computer and

* This work was supported by the Office of Naval Research contract Nonr 401(40), and the National Science Foundation contract GK-250.

Information Sciences [13]. The Second Symposium seems an apt occasion to bring this theory up to date and to attempt to integrate it with the results of a most surprising series of experiments which have been performed during the last year, both at our laboratory and elsewhere. These experiments have confronted us with an increasing accumulation of evidence that "memory," or at least a number of varieties of learned behavior, can be transfered from one brain to another by means of chemical extracts. While such a phenomenon has been hypothesized for some time in flatworms [3,10], it was never completely established to the satisfaction of the scientific community, and it is only with the more recent experiments of the last year that such a phenomenon has been demonstrated in vertebrates. These new experiments, which have employed rats, hamsters, and mice as subjects in various laboratories [1,2,6,17], have finally convinced this theorist, at least, that the phenomenon of "memory transfer" is a real one, which must be taken into account in any theoretical approach to biological memory. It is the main objective of this paper to show how the previously developed mathematical theory of memory, in terms of perceptron-type networks, can be reconciled with the data on chemical transfer.

In proposing a chemical model for memory at this time, it must be recognized that we are, in fact, entering the realm of science fiction; the present experiments, although suggestive, leave us completely in doubt as to the mechanism at work in the transfer phenomenon. While the early reports on the transfer of learned behavior suggested that RNA was the molecule responsible for the effect [1,2,6], subsequent experiments at our laboratory and elsewhere [14–17] have thrown considerable doubt on this contention, and have suggested that some form of polypeptide is a more likely candidate. This means that an explanation in terms of genetic coding mechanisms, which several theorists have attempted (c.f., Hydén [9]), is less plausible than it might seem to be at first glance. In what follows, we present an entirely different model which, although speculative, has the virtue of being complete in the sense that it shows on the one hand how the molecular mechanism might operate to modify selected synapses between simultaneously active neurons, and on the other hand how such modifications could lead to the storage and recall of experiential information in a form which would permit the organism to respond to it selectively. The theory is sufficiently rigorous and quantitative that numerical predictions as to the information capacity of a network can be made, and it may serve as a guide in the design of experiments to test the biochemical hypotheses.

We begin this exposition by reviewing the main features of the mathematical theory presented in the previous paper [13], although without repeating the detailed analysis which has been adequately handled there. The following section then discusses some of the work which has been done during the last

three years on improving the model of the *C*-network which provides the sequential control for the storage and recall of series of events. Section III also includes some illustrations of digital simulation studies which have been carried out on the entire system. We then proceed to an examination of the conditions which must be satisfied at a microscopic or molecular level if this model is to be valid, and propose a molecular model which is compatible both with the memory transfer experiments and with the mathematical properties required for the *C*-system to operate.

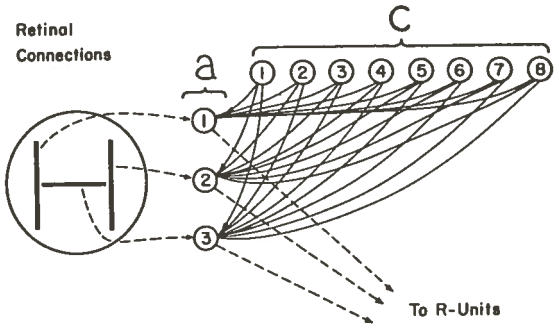
II. A Review of the Mathematical Model

In the previous paper [13], two somewhat different mathematical models were developed in parallel, which were designated the "asymmetric" and the "symmetric" model. It was shown that the asymmetric model, in which only connections terminating on active neurons (or *A*-units) are modified, is more efficient under plausible conditions of biological activity (in which, at most, a few percent of the neurons would be active at any time) than the symmetric model, in which connections to inactive units are also subject to modification. In the following presentation, therefore, we limit ourselves to the asymmetric case.

Before presenting the equations which govern this system, a concrete illustration of a miniature *C*-system network may be helpful in seeing how it works. Such a system is shown in Fig. 1. It contains three *A*-units, of which the first is activated by a vertical line on the left side of the visual field, the second by a vertical line on the right side of the field, and the third by a horizontal line across the middle. Thus a figure shaped like the letter "H" would activate all three units. There are eight *C*-units, and, because this is a very small system in which realism has been sacrificed to some extent to provide the distribution properties which would normally be found in much larger networks, we assume that the system is fully connected, each *C*-unit sending a connection to each of the three *A*-units. We assume a sequence of four stimuli, $S_1, \ldots, S_4$, which are illustrated in the figure. Each of these stimuli activates a unique pattern in the association system, and could be readily discriminated by an *R*-unit which might form part of this system in a perceptron [12]. We assume that the initial weights of all *C*-to-*A* connections are 4, and that in each state of the *C*-system, 50% of its units are active. Assume the threshold of the *A*-units to be 18. The weights of the *C*–*A* connections are modified according to a γ-system rule [12] which requires that the sum of all weights terminating on any one unit remain constant. In this case, the rule states that whenever a *C*-unit and an *A*-unit to which it is connected are both active simultaneously, the connection between them will gain one unit

of weight, the remaining connections to the same A-unit losing a corresponding decrement which just balances the gain in the active connection.

We assume that the C-system is initially in state C_1, concurrently with the presentation of stimulus S_1; this is followed by state C_2 with stimulus S_2, C_3 with S_3, and C_4 with S_4. The active C-units in each state are shown in



(a)

Stimulus:	S_1	S_2	S_3	S_4
Retinal pattern:	‖	⊢	⊣	H
a_1	1	1	0	1
a_2	1	0	1	1
a_3	0	1	1	1

(b)

C-states:	C_1	C_2	C_3	C_4
c_1	1	1	1	1
c_2	1	1	0	0
c_3	1	0	1	0
c_4	1	0	0	1
c_5	0	1	0	0
c_6	0	1	1	1
c_7	0	0	0	1
c_8	0	0	1	0

(c)

FIG. 1. Illustration of a small C-system model. (a) A–C network, showing stimulus features activating each A-unit. (b) Four stimuli, and A-unit activity vectors for each. (c) Succession of four C-states assumed to occur during stimulus sequence.

Fig. 1(c). We are not concerned at the moment with the mechanism responsible for setting the C-system to different states, but it is important to note that the four states shown in the figure are pairwise independent, as they would be in a much larger system in which a certain percentage of units was activated at random in each successive state. In this case, this means that every pair of C-states has two active units in common, and two inactive units in common.

