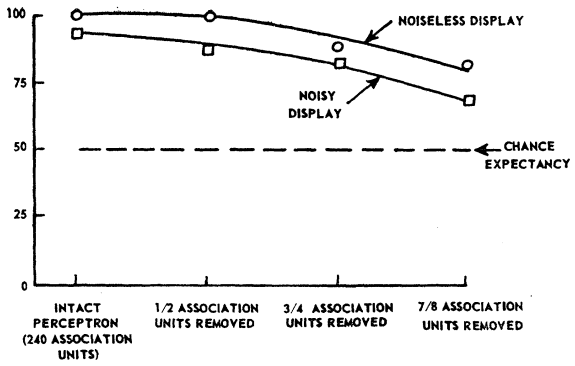

FIG. 13. Effect of association unit removal on trained "E"–"X" discrimination.

simple perceptron of Fig. 4 is no longer adequate. We shall find certain temporal effects in the paper which follows, but for others it is necessary to introduce time delays into the system.[1] A speech recognizing perceptron which utilizes such delays is currently being built at Cornell University.

Other activities now in progress[1] include quantitative studies of cross-coupled and multi-layer systems (by means of analysis and digital simulation), studies of selective attention mechanisms, the effects of geometric constraints on network organization, new types of reinforcement rules, and attempts at relating this research to biological data. Work is also in progress on development of electrolytic and other low-cost integrating devices and additional electronic components necessary for the construction of large-scale physical models.

It is clear that we are still far from the point of understanding how the brain functions. It is equally clear, we believe, that a promising road is open for further investigation.

REVIEWS OF MODERN PHYSICS VOLUME 34, NUMBER 1 JANUARY, 1962

Analysis of a Four-Layer Series-Coupled Perceptron. II*

H. D. BLOCK, B. W. KNIGHT, JR., AND F. ROSENBLATT

Cornell University, Ithaca, New York

1. INTRODUCTION

THE preceding paper[1] presented motivation and background for the general subject of perceptrons and gave some analysis and results for a simple three-layer perceptron. While it has been shown there that it is possible to associate any arbitrary set of responses to an arbitrary set of stimuli in a simple three-layer perceptron, such a perceptron characteristically requires a large representative sample of each kind of pattern (e.g., letters "A" and "B"), covering all parts of the retina, before it will recognize an arbitrarily positioned stimulus which is similar to one which it has seen before. In other words, a three-layer perceptron has no concept of "similarity" based on any criterion other than the intersections of sets of retinal elements. In a previous paper,[2] Rosenblatt has shown that a "cross-coupled perceptron," in which A units are connected to one another by modifiable connections, should tend to develop an improved similarity criterion for generalizing responses from one stimulus to another when exposed to a suitably organized environment. In this paper a simpler network, consisting of four layers of units but without cross coupling, is analyzed in a more rigorous fashion, and is shown to possess the same property.

The perceptron of the present paper is "self-organizing" in the sense that during the training period the experimenter does not tell the machine the category of each stimulus. As the analysis below will show, the only contact between the experimenter and the machine is the presentation of the stimuli.

2. THE MODEL

The model to be analyzed here is a four-layer perceptron of the schematic type S–A^{I}–A^{II}–R, as indicated in Fig. 1.

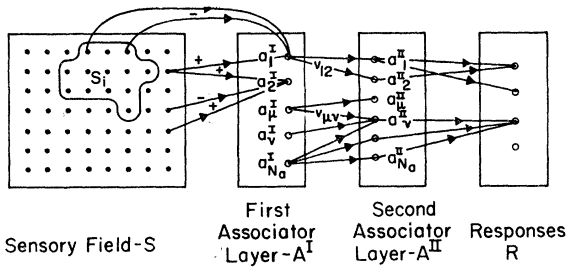

FIG. 1. Organization of four-layer series-coupled perceptron.

* Research sponsored by the Office of Naval Research.

[1] H. D. Block, Revs. Modern Phys. 34, 123 (1962).

[2] See, F. Rosenblatt, in *Self-Organizing Systems*, edited by M. Yovits and S. Cameron (Pergamon Press, New York, 1960).

There are n possible stimuli to be presented to the sensory field: S_1, S_2, $\cdots$, S_i, $\cdots$, S_n. The values of the S to A^{I} connections do not change with time. The A^{II} units are in one-to-one correspondence with the A^{I} units and have threshold θ. An active A^{I} unit $a_\mu{}^{\mathrm{I}}$ delivers a signal of θ to its corresponding A^{II} unit $a_\mu{}^{\mathrm{II}}$ and also a signal $v_{\mu\nu}$ to $a_\nu{}^{\mathrm{II}}$ for $\nu=1, 2, \cdots, N_a$ [see Eq. (1) below]. An inactive unit puts out no signal. We shall use the indices i, j, k throughout to designate various stimuli, while the indices μ and ν will be used to designate associator units. The values of $v_{\mu\nu}$ are initially zero and change with time as follows. Stimuli are presented at times 0, Δt, $2\Delta t$, $3\Delta t$, $\cdots$. If $a_\mu{}^{\mathrm{I}}$ is active at time t and $a_\nu{}^{\mathrm{II}}$ is active at time $t+\Delta t$, then $v_{\mu\nu}$ receives an increment $(\eta\cdot\Delta t)$; otherwise it does not receive this increment. At the same time each $v_{\mu'\nu'}$ is decremented by $(\delta\cdot\Delta t)v_{\mu'\nu'}$. [See Eq. (5), below.] These two effects represent a facilitation of used pathways and a decay, respectively. The A^{II} units are connected to response units with connections whose weights may be varied according to one of the standard rules of reinforcement. There is no time delay of transmission of signals through the system.

3. ANALYSIS

From the viewpoint of the R units we have a simple three-layer perceptron as described in the preceding paper[1] with A^{II} as the associator set. Since the behavior of such a system is well understood in terms of the sets of associators in A^{II} activated by the various stimuli, our main concern here is to find the nature of these sets.

The set of associators in A^{I} responding to S_i is denoted by $A^{\mathrm{I}}(S_i)$; the set responding to both S_i and S_j is $A^{\mathrm{I}}(S_i)\bigcap A^{\mathrm{I}}(S_j)$. The number of associators in A^{I} responding to both S_i and S_j is denoted by $n_{ij}{}^{\mathrm{I}}$ and is equal to the number of elements in $A^{\mathrm{I}}(S_i)\bigcap A^{\mathrm{I}}(S_j)$:

$$n_{ij}{}^{\mathrm{I}}=\sum_\mu e_{\mu i}e_{\mu j}$$

where

$$e_{\mu i}=\begin{cases}1 & \text{if } a_\mu{}^{\mathrm{I}}\in A^{\mathrm{I}}(S_i)\\ 0 & \text{if } a_\mu{}^{\mathrm{I}}\notin A^{\mathrm{I}}(S_i).\end{cases}$$

None of these quantities change in time.

Let $\alpha_\mu{}^{(i)}(t)$ denote the total input signal to association unit $a_\mu{}^{\mathrm{II}}$ at time t, if stimulus S_i were to be applied to the sensory field at time t. Then

$$\alpha_\nu{}^{(i)}(t)=\theta e_{\nu i}+\sum_\mu v_{\mu\nu}(t)e_{\mu i}. \tag{1}$$

Let

$$\beta_\nu{}^{(i)}=\theta e_{\nu i} \tag{2}$$

and

$$\gamma_\nu{}^{(i)}(t)=\sum_\mu v_{\mu\nu}(t)e_{\mu i}. \tag{3}$$

Then

$$\alpha_\nu{}^{(i)}(t)=\beta_\nu{}^{(i)}+\gamma_\nu{}^{(i)}(t). \tag{4}$$

Note that $\beta_\nu{}^{(i)}$ is θ or 0 according as $a_\nu{}^{\mathrm{I}}$ is in $A^{\mathrm{I}}(S_i)$ or not; it does not change with time. On the other hand $\gamma_\nu{}^{(i)}(t)$ represents the effect of the variable (A^{I} to A^{II}) connections whose values are $v_{\mu\nu}(t)$.

Suppose that at time t_0 stimulus S_j is presented and at time $t_0+\Delta t$ stimulus S_k is presented. Then the consequent change in $v_{\mu\nu}$ will be

$$v_{\mu\nu}(t_0+2\Delta t)-v_{\mu\nu}(t_0+\Delta t)$$
$$=(\eta\cdot\Delta t)(e_{\mu j})\phi[\alpha_\nu{}^{(k)}(t_0+\Delta t)]-(\delta\cdot\Delta t)v_{\mu\nu}(t_0+\Delta t), \tag{5}$$

where

$$\phi(x)=\begin{cases}0 & \text{if } x<\theta\\ 1 & \text{if } x\geqq\theta.\end{cases}$$

From (3) and (5) we get

$$\gamma_\nu{}^{(i)}(t_0+2\Delta t)-\gamma_\nu{}^{(i)}(t_0+\Delta t)$$
$$=\sum_\mu[v_{\mu\nu}(t_0+2\Delta t)-v_{\mu\nu}(t_0+\Delta t)]e_{\mu i}$$
$$=(\eta\cdot\Delta t)\phi[\alpha_\nu{}^{(k)}(t_0+\Delta t)]$$
$$\times\sum_\mu e_{\mu j}e_{\mu i}-(\delta\cdot\Delta t)\sum_\mu v_{\mu\nu}(t_0+\Delta t)e_{\mu i}.$$

Hence

$$\gamma^{(i)}(t_0+2\Delta t)-\gamma^{(i)}(t_0+\Delta t)$$
$$=(\eta\cdot\Delta t)\phi[\alpha^{(k)}(t_0+\Delta t)]n_{ij}{}^{\mathrm{I}}-(\delta\cdot\Delta t)\gamma^{(i)}(t_0+\Delta t), \tag{6}$$

where, for brevity, we have dropped the subscript ν, and will remember that the γ and α refer to any particular associator $a_\nu{}^{\mathrm{II}}$. Now suppose the sequence of stimuli $S_{j_0}, S_{j_1}, \cdots, S_{j_M}$ is presented at the successive times t, $t+\Delta t$, $\cdots$, $t+M\Delta t$. In Eq. (6) we take $t_0=t+m\Delta t$, $[m=0,1,2,\cdots,(M-1)]$, $j=j_m$, $k=j_{m+1}$; and obtain

$$\gamma^{(i)}[t+(m+2)\Delta t]-\gamma^{(i)}[t+(m+1)\Delta t]$$
$$=(\eta\cdot\Delta t)\phi\{\alpha^{j_{m+1}}[t+(m+1)\Delta t]\}n_{ij_m}{}^{\mathrm{I}}$$
$$-(\delta\cdot\Delta t)\gamma^{(i)}[t+(m+1)\Delta t]. \tag{7}$$

Summing on m from 0 to $M-1$, we get the change in $\gamma^{(i)}$ due to the entire sequence of stimuli:

$$\gamma^{(i)}[t+(M+1)\Delta t]-\gamma^{(i)}(t+\Delta t)$$
$$=\sum_{m=0}^{M-1}[(\eta\cdot\Delta t)\cdot\phi\{\alpha^{(j_{m+1})}[t+(m+1)\Delta t]\}n_{ij_m}{}^{\mathrm{I}}$$
$$-(\delta\cdot\Delta t)\gamma^{(i)}[t+(m+1)\Delta t]]. \tag{8}$$

We divide by $M\Delta t$ and let Δt approach zero to obtain

$$\frac{d\gamma^{(i)}}{dt}=\sum_{m=0}^{M-1}\frac{\eta}{M}\phi[\alpha^{(j_{m+1})}(t)]n_{ij_m}{}^{\mathrm{I}}-\delta\gamma^{(i)}(t). \tag{9}$$

Let F_{jk} be the number of times the pair S_jS_k occurs in the given sequence $S_{j_0}, S_{j_1}, \cdots, S_{j_M}$; and let $f_{jk}=F_{jk}/M$ be the average frequency of the pair S_jS_k. Then from (9) we get

$$\frac{d\gamma^{(i)}}{dt}=\sum_j\sum_k\eta f_{jk}\phi[\alpha^{(k)}(t)]n_{ij}{}^{\mathrm{I}}-\delta\gamma^{(i)}(t). \tag{10}$$