TABLE I

MATRICES OF WEIGHTS AND SIGNALS FROM C-STATES TO EACH A-UNIT, OBTAINED AFTER EACH SUCCESSIVE STIMULUS HAS OCCURRED[a]

		Weights of connections								Total signals from C-states to A-units			
		c_1	c_2	c_3	c_4	c_5	c_6	c_7	c_8	C_1	C_2	C_3	C_4
Initial state:	a_1	4	4	4	4	4	4	4	4	16	16	16	16
	a_2	4	4	4	4	4	4	4	4	16	16	16	16
	a_3	4	4	4	4	4	4	4	4	16	16	16	16
		c_1*	c_2*	c_3*	c_4*	c_5	c_6	c_7	c_8	C_1	C_2	C_3	C_4
After S_1:	a_1*	5	5	5	5	3	3	3	3	20	16	16	16
	a_2*	5	5	5	5	3	3	3	3	20	16	16	16
	a_3	4	4	4	4	4	4	4	4	16	16	16	16
		c_1*	c_2*	c_3	c_4	c_5*	c_6*	c_7	c_8	C_1	C_2	C_3	C_4
After S_2:	a_1*	6	6	4	4	4	4	2	2	20	20	16	16
	a_2	5	5	5	5	3	3	3	3	20	16	16	16
	a_3*	5	5	3	3	5	5	3	3	16	20	16	16
		c_1*	c_2	c_3*	c_4	c_5	c_6*	c_7	c_8*	C_1	C_2	C_3	C_4
After S_3:	a_1	6	6	4	4	4	4	2	2	20	20	16	16
	a_2*	6	4	6	4	2	4	2	4	20	16	20	16
	a_3*	6	4	4	2	4	6	2	4	16	20	20	16
		c_1*	c_2	c_3	c_4*	c_5	c_6*	c_7*	c_8	C_1	C_2	C_3	C_4
After S_4:	a_1*	7	5	3	5	3	5	3	1	20	20	16	20
	a_2*	7	3	5	5	1	5	3	3	20	16	20	20
	a_3*	7	3	3	3	3	7	3	3	16	20	20	20
		c_1*	c_2*	c_3*	c_4*	c_5	c_6	c_7	c_8	C_1	C_2	C_3	C_4
After repetition of S_1 with state C_1	a_1*	8	6	4	6	2	4	2	0	24	20	16	20
	a_2*	8	4	6	6	0	4	2	2	24	16	20	20
	a_3	7	3	3	3	3	7	3	3	16	20	20	20

[a] Asterisks indicate active units.

Now suppose the four stimuli occur in order, while the C-system is made to run through the four states shown in Fig. 1, and the connections are modified according to the rule given. The successive values of the 24 C–A connections are given by the succession of matrices shown in Table I. Now suppose, without further modification of the connections, and without any stimuli present on the retina, the C-system is made to run through the same four states a second time. The signals to each A-unit from each of the four C-states will then correspond to the values given in the four right-hand columns of Table I. Note that the responding A-units in each case (those for which the signal is greater than the threshold of 18) are identical to those originally activated by the stimulus which accompanied the C-state when it first occurred. This is true not only for the final condition of the C-system, but at any time during the recording process. Note also that the addition of a new stimulus, or a repetition of one which occurred earlier, does not in any way change the signals from other C-states to the A-units: after stimulus S_1 has occurred, for example, a repetition of state C_1 would transmit a signal of 20 to the appropriate A-units, but states C_2, C_3, and C_4 all continue to transmit total signals of 16 to each A-unit, just as they did at the outset. Finally, we note that it is not at all necessary to discontinue the memory operation during recall, as we have supposed here; if it continues to operate, it will merely reinforce the bias previously introduced, without modifying any other stored information.

The equations previously found to characterize networks of this type [13] are repeated here in summary form:

It is assumed that the basic perceptron network (the connections from the A-units to some R-unit) has been previously taught to identify stimulus S_x which occurs at some point during the sequence of stimuli presented for recording by the C-system. After the sequence of stimuli has been presented, the succession of C-states is assumed to be regenerated (by one of several mechanisms to be considered in the following section). The probability $P(R_x)$ that the correct response of the R-unit to stimulus S_x will now be evoked by the C-state which originally accompanied S_x is to be computed for a system with a large number of A-units. The following terms are defined:

N_a = number of A-units
N_1 = number of A-units activated by S_x ("proper units")
N_0 = number of A-units inactive for S_x ("improper units")
N_c = number of C-units
M = fraction of C-units connected to each A-unit ($0 < M \leq 1$)
Q_a = proportion of A-units activated by a stimulus
$\hat{Q}_1$ = the fraction of stimuli to which one of the "proper units" responds
$\hat{Q}_0$ = the fraction of stimuli to which an "improper unit" responds
Q_c = proportion of C-units active in any one state
$E(u_x)$ = expected value of signal to R-unit when S_x occurs
$\sigma(u_x)$ = standard deviation of the signal (u_x) to the R-unit

$\Phi(z)$ = cumulative normal distribution function, integrated from $-\infty$ to z
$\Phi'(z)$ = normal density function
t = number of stimuli in the recorded sequence.

The values of $\hat{Q}$ for different A-units will all approach Q_a in a sufficiently heterogeneous and nonrepetitive environment. In a stereotyped environment, however, in which a small number of stimuli keep recurring, or in which particular stimulus features reoccur many times, $\hat{Q}_1$ is likely to be much greater than Q_0. (The following expressions are simplified, in that they assume only these two values of $\hat{Q}$; in a more general treatment, there should be a separate term for each possible value of $\hat{Q}$ corresponding to different subsets of A-units.) The expectation and variance of signals to the R-unit [$E(u_x)$ and $\sigma^2(u_x)$] depend on how the perceptron was trained in the discrimination of stimulus S_x, and can be computed for a number of representative cases (see Rosenblatt, [13]). The probability of obtaining a correct response for the recall of the test stimulus S_x is then given by the expression

$$P(R_x) = \sum_{n_1=0}^{N_1} \sum_{n_0=0}^{N_0} \binom{N_1}{n_1}\binom{N_0}{n_0} P(n_1, n_0)\Phi[Z(n_1, n_0)], \tag{1}$$

where

$$Z(n_1, n_0) = \frac{[n_1/Q_aN_a] - [n_0/(1 - Q_a)N_a]}{[(n_1 + n_0)/Q_aN_a]^{1/2}} \frac{E(u_x)}{\sigma(u_x)} \tag{2}$$

$$P(n_1, n_0) = \int_{-\infty}^{\infty} F_1^{n_1}(x)[1 - F_1(x)]^{N_1 - n_1} F_0^{n_0}(x)[1 - F_0(x)]^{N_0 - n_0}\, dx \tag{3}$$

$$F_i(x) = \Phi\left(\frac{xM^{1/2} + h_i}{4\hat{Q}_i t}\right) \tag{4}$$

$$h_i = \left[\frac{MN_c(1 - Q_c)}{4\hat{Q}_i t}\right]^{1/2} \tag{5}$$

The function Z represents the expected ratio of the expected value to the variance of the signal to the R-unit when n_1 "proper units" and n_0 "improper units" are reactivated by signals from the C-system. $P(n_1, n_0)$ is the probability that these numbers of units will, in fact, be activated. It is seen that this is a function of h which is the ratio of the expected signal from the C-system to the standard deviation of that signal, for any one A-unit, at the time that the proper C-state for recall of stimulus S_x occurs. The probability $P(n_1, n_0)$ would simply be equal to $\Phi^{N_1}(h_1) \cdot \Phi^{n_0}(h_0)$ if the signals to the different A-units could be assumed to be statistically independent. Correlation effects, however, come from two sources: one is the fact that the sets of C-units connected to different A-units are likely to have a fraction M of their units in common; the other is the effect of frequent joint activation (and correlated

reinforcement) of those A-units which respond to recurrent stimuli in a repetitive environment. These two effects lead to the appearance of the quantities M and $\hat{Q}_i$ in expression (4), and the relatively complicated expression (3) for the probability. The analysis which leads to this result is presented in detail in Rosenblatt [13]. It is assumed in this analysis that the threshold of the A-units is maintained at a level which guarantees a uniform Q_a, regardless of the source and magnitude of signals to the A-units.

An approximation, which holds accurately for large values of N_a, is given in Ref. 13, which replaces the summations in Eq. (1) by appropriate integrals. Numerical computations have been done chiefly by means of that simplification. This treatment also permits the computation of an asymptotic limit for $P(R_x)$ as the number of A-units increases. Specifically, we find that if the discrimination of S_x from other stimuli was originally learned perfectly, and with $h_1 = h_0 = h$,

$$\lim_{N_a \to \infty} P(R_x) = \Phi \frac{h}{M^{1/2}} = \Phi\left[\frac{N_c(1 - Q_c)}{4Q_a t}\right]^{1/2} \tag{6}$$

From this it is easy to see that the performance will be improved if Q_a is small and N_c large; also that the probability will asymptotically approach 0.5 as t increases.

A number of representative tables have been computed from these equations [13], which demonstrate that extremely long sequences can be stored and recalled correctly by networks of this type. For example, assuming an ideal limiting case in which $\hat{Q}_i \approx 0$, we find that a network with 1000 A-units and 1000 C-units, with 1000 connections per A-unit, could record a sequence of 10^5 stimuli before the probability of correct recall of a test stimulus falls to 0.994. With 10^9 A-units, 10^9 C-units, and 1000 connections per A-unit, as before, a sequence of 10^{11} stimuli could be stored with the same performance level. If $\hat{Q}_i$ is taken at a more reasonable level for a neurological system, say 0.05 (so that we expect, on the average, about 5% of the stimuli to activate any one A-unit), then we find that with 1000 connections per A-unit, if the number of stimuli recorded is kept equal to the number of A- and C-units, the probability of correct recall will remain about 0.997. This means, in effect, that each additional A-unit and C-unit pair added to the system permits the recording and recall of one additional stimulus without lowering the performance probability. As Q_a (or $\hat{Q}$) increases, the number of A and C-units required per stimulus to maintain a given probability level, also increases. In the previous paper [13], it was shown that as the system saturates, the information stored per connection (in bits) approaches

$$H_\infty = \frac{1 - Q_c}{4Q_a \pi \ln 2}. \tag{7}$$

For Q_c close to zero and Q_a at 0.05, this gives 2.296 bits per connection as the limiting information density. This means that as further stimuli are shown to the system, the total amount of information stored remains constant, the loss per stimulus being reflected in the diminishing probability of correct recall.

It should be noted that these results assume that there is no deterioration or decay of stored information as long as it is left alone; connections which are not modified as a result of neural activity are assumed, in this model, to maintain their weights without decrement. If we were to assume a slow decay function, such as an exponential return to the initial, unmodified conditions, then the equations given would still hold accurately for short sequences, but for longer sequences we would be obliged to introduce a time-dependent weighting function, which would favor the more recent stimuli at the expense of the earlier ones.

It has also been tacitly assumed that the succession of C-system states can be generated and regenerated without error. This is a somewhat implausible assumption for any real system by which such state sequences might be generated, as will be seen in the following section.

The most important mathematical property of the model just described is the fact that the expected signal to any A-unit from a C-state which has not been previously reinforced, remains equal to its initial value, regardless of how many stimuli have been recorded in the network. This effect, which was illustrated in the small network of Fig. 1, is what makes it possible to store such long sequences without mutual interference between previously recorded events. The deterioration in performance which ultimately occurs is not due to the accumulation of any systematic bias, but rather to the accumulation of variance in the signals from the random intersections of sets connected to different A-units. This property also makes the system highly resistant to the addition of random noise or to the extirpation of portions of the C-system, as long as the removal is not systematically correlated with any of the C-states which have been employed. Thus the phenomenon of "distributed memory" is admirably illustrated by a network of this type. This independence effect depends basically on two conditions which must be satisfied in the initial design of the model: (1) The intersections between any two C-states should have an expected value equal to the product of the measures of the active sets; i.e., the states should have the same intersection properties that they would satisfy if they were chosen at random. Also, the connections from A-units to C-units should not be systematically correlated with any particular C-states, thus guaranteeing that the intersection property just stated holds within the set of C-units connected to any one A-unit, as well as for the system as a whole. (2) The weight modifications should follow the γ-system rule, whereby the gain in active connections to any given unit is just balanced by a compensating loss in the weights of inactive connections to the same unit.

It is the combination of the γ-system with the quasi-random assignment of C-units to different states which guarantees the type of behavior which we have observed.

III. Models for the C-Network and Simulation Results

In the original version of this theory (Rosenblatt [13]) it was shown that the required properties of state sequences in the C-system could be obtained by a randomly coupled network with 1 : 1 connections between the units. The system is started by turning on some arbitrarily chosen fraction Q_c of the units, and is then allowed to run freely through successive states. As long as the unit-to-unit transmission remains synchronized, the level of activity will remain exactly what it was initially, and if the initial state was chosen at random, the successive states will satisfy the conditions described in the preceding section. The identical sequence will be regenerated if the system is set back to its first state. Although logically satisfactory as an illustration of the type of system which we are seeking, it was clear from the outset that such a network was totally implausible as a biological model. Its requirements for perfect synchronization, as well as the one-to-one constraint on the connections, make it unacceptable. Several alternative models were proposed, which would advance their state only when signaled to do so by the momentary interruption of their activity by an inhibitory signal. The assumptions which went into these models, however, were speculative in the extreme, and a more satisfactory model was sought for at least two years following the publication of that paper.

The first family of models to be investigated assumed that the C-network might consist of a randomly connected network of neurons, with inhibitory connections between the cells. It was assumed that a steady (excitatory) drive signal to all of the C-units would normally tend to sustain some level of activity in the network, and that the signal to change states would consist of a momentary suppression of this drive signal. The functioning of this network is as follows: Assume that initially some set of units C_1 is activated by the drive signal. Those units which are most strongly activated, and which receive a minimum of inhibition from other active units, will tend to remain active, while the remainder of the network (consisting of the majority of the C-units, if it is strongly coupled) will be inhibited, and will remain inactive. As long as the drive signal continues without interruption, the same dominant set of C-units will remain on, and the state C_1 continues. But the inhibitory signals which are continuously bombarding the remaining neurons in the network will have the effect of "priming" them, so that if the drive signal is momentarily stopped, not only does the previously active set stop firing, but the neurons which were previously most strongly inhibited will now tend to become active, as a result of the well-known post-inhibitory rebound phenomenon in biological neurons. Thus when the drive signal is restored, a new set of neurons are already firing, and have inhibited the remainder of the system. This constitutes state C_2, which will again remain stable until the drive is interrupted a second time when the neurons most strongly inhibited by C_2 will emerge as the elements of the third active set, and so on.