Defining the matrix

$$K_{ij}=\sum_{k=1}^{n}n_{ik}{}^{\mathrm{I}}f_{kj}$$

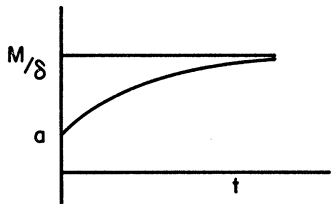

FIG. 2. Graph of the solution of $x=(M/\delta)-e^{-\delta t}[(M/\delta)-a]$.

we have from (10)

$$\frac{d\gamma^{(i)}}{dt}=\eta\sum_j K_{ij}\phi[\beta^{(j)}+\gamma^{(j)}(t)]-\delta\gamma^{(i)}(t). \qquad (11)$$

This is a system of nonlinear differential equations for $\gamma^{(1)}(t), \cdots, \gamma^{(n)}(t)$, with initial conditions $\gamma^{(i)}(0)=0$.

If the f_{kj} vary with t, then K_{ij} are time dependent but in any case they are non-negative and bounded; ϕ is non-negative, monotone increasing in γ, bounded and continuous on the right. We shall treat here only the case in which the K_{ij} are constants.

Before discussing the solution of (11) we consider the equilibrium equation

$$\gamma^{(i)}=\frac{\eta}{\delta}\sum_j K_{ij}\phi[\beta^{(j)}+\gamma^{(j)}]. \qquad (12)$$

The system of equations (12) may have more than one solution. However we shall show that there is a unique *minimal* solution (by this we mean a solution, none of whose components $\gamma^{(i)}$ exceed the corresponding components of another solution); and this minimal solution is obtained in a finite number (at most n) of iterations of (12), starting with all $\gamma^{(j)}=0$ on the right side and then finding the new values of $\gamma^{(j)}$ from (12), putting these back into the right side and so on. That is, we take $\gamma_0{}^{(i)}=0$ and

$$\gamma_{m+1}{}^{(i)}=\frac{\eta}{\delta}\sum_j K_{ij}\phi[\beta^{(j)}+\gamma_m{}^{(j)}]. \qquad (13)$$

We prove first that the process terminates in at most n iterations. This can be seen from the following considerations. Since the right-hand side of (13) is non-negative and $\gamma_0{}^{(i)}=0$, it follows that $\gamma_1{}^{(i)}\geqq\gamma_0{}^{(i)}$. Now since the right-hand side of (13) is a nondecreasing function of the γ's, it follows that

$$\gamma_2{}^{(i)}\geqq\gamma_1{}^{(i)}, \cdots, \gamma_{m+1}{}^{(i)}\geqq\gamma_m{}^{(i)}.$$

Therefore also

$$\phi[\beta^{(j)}+\gamma_{m+1}{}^{(j)}]\geqq\phi[\beta^{(j)}+\gamma_m{}^{(j)}],$$

that is, successive ϕ's cannot decrease. If, at a particular step no ϕ increases then we are at a solution. The ϕ's have only the values zero or one, so even if only a single ϕ changes at each step, the process terminates in at most n steps.

The solution of (12) thus obtained we denote by $\gamma^{(i)*}$. We now prove that this solution is minimal. Let

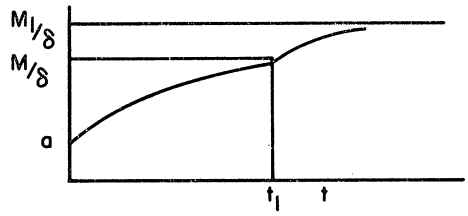

FIG. 3. Graph of the solution of $x=(M/\delta)-e^{-\delta t}[(M/\delta)-a]$ if at time $t=t_1$ we replace M by $M_1>M$.

$\tilde{\gamma}^{(i)}$ be any solution of the equilibrium equation (12). Clearly $\tilde{\gamma}^{(i)}\geqq 0$. Then for the iteration process (13), we have $\gamma_0{}^{(i)}\leqq\tilde{\gamma}^{(i)}$, for all i. Since the right-hand side of (13) is a monotone function of $\gamma^{(i)}$, we have

$$\gamma_1{}^{(i)}=\frac{\eta}{\delta}\sum_j K_{ij}\phi[\beta^{(j)}+\gamma_0{}^{(j)}]$$

$$\leqq\frac{\eta}{\delta}\sum_j K_{ij}\phi[\beta^{(j)}+\tilde{\gamma}^{(j)}]=\tilde{\gamma}^{(i)}.$$

Similarly, $\gamma_m{}^{(i)}\leqq\tilde{\gamma}^{(i)}$, hence $\gamma^{(i)*}\leqq\tilde{\gamma}^{(i)}$. Hence $\gamma^{(i)*}$ is minimal.

To avoid consideration of a special pathological case, we make a mild assumption. Consider the sum

$$(\eta/\delta)\sum_{j\in\mathcal{S}} K_{ij}$$

taken over a subset $\mathcal{S}$ of the possible values of j, $(1,2,\cdots,n)$. We assume that no such sum is equal to θ. This is not a serious assumption, since by an arbitrarily small change in η/δ we can satisfy this requirement.

Now suppose that the $\gamma^{(i)}(t)$ satisfy the system of differential equations (11) and the initial conditions $\gamma^{(i)}(0)=0$. Then we assert that the $\gamma^{(i)}(t)$ are nondecreasing and $\lim_{t\to\infty}[\gamma^{(i)}(t)]=\gamma^{(i)*}$. That is, the solution obtained by the iterative process (13) is indeed the solution of the differential equation (11), with initial conditions zero in each case.