Extensive simulation studies of this type of network were carried out on the IBM 7090 and 7094 computers, using neuron models which incorporated many of the known details of a neuron's response including temporal summation, adaptation to excitatory and inhibitory signals, and realistic bounds on such quantities as the magnitude of the membrane potential, maximum hyperpolarization, excitatory and inhibitory post-synaptic potentials, and other relevant features. The general results, though disappointing, were instructive. It was found that a steady state could indeed be maintained in such a network, provided the inhibitory coupling was strong enough to completely suppress the activity of the "off" neurons, allowing complete dominance of a small subset. In some cases, activity would fluctuate between several subsets of competing neurons, but with appropriate choice of parameters, this could generally be prevented. It was also found that parameters could be found in which the system would advance through a succession of states, as predicted, each state in turn becoming stable until the drive was interrupted. Although the networks simulated were small (rarely over twenty neurons) it appeared plausible that in large systems, very long state sequences could in principle be obtained before returning to a previous state.

The main difficulty with this model was found to be in the sensitivity of the state succession to minor variations in timing, and to noise effects which manifested themselves during the transitional period between one state and the following one. As long as the drive was on, and a particular state remained dominant, it might be held indefinitely, and would be highly noise resistant under proper parametric conditions. But in the choice of the next state, the decision as to which state would emerge was found to depend upon very slight differences between the signals from competing neurons. Suppose, for example, neuron c_1 has mutually inhibitory connections with c_2. Then if c_1 fires slightly in advance of c_2, it will become dominant, together with any neurons which might be supported by its activity, while if c_2 fires first, it becomes dominant, keeping c_1 from ever firing at all. Thus differences of less than a millisecond in the time of firing of a particular neuron could seriously alter the composition of the following active set. Although many seemingly plausible treatments of this effect were tried in attempts to cure it, none were successful, and it now appears that the phenomenon is inherent in the uniform randomness of the design, whereby an active set not only primes those neurons which are supposed to follow it, but a great many unwanted nuerons as well,

many of which have nearly as strong a tendency to become active as do the proper neurons. This close competition between wanted and unwanted neurons makes the system inescapably sensitive to noise effects or minor variations in timing whenever a transition is to occur. Thus, while a state sequence having the required properties can be generated successfully, it is unlikely to repeat itself for more than one or two successive states on a second run from the same starting point.

This finding led us to an examination of increasingly constrained networks, in which each active set would prime a selected following set consisting of noncompetitive neurons, which would therefore have a clear superiority as the next state of the system. While a number of possible models can be constructed in this fashion, with "pre-wired" connections, the most interesting possibility which emerged was that of using the same memory principle as in the C–A connections, to enable each C-state to "learn" to select an appropriate following state. An example of such a network is shown in Fig. 2. The basic organization of the network [Fig. 2(a)] consists of a perceptron with a sensory system (A-units and R-units) plus a C-system which is subdivided into two kinds of units: an excitatory set (C_E) and an inhibitory set (C_I). Connections to the A-units are drawn from the excitatory set, and follow the same adaptive rules as before. The E-units also have adaptive connections to other C-units, both E and I, at random. The inhibitory (I) set (which has fixed output connections terminating on random sets of E-units) does not affect the A-units directly, but helps in determining the sequence of C-states, as shown in Fig. 2(b). A random generator (or any other mechanism for state selection, such as a vector of signals from the currently active R-units) is used to select an initial state C_1 consisting of the active sets E_1 and I_1. The set I_1 strongly inhibits some set of E-units, designated F_1 in the figure. While E_1 and I_1 are both on, their interconnections are strengthened in accordance with the γ-system rule, so that, eventually, turning on only a portion of the E-set is likely to activate the entire E_1 and I_1 state. Note, however, that due to the use of the γ-system, turning on an independently chosen E-set will not have any net effect (other than random noise) on the units of state C_1. The state may now be advanced to C_2. This occurs when the random generator selects a new state (E_2, I_2), by transmitting some new signal vector to the C-system. As soon as the former state C_1 is no longer maintained and I_1 ceases to fire, the "follower set" F_1 is activated by a post-inhibitory rebound phenomenon. For a short period therefore (say about 100 msec), F_1 continues to fire jointly with the new set E_2, and becomes integrated with it (and with I_2 as well) by the augmentation of its excitatory connections. I_2, in turn, now selects a follower set F_2, which will subsequently become integrated with a new randomly chosen set, E_3 and I_3.

The main advantages of this system are, first, that the following state which

is "primed" by any given C-state is highly selective, and the primed units (the members of the F-set) do not compete with one another for dominance, but rather tend to support one another through the development of excitatory interconnections. Thus, having turned on a portion of a previous C set, the remainder of that set tends to be reactivated in its entirety, just as a set of

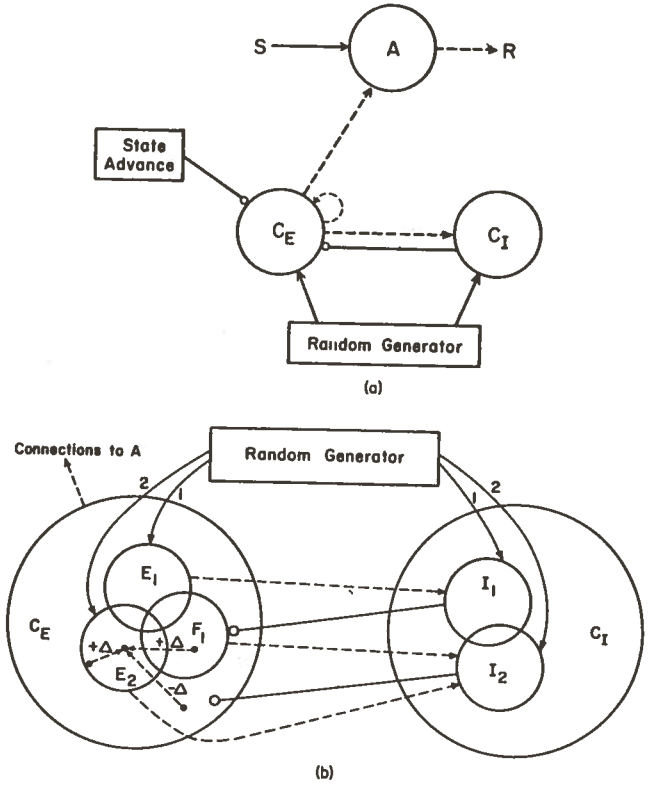


FIG. 2. Adaptively coupled C-system: (a) basic set organization; (b) state-sequence generation.

A-units is turned on by a previously associated C-state. There is, in fact, no logical difference between the use of a C-state to turn on an A-state and the use of the same C-state to turn on an E-unit or I-unit which was previously associated with it. As with the A-system, there should be no net effect on any C-units from a new, randomly chosen C-state, and "improper units" will not be affected. A threshold servo may be used, as in the association system, to maintain a desired level of activity (Q_c) at all times. With proper choice of parameters, the activation of only a few percent of the proper C-units will

define a state sufficiently so that the remainder will "pop in," in the manner of a flip-flop. The second advantage of this system is the possibility of arbitrarily selecting the sequence of C-states by means of the external driving function. While the F-sets are a deterministic consequence of the preceding state, the portions designated E_i are freely selected, and are tied to the preceding states only through the F's. Thus a preceding sequence can be "edited," sections deleted and new sections interpolated, by forcing a new successor state to accompany a previous F-set. This forcing might be done by the use of adaptive connections from R-units to key states in C_E which initiate particular sequences. If this is done, note that an F-state can become coupled to two or more alternative E-sets, but that with a threshold servo to keep Q_c constant, only one of these is likely to become dominant at one time due to the excitatory cross-coupling which develops within each E-set and the weakened connections between them. Thus one of two rival associations (but not a mixture of both) will tend to become dominant at one time.

In "playing back" previously established sequences, the function of the random state generator is replaced by the "state advance" system, which merely suppresses the activity of the existing E-state whenever the C-system is to be advanced to the next state. This permits the primed F-set to become active, and this in turn activates the associated E- and I-units.

While the internal structure which results in such a system may become quite complicated, note that the initial constraints necessary to construct it are strikingly simple, and concern only the statistical parameters of the system. While no quantitative studies of this system have been completed at

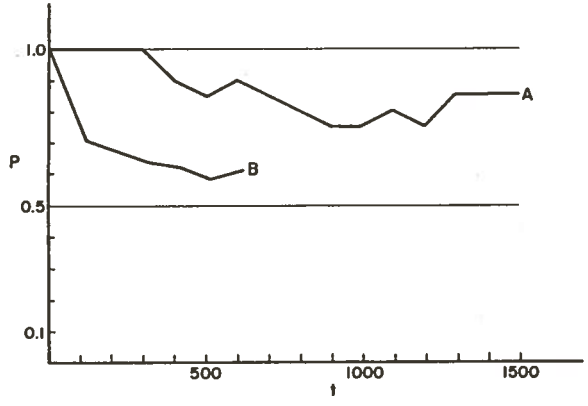


FIG. 3. Simulation results, showing $P(R_x)$ as function of length of stored sequence (t). Each curve is for mean of 10 perceptrons, with 100 A-units and 100 C-units each, fully coupled. $Q_c = 0.14$. Curve A: sequence consists of horizontal bar, followed by vertical bar, followed by t random dot stimuli. Curve B: sequence consists exclusively of alternating horizontal and vertical bar.

this time, it seems likely that a wide latitude of admissible parameters will prove to be workable.

Several simulation studies of complete C-system perceptrons have now been completed (on the 7094 computer), which indicate that our quantitative predictions of performance actually hold quite well for systems of about 100 A-units and C-units, with sequences of simple geometric stimuli. Some examples of curves obtained in these experiments are given in Figs. 3 and 4.

For curve A in Fig. 3, the perceptron was first trained to distinguish a single horizontal bar (4 units wide and 20 long on a 20 × 20 retina) from a single vertical bar. These two bars formed the first two members of a stimulus

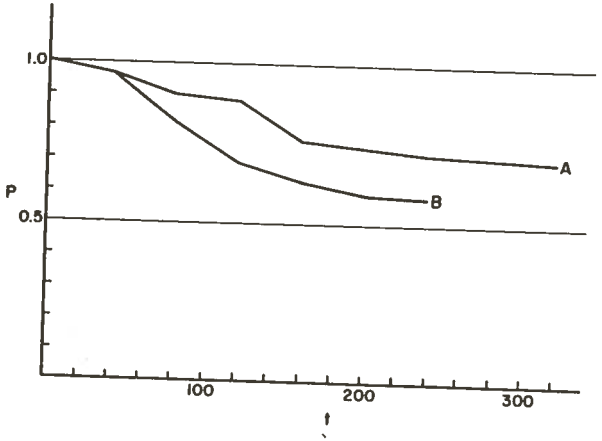


FIG. 4. Simulation results, for same perceptrons as Fig. 3. Curve A: sequence consists of twenty horizontal and twenty vertical bars, followed by t random dot stimuli. Curve B: sequence consists exclusively of repeated cycles of twenty horizontal and twenty vertical bars.

sequence which was recorded in the C-system, the remainder of the sequence being made up of random dot stimuli of 80 retinal points each. The curves show the average performance of ten perceptrons, with 100 A-units and 100 C-units in each, which are fully coupled. The A-units had three excitatory and one inhibitory connection from the retina, with activity held at 9% by means of a threshold servo. The C-system was a free-running binomial cross-coupled network (see Rosenblatt [12]), with three excitatory and one inhibitory connections to each C-unit, originating from randomly chosen points in the set. These parameters give a value for Q_c of about 0.14. (Other curves obtained for Q_c of 0.37 show a poorer performance, as anticipated from the theory.) The curve shows the proportion P of correctly recalled identifications of the two test stimuli as a function of the number of stimuli stored in the memory, when the first two C-states were repeated. A test was run after every 100

additional stimuli were recorded. The second curve B shows the effect of a stereotyped sequence consisting entirely of the horizontal bar alternating with the vertical bar (H, V, H, V, ...), the first twenty stimuli being tested to obtain an estimate of P.

Figure 4 shows analogous curves, with the same parameters, but with a set of twenty different horizontal bars and twenty different vertical bars followed by random stimuli for curve A (the heterogeneous environment case), and continuously repeated cycles of all forty horizontal and vertical bars for curve B (the stereotyped environment case). Note that the difference between the two cases is less than in Fig. 3 where the entire sequence for curve B is made up of only two stimuli, but that the overall performance for the heterogeneous case is poorer, due to the greater difficulty of the discrimination problem in the 40-bar environment. It is noteworthy that for the easy problem of identifying the single horizontal and vertical bar in the sequence of random stimuli [Fig. 3, A], none of the ten perceptrons made a single mistake for recorded sequences of up to 300 stimuli.

These results demonstrate that the general scheme of the C-system model, employing a γ-system memory mechanism, is indeed capable of recording and reproducing long sequences of stimuli successfully. A number of other studies now in progress suggest that such systems may be consulted at randomly chosen states by means of adaptive connections from the R-units, and may participate in much more elaborate cognitive processes than the problems of simple recall which are illustrated here. For present purposes, however, the preceding demonstration seems sufficient to demonstrate the adequacy of the γ-system mechanism to produce many of the chief phenomena of human long-term memory. In the following section, we show that the γ-system is capable of being generated by a plausible (albeit completely hypothetical) biochemical mechanism, and that this mechanism might readily yield the phenomena of "memory transfer" which we have demonstrated in our recent experiments.