First we shall show that $d\gamma^{(i)}/dt\geqq 0$. Moreover,

$$\text{if} \quad \gamma^{(i)}(t)>0 \quad \text{then} \quad d\gamma^{(i)}(t)/dt>0. \qquad (A)$$

As a preliminary step, consider the nature of the solution of the equation $dx/dt=M-\delta x$, where M and δ are positive constants and $x(0)=a$, where $0\leqq a<M/\delta$. The solution $x=(M/\delta)-e^{-\delta t}[(M/\delta)-a]$, has the appearance of Fig. 2. The solution approaches M/δ monotonely from below, and $dx/dt>0$ for all $t>0$. If at time $t=t_1$, we replace M by $M_1>M$ the solution appears as Fig. 3.

As t goes from 0 to t_1 the solution approaches M/δ monotonely from below; as t increases beyond t_1 the solution approaches M_1/δ monotonely from below. The solution is continuous; so is its derivative, except at t_1, where the left- and right-hand derivatives are not equal; but both are positive.

If instead of $M>\delta a\geqq 0$, we take $M=a=0$, the solution $x(t)=0$ for $0\leqq t\leqq t_1$.

We now proceed to the proof of (A). Let

$$M^{(i)}(t)=\eta\sum_j K_{ij}\phi[\beta^{(j)}+\gamma^{(j)}(t)].$$

Then (11) can be written

$$d\gamma^{(i)}/dt=M^{(i)}(t)-\delta\gamma^{(i)}(t),$$

where here and in the following paragraph, i is a generic index of the set $(1,2,\cdots,n)$, while j and k will refer to specific indices to be defined below.

Each function $M^{(i)}(t)$ can take on at most 2^n possible values. Let k be a specific value of i and suppose first that $M^{(k)}(0)>0$. The only times at which $M^{(i)}(t)$ can change its value are when one of the $\gamma^{(i)}$ (indeed one whose corresponding $\beta^{(i)}=0$) reaches the value θ. Suppose the first time at which this happens is $t_1>0$. Suppose then that $\gamma^{(j)}(t_1)=\theta$. Since in the interval $0<t<t_1$ all $d\gamma^{(i)}/dt\geqq 0$, we have $M^{(i)}(t_1)\geqq M^{(i)}(t_0)$. Thus the solution $\gamma^{(k)}(t)$ appears as in Fig. 3; in particular for all k such that $M^{(k)}(0)>0$, we have $\gamma^{(k)}(t_1)<M^{(k)}(0)/\delta\leqq M^{(k)}(t_1)/\delta$; and for the others $\gamma^{(i)}(t_1)\leqq M^{(i)}(t_1)/\delta$. Furthermore, since both the left and right derivatives of $\gamma^{(j)}$ at t_1 are positive we have, for $t>t_1$ and sufficiently close to t_1, $\gamma^{(j)}(t)>\theta$, so that it will not be until time t_2, with $t_2>t_1$, that there will again be a $\gamma^{(i)}(t)$ having the value θ. In the interval $t_1<t<t_2$ we have the same pertinent conditions as we had in the interval $0<t<t_1$; namely $d\gamma^{(i)}/dt=M^{(i)}(t_1)-\delta\gamma^{(i)}(t)$, with initial values $\gamma^{(i)}(t_1)\leqq M^{(i)}(t_1)/\delta$ and in particular $\gamma^{(k)}(t_1)<M^{(k)}(t_1)/\delta$. Thus in the interval $(t_1<t<t_2)$ we again have $d\gamma^{(i)}/dt\geqq 0$, and $d\gamma^{(k)}/dt>0$. We repeat the same argument for the successive intervals (t_2,t_3), (t_3,t_4) and so on. Since the $\gamma^{(i)}(t)$ are monotone there are at most n such intervals.

If $M^{(k)}(0)=0$ then $\gamma^{(k)}(t)=0$ for $0\leqq t\leqq t_1$. If $M^{(k)}(t_1)>0$, then we use the previous argument starting at $t=t_1$; otherwise $\gamma^{(k)}$ remains zero at least until t_2 and so on. In any case we have proved (A).

Next we show that

$$\lim_{t\to\infty}\gamma^{(i)}(t)=\gamma^{(i)*};\quad i=1,2,\cdots,n. \tag{B}$$

Since, from the proof of (A) it is clear that each $\gamma^{(i)}(t)$ is monotone and bounded, $\lim_{t\to\infty}\gamma^{(i)}(t)$ exists; call it $\gamma^{(i)\ddagger}$; it is a sum of the form

$$(\eta/\delta)\sum_{j\in S}K_{ij}$$

(which was assumed at the outset to be unequal to θ) and thus $\gamma^{(i)\ddagger}\neq\theta$. Therefore, $\phi[\beta^{(j)}+\gamma^{(j)}]$ is continuous when $\gamma^{(j)}=\gamma^{(j)\ddagger}$. Letting $t\to\infty$ in Eq. (11) we see that $\gamma^{(i)\ddagger}$ is a solution of the equilibrium equation (12). Hence $\gamma^{(i)\ddagger}\geqq\gamma^{(i)*}$, since $\gamma^{(i)*}$ is minimal. We next show that for all $t\geq 0$, $\gamma^{(i)}(t)\leqq\gamma^{(i)*}$.

Note that initially $\gamma^{(i)}(0)\leqq\gamma^{(i)*}$. Suppose that t_1 is the first time at which some $\gamma^{(k)}(t)=\gamma^{(k)*}$. From (11) and the fact that ϕ is nondecreasing we see that at t_1

$$\begin{aligned}\frac{d\gamma^{(k)}}{dt}&=\eta\sum_j K_{kj}\phi[\beta^{(j)}+\gamma^{(j)}(t_1)]-\delta\gamma^{(k)}(t_1)\\&\leqq\eta\sum_j K_{kj}\phi[\beta^{(j)}+\gamma^{(j)*}]-\delta\gamma^{(k)}(t_1)\\&=\delta\gamma^{(k)*}-\delta\gamma^{(k)}(t_1)=0,\end{aligned}$$

i.e., $d\gamma^{(k)}/dt\leqq 0$ at $t=t_1$. If $\gamma^{(k)*}>0$, we see from (A) that $d\gamma^{(k)}/dt>0$ at t_1, a contradiction. Suppose that $\gamma^{(k)*}=0$, so that also $t_1=0$. Then, as long as no $\gamma^{(i)}(t)$ reaches a nonzero $\gamma^{(i)*}$, we have, since ϕ is a nondecreasing function of its argument,

$$\begin{aligned}M^{(k)}(t)&=\eta\sum_j K_{kj}\phi[\beta^{(j)}+\gamma^{(j)}(t)]\\&\leqq\eta\sum_j K_{kj}\phi[\beta^{(j)}+\gamma^{(j)*}]=\delta\gamma^{(k)*}=0.\end{aligned}$$

Hence over this period $\gamma^{(k)}(t)=0$. But no nonzero $\gamma^{(i)*}$ can ever be attained by $\gamma^{(i)}(t)$, since, by the above argument, we would have $d\gamma^{(i)}/dt\leqq 0$ at the first time it occurred, in contradiction to (A).