IV. A Biochemical Model for the γ-System

In our 1963 paper, we proposed a biochemical model which might produce the logical effect of a γ-system as required by the mathematical theory. This model assumed that pores in the subsynaptic membrane at inhibitory synapses might be blocked as a result of chemical changes occurring during the correlated activity of the presynaptic inhibitory neuron (a C-unit) and the post-synaptic cell (an A-unit) thus leading to an over-all gain in the net excitatory signal to an A-unit. The effect of the blockage upon synaptic functioning depends on the finding that an inhibitory transmitter substance seems to open pores of limited size in the post-synaptic membrane, thus selectively

increasing its permeability to potassium, but not to sodium (c.f. Eccles [5]). In order to obtain the γ-system effect, it was assumed that the blocking molecule, called a "recorder substance," was held in an equilibrium concentration by means of an antagonist, so that only a certain percentage of the total synaptic sites on a cell could be blocked at any one time. This latter mechanism now seems particularly dubious, and the entire theory, which depends on the postulation of four types of molecules which have yet to be discovered, seems somewhat tenuous as an explanation of long-term memory, although the underlying concept of synaptic blockage of inhibitory sites still seems attractive as a possible mechanism for short-term memory, where the γ-system rule need not be strictly followed.

The main impetus to seek a new biochemical model came from the "memory transfer" experiments mentioned in our introduction. The previous model seemed incompatible with this phenomenon, and it was felt that in trying to meet the challenge of finding a plausible theory for the transfer effects, and, at the same time, satisfying the conditions for the γ-system model, we might possibly come closer to depicting the true state of affairs. The model which has resulted is, of course, still in the realm of science fiction, but it has suggested a number of laboratory experiments and has so far correctly predicted several of the properties of the active factor found in our transfer studies.

We have seen that the most essential mathematical property that must be satisfied by the model is that of "conservation of the weights" to any given unit, represented by the γ-system. In addition to this, the transfer phenomenon now imposes several additional conditions:

(1) Since the modifications necessary for storage of information in our model are specific to particular combinations of pre- and post-synaptic neurons, the information-carrying molecules must somehow identify particular pre- and post-synaptic cell combinations, either individually or by sets.

(2) The information carrying molecules must either be capable of reaching and acting selectively upon their corresponding active sites, or must induce the formation of analogously coded molecules which selectively affect these sites. In particular, the molecules or their products must be capable of crossing the blood-brain barrier in sufficient quantity to produce the observed effects.

(3) The observed effects seem to take a period of at least a few hours to appear, following injection, and typically reach their peak only after 12 h (Rosenblatt and Miller [16]). The time course of any process postulated by the model must be commensurate with this.

We assume that some process such as synaptic blocking is responsible for short-term retention of memory, with a completely separate process for long-term memory, which is what we shall describe here. It is assumed that long-term memory involves the preferential gain in stability of connections from

certain presynaptic cells (e.g., *C*-units) to certain post-synaptic cells (e.g., *A*-units or *R*-units). It is assumed that when an event is recorded, there is a period of a few hours to a few days during which some of the structurally weaker connections to the post-synaptic cell are lost, and are replaced in a competitive fashion by new endings emanating from the appropriate set of presynaptic cells. The stability of the "correct" cell junctions, which are thus formed, is assumed to be enhanced by the production of an adhesive molecular complex, specifically coded for both the pre-synaptic and post-synaptic membranes of the appropriate cell pairs, and made up of constituents released by the pre- and post-synaptic cells when they are jointly active.

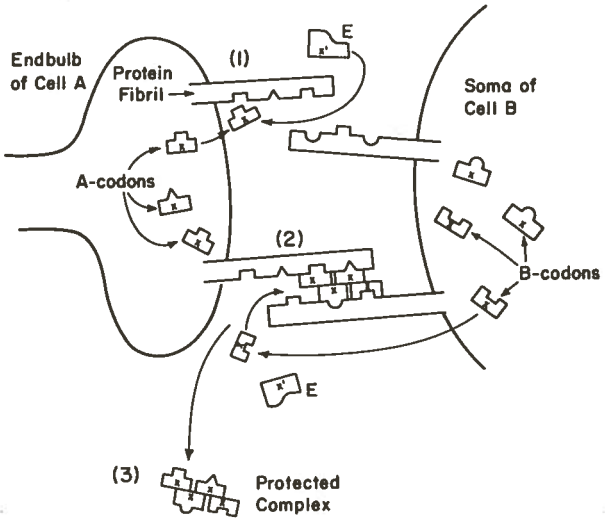


FIG. 5. Molecular memory trace mechanism.

The details of this process are illustrated in Fig. 5. This shows a highly schematic representation of a single synaptic region, including an endbulb coming from cell *A*, the synaptic cleft, and the subsynaptic membrane belonging to cell *B*. It is now known that a large fraction of synapses in the central nervous system contain a network of "intrasynaptic filaments," or fibrils, about 100 Å or less in diameter, and spaced about 200 Å apart; these are readily seen in many electron microscope preparations. Shortly after preparing a preliminary version (about a year ago) of the picture which appears in Fig. 5, I was intrigued to find a report of some electron microscope work by Gray (in Robertson [11]), in which he claims to have observed intrasynaptic fibrils emanating from both the pre- and post-synaptic membrane and forming a hooklike junction within the synaptic cleft. The resemblance of Gray's drawing to our postulated structure is most striking. It should also

be borne in mind that the endbulb and subsynaptic membrane typically form a structural unit strong enough so that when an endbulb is torn away from a cell, as in differential centrifugation procedures (cf. Gray and Whittaker [7]), it tears off a piece of subsynaptic membrane which continues to adhere to it. Since the synaptic cleft itself is typically some 200 Å wide, additional structural elements, such as the intrasynaptic filaments, must be responsible for the structural strength of the synaptic junction. For present purposes, however, the postulation of these filaments is a convenient, rather than a necessary assumption for the theory.

In Fig. 5, these filaments are assumed to be made up of structural protein, with template regions which specifically characterize the cell from which they originate. The specificity of this characterization, which is one of the more startling claims of this theory, requires some examination. There is already ample evidence of an extreme degree of specificity in some of the innate connections of the vertebrate brain. For example, Hubel and Wiesel [8] have reported in the visual cortex of the cat, various types of cells which respond to lines and edges in a particular location and orientation in the visual field, and moving in a given direction within specific velocity limits. These cells typically respond binocularly, the identical conditions which activate the cell for the left eye being those which activate it for the right eye. This already calls for a highly sophisticated mechanism for growing synaptic connections from the two eyes (by way of the lateral geniculate and other intervening stages) so that just the right types of cells from both retinae are ultimately connected to the same neuron, by way of the same types of intervening analyzing mechanisms. It should be noted, moreover, that such a cell in the visual cortex may have as its immediate neighbor a cell which responds to a substantially different set of stimulus conditions. To date, the only assumption which seems adequate to account for this degree of specificity is that each type of cell and function which it subserves, as well as its location in the nervous system, is represented by a chemical code which can be detected and matched by the ingrowing fibers. The postulation of continued growth or replacement of these fibers in the adult brain is still conjectural, but receives increasing support from observations of degeneration and regeneration,* and studies of growth of new connections in tissue culture (c.f. Crain [4]). In our model we assume that each of the specific template sites on the neurofibrils is the result of induction of a specific gene during the differentiation of the cell. If some choice of 50 out of 100 genes were specifically induced in each cell, then this would permit the coding of $\binom{100}{50}$ different protein structures—a

* Of particular interest here are Wiesel and Hubel's studies on young kittens with light deprivation of one or both eyes resulting in changes in the responses of neurons in the visual cortex suggesting a competition among active fibers to active cells, as postulated in our model.

number far greater than the number of cells in the human central nervous system.

In each cell (*A* and *B*, in Fig. 5) it is assumed that a set of polypeptide fragments (or other molecular types) designated "codons" in the figure, by analogy to the genetic terminology, are produced, complementing the protein of that cell's intrasynaptic fibrils. It is conceivable that the fibril protein itself might serve as a template for the production of the appropriate codons. Thus if the fibril protein of cell *A* contains fifty different kinds of template sites, a corresponding fifty codon species will be produced in the cell, and it is supposed that they are stored in the endbulb, either freely in the cytoplasm or possibly in vesicles, along with transmitter substances. Whenever cell *A* fires, a small batch of these codons are assumed to be discharged into the synaptic cleft (again either by passive diffusion through the depolarized membrane, or by discharge of a vesicle, as in transmitter release). Similarly, when cell *B* fires, its characteristic codons are released. Outside the cell, the normal fate of these molecules is to be destroyed rapidly by a proteolytic enzyme (marked *E* in the Figure). This state of affairs is shown in the upper portion of the figure, in the region marked (1). The enzyme is assumed to attack the codon at its active site, marked *X*, which remains exposed when the codon latches to its corresponding template on one of the neurofibrils.

If both cells fire in unison or in immediate succession, however, a process illustrated at location (2) in the figure is assumed to take place. Each cell discharges its codons into the cleft, as before, and these tend to bind to the corresponding receptor sites on the proteins of cells *A* and *B*. Ordinarily, they would promptly be destroyed by the enzyme *E*. When a pair of presynaptic and post synaptic fillaments happen to cross one another in close proximity, however, the codons which are lined up back-to-back are assumed to form disulfide bonds (or other readily forming bonds) which combine them into a complex. In this form, they not only provide a covalently bonded bridge between the fibrils of cell *A* and the fibrils of cell *B*, but they are protected from enzyme *E*. Note that once such a structure is initiated, the activity of either cell alone is likely to add to it by providing additional codons which latch into the free positions at the ends, where they form part of the protected structure.

From time to time, portions of this complex may break off due to degeneration of synapses, enzyme action, or physical damage as in the preparation of brain homogenates. If one of these pieces should find its way to a matched set of templates on either a presynaptic or postsynaptic filament, it will bind to it, permitting a complementary filament to bind to the other side. As soon as this occurs, a new bridge has been established, and will continue to grow by accretion of additional codons from each of the two cells whenever they fire. In the normal, intact brain, this mechanism is the one which ultimately

leads to an increase in the effectiveness of cell *A* upon cell *B*, i.e., to an increment in the weight of the *A–B* connection. Suppose, for example, that cell *B* has a fixed number (4) of sybsynaptic sites available on its somatic membrane, as shown in Fig. 6. Two cells, *A* and *C*, each occupy two of these sites at the

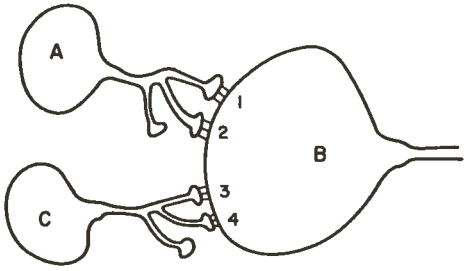


FIG. 6. Competition for synaptic sites.

outset, as shown in the figure, and each is assumed to have one extra unconnected ending available in the vicinity. Now suppose *A* and *B* are jointly active for a short period, resulting in the initiation of a number of adhesive complexes at the two *A–B* synapses. This leads to a strengthening of the *A–B* synapses, structurally, relative to the *C–B* synapses, but it does not alter *A*'s effect upon *B*, which depends entirely on the amount of transmitter substance released by *A* at its synaptic junctions. It does, however, stabilize *A*'s synapses, so that if a general decay process is initiated (either by controlled enzymatic triggering or as a result of a normal turnover process) one of the *C–B* synapses is likely to be lost before either of the *A–B* synapses. The "spare" nerve endings from cells *A* and *C*, which are assumed to be continuously probing the neighborhood for available points of attachment, now compete for the vacant synaptic site. If a fiber from cell *C* happens to make contact first, the previous situation is restored; but if the extra *A* fiber is the one to make contact, free pieces of adhesive complex from the neighboring sites, which are likely to occur in the local medium, will tend to stabilize it relative to the surviving *C–B* synapse, so that cell *A* now has three stable synapses on *B*, while cell *C* has only one relatively unstable synapse on *B*.

While the number of synapses in this illustration is unrealistically small (typical CNS cells might have over a thousand synapses), it serves to illustrate the principle of "survival of the stickiest" upon which this model depends. Note that a γ-system has been strictly observed, since it is assumed that each postsynaptic cell is limited to a certain fixed number of available sites for attachment of endbulbs, and this available space is at all times fully occupied by competing fibers. The conservation of space on the cell surface more or less guarantees that the γ-system will hold; even if the entire cell changes in

size, gaining new sites or losing old ones, they would tend to be reapportioned in such a way that the previously established control-biases would be maintained. In the extreme case of the death of one cell and its replacement by an analogously coded substitute, we might expect the mechanism for regeneration of analogous adhesive structures to reduplicate the bias which was previously introduced.