Thus we have shown that: if $\gamma^{(i)*}>0$, then $\gamma^{(i)}(t)<\gamma^{(i)*}$; and if $\gamma^{(i)*}=0$, then $\gamma^{(i)}(t)=\gamma^{(i)*}$. In general $\gamma^{(i)}(t)\leqq\gamma^{(i)*}$.

Hence $\gamma^{(i)\ddagger}=\lim_{t\to\infty}\gamma^{(i)}(t)\leqq\gamma^{(i)*}$ and (B) follows.

From this point on we shall be concerned with the steady-state values $\gamma^{(i)*}$, and, for brevity, we shall drop the *. In the terminal condition the associator a_ν^{II} is activated by S_i if $\beta_\nu^{(i)}+\gamma_\nu^{(i)}\geqq\theta$. The set of A^{II} associators which are activated by stimulus S_i will be denoted by $A^{\mathrm{II}}(S_i)$. In the initial state, the set $A^{\mathrm{II}}(S_i)$ is denoted by $A_0{}^{\mathrm{II}}(S_i)$, and in the terminal state by $A_\infty{}^{\mathrm{II}}(S_i)$. The number of A^{II} associators which are activated by both S_i and S_j is called $n_{ij}{}^{\mathrm{II}}$ and is equal to the number of units in $A^{\mathrm{II}}(S_i)\bigcap A^{\mathrm{II}}(S_j)$.

Once the $n_{ij}{}^{\mathrm{II}}$ are known, the behavior of the perceptron can be predicted, as outlined in the preceding paper. To determine the $n_{ij}{}^{\mathrm{II}}$ we might proceed as follows. First the set of A^{I} associators is broken into the smallest subdivisions of the Venn diagram representing the sets responding to different combinations of stimuli. Each of these cells is characterized by a certain β vector. For each such β vector we solve Eq. (12) for the terminal values of $\gamma^{(i)}$. Here we regard $n_{ik}{}^{\mathrm{I}}$ and f_{kj} as given. (In the present paper $n_{ik}{}^{\mathrm{I}}$ represents the actual number of A^{I} units responding to both S_i and S_k in the particular perceptron being studied. In general, if the S to A^{I} connections are chosen at random, subject to statistical constraints, then the $n_{ik}{}^{\mathrm{I}}$ represent random variables whose expected values are $N_aQ_{ik}{}^{\mathrm{I}}$, where $Q_{ik}{}^{\mathrm{I}}$ is the probability that an associator $a_\mu{}^{\mathrm{I}}$ is activated by both S_i and S_k. These random variables have been studied rather thoroughly for several main classes of networks.[3]) Initially, $n_{ij}{}^{\mathrm{II}}=n_{ij}{}^{\mathrm{I}}$. Knowing $\beta^{(i)}$ and $\gamma^{(i)}$ we can compute the region of the A^{II} Venn diagram to which

[3] F. Rosenblatt, *Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms* (Spartan Books, Washington, D. C., 1961).

each of the A^{II} units moves. Then we have the complete terminal distribution in the Venn diagram of A^{II} and hence in particular the n_{ij}^{II}. While this program is simple in principle the actual computations can be quite formidable. It can be seen in advance that the motion will be for associators to go into regions of higher intersections but points which are initially outside all the $A^{II}(S_i)$ will stay outside all the $A^{II}(S_i)$.

The analysis is approximate, of course, in that we have replaced the difference equation (8) by the differential equation (9). For small increments of reinforcement $\eta\Delta t$ and small increments of decay $(\delta\Delta t)v$ at each step the behavior of the discretized system will be approximated by that of the continuous one.

In this way the performance of these perceptrons can be analyzed for a variety of particular cases.[3,4] We shall instead consider here two "training programs" in which, by a suitable choice of parameters, the motion of the β vectors can be controlled and the performance predicted.

4. APPLICATIONS

Let P_j denote the fraction of occurrences of S_j in the given sequence $S_{j_0}, S_{j_1}, \cdots, S_{j_M}$ and let P_{jk} denote number of times S_k immediately follows S_j divided by the number of times S_j occurs. Then in a long sequence the equilibrium equation (12) takes the form

$$\gamma^{(i)}=\frac{\eta}{\delta}\sum_j\sum_k n_{ij}^{I}P_jP_{jk}\phi[\beta^{(k)}+\gamma^{(k)}], \qquad (14)$$

where P_j corresponds to the probability of S_j, and P_{jk} corresponds to the transition probability

$$S_j \rightarrow S_k=\text{Prob}(S_jS_k|S_j).$$

Training Program I

The stimuli are divided into two classes: $\{S_1,S_2,\cdots,S_K\}$ is class X, while $\{S_{K+1},\cdots,S_n\}$ is class Y. There is assumed to be no appreciable difference in the retinal overlaps; we assume $n_{ij}^{I}=(q+s\delta_{ij})$, where $s>0$, $q\geqq 0$. Thus the diagonal elements of the n_{ij}^{I} matrix are all $(q+s)$ and all other elements are q. (Note that by raising thresholds of the A^{I} units, the ratio q/s can be made as small as desired.) We shall assume that the probability of transition to a member of the same class is p, nearly unity, and to a member of the opposite class is $(1-p)$, nearly zero. Inside a class all members are equally likely. Thus

$$P_j=\begin{cases}1/2K & \text{for } S_j \text{ in } X\\ 1/2L & \text{for } S_j \text{ in } Y\end{cases}$$

where $L=n-K$;

$$P_{jk}=\begin{cases}p/K & \text{for } S_j \text{ in } X,\ S_k \text{ in } X,\\ (1-p)/L & \text{for } S_j \text{ in } X,\ S_k \text{ in } Y,\\ p/L & \text{for } S_j \text{ in } Y,\ S_k \text{ in } Y,\\ (1-p)/K & \text{for } S_j \text{ in } Y,\ S_k \text{ in } X.\end{cases}$$