This model seems to provide a clear biological purpose for the transferability of acquired memory traces, since the model depends on the migration of complexes from existing sites to new sites for its performance. Moreover, several predictions can be made from the model which have a bearing on "memory transfer" experiments now in progress. First, the molecular weight of a complex required potentially to encode all of the synapses of the central nervous system can be calculated from any one of a number of initial assumptions. If we assume, for example, that each codon consists of seven amino acid residues (with a mean molecular weight of about 100 per residue), and that five of the seven amino acids characterize the binding site (the remaining two being cysteine, responsible for forming the disulfide bonds in the adhesive complex), then if only three kinds of amino acids are permitted as "code letters," there would be $3^5 = 243$ possible types of codon, with a molecular weight of 700. Four such codons in a complex would be sufficient to identify over 10^9 distinctive cell types, and eight would identify 10^{18} combinations of presynaptic and postsynaptic neurons, with a molecular weight of 5600. By experimenting on paper with several such models, we have typically found molecular weights in the neighborhood of 5000 or slightly less to be sufficient for the coding of all synapses in the human brain. By assuming species of molecules other than polypeptides, or by allowing a greater number of amino acid types to participate in active site designation, more information can be stored for lower molecular weights, but this is compensated by the necessity of allowing for the typical excess structure other than the active sites found in most biological molecules. In any case, the assumption of 5000 as a plausible weight seems to agree well with crude molecular weight determinations which were subsequently carried out by Sephadex chromatography on some of our extracts [15].

Another consideration which should be borne in mind in considering the probable biochemistry of the adhesive substance is that the codons must not be allowed to form complexes spontaneously within the parent cell. Several solutions to this problem are possible. One is that the binding sites of presynaptic and postsynaptic codons (assumed above to be cysteine in both cases) may be complementary rather than identical, so that they could only form complexes when combined in the synaptic cleft. Another possibility is that the codons assume a protected tertiary structure within the cell, the cysteine site only being exposed when the codon binds to the extracellular protein

filament. The stripping of a chemical mask from the binding site is a third possibility, suggesting a possible mechanism for enzymatic control over the memory process.

Since most of the adhesive substance in the brain is assumed to be bound, at any one time, to the intrasynaptic filaments, we might predict that a preparation of nerve-ending particles (obtained by differential centrifugation from a sucrose homogenate of brain tissue) would yield a highly active preparation in a transfer experiment. While this prediction remains to be tested, we have found that most of the activity seems to be associated with the precipitated particle fractions of brain homogenates after centrifugation, rather than remaining in the soluble phase. Prolonged washing with water or detergents, however, has been shown to progressively potentiate the soluble phase at the expense of the particle fractions [16]. This is at least consistent with, although not a proof of, the hypothesis of membrane binding.

As I have tried to emphasize throughout this paper, we can do no more at this time than speculate about the nature of the mechanism responsible for the "memory transfer" phenomenon. The purpose of such speculation is, on the one hand, to see how far we might have to modify previous thinking in order to accomodate the new findings, and on the other hand to provide some suggestions as to reasonable directions for biological experimentation. The transfer phenomenon gives us, for the first time, a technique by which many theoretical questions about the mechanisms of memory may ultimately be answered. Since the return to be expected at this time from further experimentation is likely to prove of far greater value than that to be obtained from further speculation, it seems reasonable to hold intensive work on the development of more detailed models in abeyance until some of the questions already raised here have been answered.

ACKNOWLEDGMENTS

The author is indebted to Trevor Barker and Rodman G. Miller for their assistance in the computer simulation of the various *C*-systems discussed in this paper, and to Professor R. D. O'Brien and the many members of his staff at Cornell University who have contributed to the experimental program on memory transfer. The Courant Institute of Mathematical Sciences and the AEC Division of Research have also been most helpful in providing computational facilities for much of the work reported here.

REFERENCES

[1] Babich, F., Jacobson, A. L., Bubash, S., and Jacobson, A., *Science* **149**, 656 (1965).
[2] Babich, F., Jacobson, A. L., and Bubash, S., *Proc. Natl. Acad. Sci. U.S.* **54**, 1299 (1965).
[3] Corning, W. C., and John, E. R., *Science* **134**, 1363 (1961).
[4] Crain, S., and Peterson, E. R., *Science* **141**, 427 (1963).

[5] Eccles, J. C., "The Physiology of Nerve Cells." Johns Hopkins Press, Baltimore, Maryland, 1957.
[6] Fjerdingstad, E. J., Nissen, Th., and Røigaard-Petersen, H. H., *Scand. J. Psychol.* **6**, 1 (1965).
[7] Gray, E. G., and Whittaker, V. P., *J. Anat.* **96**, 79 (1962).
[8] Hubel, D. H., and Wiesel, T. N., *J. Physiol.* **160**, 106 (1962).
[9] Hydén, H., *in* "Macromolecular Specificity and Biological Memory" (F. O. Schmidt, ed.). MIT Press, Cambridge, Massachusetts, 1962.
[10] McConnell, J. V., *J. Neuropsychiat.* **3**, 42 (1962).
[11] Robertson, J. D., *Neurosciences Res. Progr. Bull.* **3**, No. 4, 3 (1965).
[12] Rosenblatt, F., "Principles of Neurodynamics." Spartan Books, Washington, D.C., 1962.
[13] Rosenblatt, F., *in* "Computer and Information Sciences" (J. T. Tou and R. H. Wilcox, eds.). Spartan Books, Washington, D.C., 1964.
[14] Rosenblatt, F., Farrow, J. T., and Herblin, W. F., *Nature* **209**, 46 (1966).
[15] Rosenblatt, F., Farrow, J. T., and Rhine, S., *Proc. Natl. Acad. Sci. U.S.* **55**, 548, 787 (Parts I and II) (1966).
[16] Rosenblatt, F., and Miller, R. G., *Proc. Natl. Acad. Sci. U.S.* **56**, 1423, 1683 (Parts I and II) (1966).
[17] Ungar, G., and Oceguera-Navarro, C., *Nature* **207**, 301 (1965).

Some Approaches to Optimum Feature Extraction*

Julius T. Tou† and Richard P. Heydorn
INFORMATION SCIENCE RESEARCH CENTER
BATTELLE MEMORIAL INSTITUTE
COLUMBUS LABORATORIES
COLUMBUS, OHIO

The basic problem of feature extraction is divided into two general categories: intraset feature extraction and interset feature extraction. The intraset feature extraction problem is studied from three points of view; namely, estimation, clustering, and minimization of population entropy. The interset feature extraction problem is approached from the concept of divergent information. A class of linear transformations is proposed to reduce the dimensionality of measurement vectors and at the same time to maximize or minimize a performance criterion function describing the information transmitted by the patterns or some measure of error resulting from the feature extraction process. The patterns are assumed to originate from a normal multivariate distribution. The transformed patterns represent a set of feature vectors the elements of which describe the important properties of the patterns and provide the necessary information for discriminating between pattern classes. The proposed approaches are applied to feature extraction and recognition of alphabetic characters as an illustration, and computer simulation results are obtained.

I. Introduction

Among the challenging problems in the design of pattern recognition systems, two problems are of utmost importance: (1) the extraction of pattern features, and (2) the optimum classification of pattern classes. The first is concerned with the problem of what to measure, and the second deals with the problem of making optimum decisions in classification. During the past decade, considerable work has been done in solving the optimal classification problem. Various theories and techniques have been developed on the basis of modern mathematics. The literature is well documented with research reports,

* The work reported here was supported in part by the Office of Naval Research.
† Also with the Department of Electrical Engineering, The Ohio State University, Columbus, Ohio.