Then we get from (14)

$$\gamma^{(i)}=\frac{\eta}{\delta}\Big[\sum_{j=1}^{K}\sum_{k=1}^{K}+\sum_{j=1}^{K}\sum_{k=K+1}^{n}+\sum_{j=K+1}^{n}\sum_{k=1}^{K}$$

$$+\sum_{j=K+1}^{n}\sum_{k=K+1}^{n}\{(q+s\delta_{ij})P_jP_{jk}\phi(\beta^{(k)}+\gamma^{(k)})\}\Big],$$

$$=\frac{\eta}{\delta}\Big[\frac{q}{2K}\sum_{k=1}^{K}\phi(\beta^{(k)}+\gamma^{(k)})+\frac{q}{2L}\sum_{k=K+1}^{n}\phi(\beta^{(k)}+\gamma^{(k)})$$

$$+sP_i\sum_{k=1}^{n}P_{ik}\phi(\beta^{(k)}+\gamma^{(k)})\Big]. \qquad (15)$$

Let us now assume that S_x is one of the stimuli of class X. Then

$$\gamma^{(x)}=\frac{\eta}{\delta}\Big[\Big(\frac{sp+qK}{2K^2}\Big)\sum_{k=1}^{K}\phi(\beta^{(k}+\gamma^{(k)})$$

$$+\Big(\frac{s(1-p)+qK}{2KL}\Big)\sum_{k=K+1}^{n}\phi(\beta^{(k)}+\gamma^{(k)})\Big]. \qquad (16)$$

We now observe the following

(a) If $\eta(sp+qK)/2\delta K^2\geqq\theta$, then

$$A_\infty^{II}(S_x)\supseteq\bigcup_{S_j\in X}A_0^{II}(S_j).$$

In words, if the stated inequality holds, then, in the terminal condition, *each* of the stimuli of class X activates the union of all sets which were initially activated by *any* of the stimuli of class X. [See Fig. 4(a).]

The proof follows from the fact that any associator which originally responded to any of the stimuli in

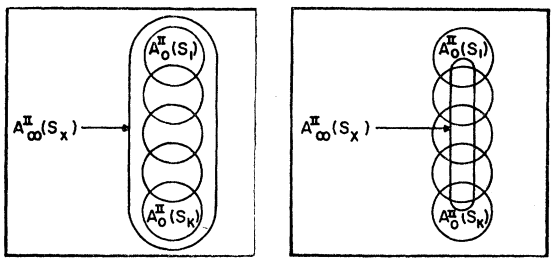

FIG. 4. Superimposed initial and terminal Venn diagrams of A^{II} layer, for two cases.

[4] H. D. Block, B. W. Knight, Jr., and F. Rosenblatt, Paper No. 1. Cognitive Systems Research Project. Cornell University, Ithaca, New York (July, 1960).

class X has $\beta^{(k)}=\theta$ for some $k\leqq K$. Hence there is a nonzero term in $\sum_{k=1}^{K}$ of (16). The postulated inequality then guarantees that the associator will be active in the terminal state.

(b) If $\eta[s(1-p)+qK]/2K\delta<\theta$, then

$$A_\infty^{\text{II}}(S_x) \subseteqq \bigcup_{S_j\in X} A_0^{\text{II}}(S_j).$$

In words, if the stated inequality holds, then, in the terminal condition, no stimulus in class X activates any associator outside of the union of sets initially activated by stimuli of class X. [See Fig. 4(b).]

The proof follows from the fact that, if we were to solve (16) by iteration, then any associator which is activated by none of the X stimuli has, on the first pass, no contribution from the term

$$\sum_{k=1}^{K} \text{ in Eq. (16).}$$

Thus on the first iteration we get from (16)

$$\gamma_1^{(x)}=\left(\frac{\eta}{\delta}\right)\left[\frac{s(1-p)+qK}{2KL}\right]\sum_{k=K+1}^{n}\phi(\beta^{(k)}+\gamma^{(k)}) \leqq \frac{\eta}{\delta}\left[\frac{s(1-p)+qK}{2K}\right]<\theta \quad (17)$$

in virtue of the assumed inequality (b). Similarly, on the next iteration, the term

$$\sum_{k=1}^{K} \text{ in Eq. (16)}$$

is again zero for such an associator and, as in (17), $\gamma_2^{(x)}<\theta$. Since there are only a finite number (indeed less than L) iterations we have $\gamma^{(x)}<\theta$ for such an associator and this associator does not become active.

(c) If the inequalities of both (a) and (b) hold then

$$A_\infty^{\text{II}}(S_x)=\bigcup_{S_j\in X} A_0^{\text{II}}(S_j).$$

Necessary and sufficient conditions for both (a) and (b) to hold are easily found. They are

(i) $s>qK(K-1)$,

(ii) $p>[Ks+qK(K-1)]/s(K+1)$,

(iii) $K^2/(sp+qK)\leqq\eta/2\theta\delta<K/(s(1-p)+qK)$.

Condition (i) ensures that a p, $(0<p<1)$ can be chosen to satisfy (ii). Condition (ii) ensures that $\eta/2\theta\delta$ can be chosen to satisfy (iii). The conditions can be put in the alternative form

(i′) $p>K/(K+1)$

(ii′) $s>qK(K-1)/[p(K+1)-K]$

and (iii).

If the parameters satisfy the stated inequalities then

$$A_\infty^{\text{II}}(S_x)=\bigcup_{S_j\in X} A_0^{\text{II}}(S_j).$$

This means that in the terminal state each stimulus of class X activates the same set of A^{II} units. Similarly there is a second set of A^{II} units activated by each member of class Y. If the A^{II} to R connections are random then, in general, one pattern of activated R units will respond to all stimuli of class X and another pattern of activated R units will respond to all stimuli of class Y. Thus the machine has dichotomized the classes, its output being in terms of this intrinsic code on the A^{II} units. Alternatively, with a single R unit and zero initial values on the A^{II} to R connections, a single corrective reinforcement applied to one stimulus of each class will serve to establish the dichotomy, yielding the correct response for all the stimuli.

The problem in which the stimuli are separated into more than two categories can be analyzed in a similar manner.[4]

Training Program II

Consider the stimuli $S_1, S_2, \cdots, S_K$ and their transforms $S_{K+1}=T(S_1)$, $S_{K+2}=T(S_2)$, $\cdots$, $S_{2K}=T(S_K)$ under some one-to-one transformation T of the retinal points. For example $S_1, \cdots, S_K$ may be in the left half of the field and T a transformation which moves them to the right half. $S_x(x=2K+1)$ is not shown during the training but is a test stimulus to be applied after the perceptron is trained. $S_y=T(S_x)$, $y=2K+2=n$. Let us assume S_x intersects $S_1, \cdots, S_L(L<K)$ to a larger extent than it does the others and hence S_y intersects mostly with $S_{K+1}, \cdots, S_{K+L}$. Specifically (cf. Fig. 5)

$$n_{xj}^{\text{I}}=\begin{cases}(q+s\delta_{xj}) & j>L\\ (q+r) & j\leqq L,\end{cases}$$

$$n_{yj}^{\text{I}}=\begin{cases}q & j\leqq K\\ (q+r) & K+1\leqq j\leqq K+L\\ (q+s\delta_{yj}) & j>K+L.\end{cases}$$

We also assume that no associator is activated by more than μ of the stimuli $S_1, S_2, \cdots S_K$, where $\mu<K/L$.

A stimulus S_i from $(S_1,\cdots,S_K)$ is picked at random and the next stimulus is the transform $T(S_i)$. Then another is picked at random from $(S_1,\cdots,S_K)$ and this is followed by its transform, and so on.

Then

$$P_j=\begin{cases}1/2K, & j\leqq 2K,\\ 0 & j>2K,\end{cases}$$

$$P_{jk}=\begin{cases}1 & j\leqq K, \quad k=K+j,\\ 1/K & j>K, \quad k\leqq K,\\ 0 & \text{otherwise.}\end{cases}$$

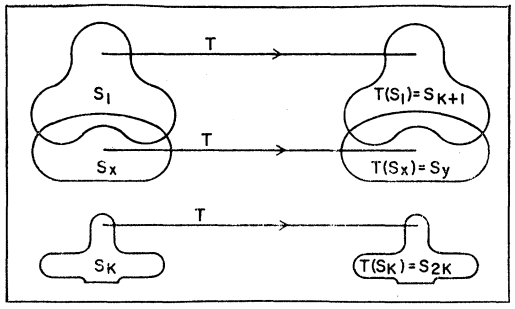

FIG. 5. Stimuli and their transforms on the retina.

From Eq. (14) we obtain

$$\gamma^{(x)}=\frac{\eta}{2K\delta}\left[\sum_{j=1}^{K} n_{xj}{}^{\mathrm{I}}\phi(\beta^{(K+j)}+\gamma^{(K+j)}) + \frac{1}{K}\sum_{j=K+1}^{2K}\sum_{k=1}^{K} n_{xj}{}^{\mathrm{I}}\phi(\beta^{(k)}+\gamma^{(k)})\right] \quad (18)$$

$$\gamma^{(x)}=\frac{\eta}{2K\delta}\left[(q+r)\sum_{j=1}^{L}\phi(\beta^{(K+j)}+\gamma^{(K+j)}) + q\sum_{j=L+1}^{K}\phi(\beta^{(K+j)}+\gamma^{(K+j)}) + q\sum_{j=1}^{K}\phi(\beta^{(j)}+\gamma^{(j)})\right].$$

Hence if

(i) $$\eta(q+r)/2K\delta\geqq\theta$$

then

$$A_\infty{}^{\mathrm{II}}(S_x)\supseteqq A_0{}^{\mathrm{II}}(S_x)+\bigcup_{j\leq L}A_0{}^{\mathrm{II}}[T(S_j)].$$

In words, if the stated inequality holds then, in the terminal state, S_x activates all those elements originally excited either by itself or by any of the transforms $T(S_1), \cdots, T(S_L)$.

(ii) If $\eta q(K+\mu-L)/2K\delta<\theta$, then

$$A_\infty{}^{\mathrm{II}}(S_x)\subseteq A_0{}^{\mathrm{II}}(S_x)+\bigcup_{j\leq L}A_0{}^{\mathrm{II}}[T(S_j)].$$

If both inequalities hold, then

$$A_\infty{}^{\mathrm{II}}(S_x)=A_0{}^{\mathrm{II}}(S_x)+\bigcup_{j\leq L}A_0{}^{\mathrm{II}}[T(S_j)].$$

Consider next $\gamma^{(y)}$. Equation (18) applies (with x replaced by y) and

$$\gamma^{(y)}=\frac{\eta}{2K\delta}\sum_{j=1}^{K}\left[q\phi(\beta^{(K+j)}+\gamma^{(K+j)}) + \frac{Lr+Kq}{K}\phi(\beta^{(j)}+\gamma^{(j)})\right].$$

Thus if

(iii) $$\frac{\eta}{2K\delta}\left(q(K+\mu)+\frac{Lr\mu}{K}\right)<\theta,$$

then $A_\infty{}^{\mathrm{II}}(S_y)=A_0{}^{\mathrm{II}}(S_y)$. If all three inequalities hold then the stimulus S_x generalizes to $T(S_1)$, but the transform $T(S_x)=S_y$ does not generalize to the stimulus S_x. Thus, when the machine has reached its terminal state the stimulus sequence S_x followed by S_y is characterized by a *decreasing* amount of activity, while the sequence S_y followed by S_x would yield an *increasing* pattern of activity. By connections having a time delay to the response units the machine can thus distinguish between a motion to the left (decreasing activity) and a motion to the right (increasing activity).

Of course the above asymmetry between $[S,T(S)]$ and $[T(S),S]$ is due to the training method. If a symmetrical training is applied, the generalization goes both ways.[4]

Necessary and sufficient conditions that all three inequalities hold are easily found, [with $r\geqq 0$, condition (iii) implies (ii)] and are:

(a) $$r>q\frac{K(K+\mu-1)}{(K-L\mu)},$$

and

(b) $$\frac{2K}{(q+r)}\leqq\frac{\eta}{\theta\delta}<\frac{2K^2}{[Kq(K+\mu)+Lr\mu]}.$$

In particular let $L=1$. Then $A_\infty{}^{\mathrm{II}}(S_x)=A_0{}^{\mathrm{II}}(S_x)+A_0{}^{\mathrm{II}}[T(S_1)]$, and $A_\infty{}^{\mathrm{II}}[T(S_x)]=A_0{}^{\mathrm{II}}[T(S_x)]$. Thus due to the intersection between $A_0{}^{\mathrm{II}}(S_y)$ and $A_0{}^{\mathrm{II}}[T(S_1)]$ the test stimulus S_x generalizes to its transform, even though neither one has occurred during the training sequence. This is the effect which was originally predicted for cross-coupled perceptrons,[2] and has since been demonstrated in digital simulation experiments. Specifically, suppose that we have another test stimulus S_z, analogous to S_x, but its chief intersection is with S_2, say also $q+r$. Then, if (a) and (b) are satisfied (with $L=1$ and $\mu<K$),

$$A_\infty{}^{\mathrm{II}}(S_z)=A_0{}^{\mathrm{II}}(S_z)+A_0{}^{\mathrm{II}}[T(S_2)]$$

and

$$A_\infty{}^{\mathrm{II}}[T(S_z)]=A_0{}^{\mathrm{II}}[T(S_z)].$$

Suppose the perceptron has zero initial values on the A^{II} to R connections. Let it be shown S_x and let all active A^{II} to R connections be given a positive increment. Then let it be shown S_z and let all active A^{II} to R connections be given a negative increment. Now if the perceptron is shown $T(S_x)$, (a stimulus it has never seen before) the input to the response unit is the number of associators in

$$A_\infty{}^{\mathrm{II}}[T(S_x)]\bigcap\{A_\infty{}^{\mathrm{II}}[T(S_1)]\bigcup A_\infty{}^{\mathrm{II}}(S_x)\}$$

minus the number of associators in

$$A_\infty{}^{\mathrm{II}}[T(S_x)]\bigcap\{A_\infty{}^{\mathrm{II}}[T(S_z)]\bigcup A_\infty{}^{\mathrm{II}}(S_z)\},$$

which, in general, is positive. Similarly if the machine is shown $T(S_z)$ the response is negative. [Note that

the same result is obtained if the final training is done in the opposite order; that is, show $T(S_x)$ and increment positively; then show $T(S_z)$ and increment negatively; now ask for the responses to S_x and to S_z.]

Thus in the terminal state, after training with the transformation applied to unrelated stimuli, the machine, when taught the response to S_x and to S_z, automatically gives the same response to $T(S_x)$ as to S_x and to $T(S_z)$ as to S_z. It has learned to identify stimuli as equivalent under the transformation T (and similarly T^{-1}).

It has been noted[2] that this ability to identify objects equivalent under a transformation is greatly enhanced by using random blobs, rather than completely random pepper and salt patterns during the initial training period. The reason for this is clear from the fact that the intersection of the pepper and salt patterns with the test letters are all approximately equal, so that r/q is small and condition (a) is violated.

5. CONCLUSIONS

We have shown that for the model described, autonomous learning is possible. In particular, from the analysis, the following types of performance are possible.

(a) The perceptron is shown a random sequence of letters of the alphabet, each letter occurring in various forms, fonts, and positions. The sequence is composed in such a way that a given letter "A," is more likely to be followed by another form or position of the same letter, A, than by a different letter. Ultimately, the perceptron will have seen a number of "runs" of each letter of the alphabet, each such run consisting of a sample of possible positions and variations. At the end the machine should assign a distinctive response to any letter presented; one response for "A" 's another for "B" 's etc. Of course, the particular assignment of responses cannot be specified in advance, since at no time does the experimenter give the machine any instructions based on his knowledge of what the letters are; he merely shows it one letter at a time, distorting and transforming it. It is not the topological similarity of the "A" 's with each other, nor the point-set overlap that is crucial here, but rather the fact that the "A" 's occur contiguously in time. Thus any set of objects that occur contiguously in time can be classified separately from any other sets whose members have the same property.

(b) The machine is shown blobs which jump from the left to the right half of the retina. It is then shown an "A" and a "B" on the right and taught to give the response R_1 to "A" and R_2 to "B". Then, shown an "A" on the left, it gives response R_1, and, for "B" on the left, R_2. It can also be designed so that on being shown any object moving to the right it gives one response; for all objects moving to the left the opposite.

6. CROSS COUPLED SYSTEMS

The more general cross-coupled systems of Fig. 2 of the previous paper[1] present several additional complications.

(a) Closed-loop reverberations are possible. Activity can go on indefinitely, independent of the stimuli presented. The question of whether these reverberations die out, stabilize, or spread to activate all units is crucial to the design.

(b) The sets $A^I(S_i)$ are no longer constant in time. This complicates the analysis. Moreover they depend on the *sequence* of stimuli preceding S_i rather than on only S_i itself. However with suitable modifications, an analysis analogous to that given here has been carried through and equations analogous to (12) obtained.[3] The application of these results to particular problems is being studied.

Cross coupled systems with feedback from the R to A units are now being studied both analytically and by simulation, with the belief that such systems or related models might be capable of dealing with the problems of figure-ground discrimination, relations of figures in complex visual fields, selective attention and recall, and temporal pattern recognition.[3]