ANALYSIS OF PERCEPTRONS

H. D. Block
Cornell University, Ithaca, New York

Summary
Perceptrons are self-organizing or adaptive systems proposed by Frank Rosenblatt as greatly simplified models for biological brains. The main objective is to begin to explain how the brain performs its functions in terms of its structural components. Consequently the methodology consists largely of investigating the behaviour of neural networks which, except for oversimplification, are not unreasonable models of the brain structure, and searching for non-trivial psychological behaviour. This is in contrast with the customary engineering approach of first deciding what function is to be performed and then designing a system to perform the desired function.

A perceptron is a network consisting of ideal neurons, similar to those of McCulloch and Pitts, connected together more or less at random, subject to certain organizational constraints and laws of growth. In this paper we give a brief introduction to the subject and describe some of the results of the mathematical analysis of several such systems; namely 'simple' perceptrons and 'four layer series-coupled' perceptrons.

Introduction
In 1956 Frank Rosenblatt proposed an adaptive or self-organizing system which he called a perceptron. The principal purpose was to provide a model which would begin to explain how the brain performs its functions in terms of its structure. It was intended that the system would be essentially consistent with the known facts of neurophysiology, except of course in the direction of oversimplification. The model was defined with sufficient precision to offer the possibility of exact analysis and prediction of its behaviour. It was hoped that this behaviour would exhibit some aspects of perception. This hope was soon justified, and the perceptron attracted interest as a pattern recognizing device. Although some such applications are being studied, the main orientation of the current research program at Cornell University is toward the brain function-structure problem. Consequently, the methodology of this research differs from that of an engineering program directed toward the development of a pattern-recognizing device, in the following way. The engineer designing a pattern-recognizer would start with the class of patterns to be recognized. He would then construct a system which would, when presented with each of the patterns, produce the desired response. In brief, the engineer starts with the function he wishes the machine to perform and then creates the structure that will perform that function. Rosenblatt's approach is somewhat the opposite. He starts with a system consisting of neurons, similar to those of McCulloch and Pitts, connected together more or less at random subject to certain organizational constraints and laws of growth. After the system has been specified it is then studied to determine how it will function when presented with stimuli. Consequently there exists a difference in viewpoint and methodology. We emphasize this point here, since it explains the formulation and approach employed in the problems to be described below.

[Admittedly, a system which exhibits no interesting behaviour will be discarded, so that a certain amount of "design" is involved in the selection of the systems to be investigated. Similarly, the engineer will systematically check his synthesis by analysis. Thus the contrast should not be pushed too far; it is really a question of emphasis.] A somewhat more detailed exposition of the background of perceptron theory can be found in Reference 4, and a rather thorough discussion in Rosenblatt's comprehensive survey report.9
Perceptrons

A perceptron is a system of the general type indicated in Figure 1. A stimulus activates a subset of the sensory elements. The activated sensory elements send impulses, with various time delays, to the associators. Some of the impulses are positive (excitatory) and some are negative (inhibitory). If the algebraic sum of the impulses arriving at an associator in a suitable time interval exceeds a certain threshold (which need not be the same for all associators) that associator becomes activated and sends out impulses, as indicated by the arrows, to other associators and/or to the response units. The amount of impulse carried by a connection (the 'value' of a connection) varies in accordance with a 'reinforcement rule'; for example a connection whose tail and head are sequentially active is reinforced so that the impulse transmitted by this connection will tend to be increased. This mechanism furnishes the 'memory' of a perceptron. After being activated a unit might suffer a "refractory period," during which it cannot again be activated. Similarly, the response units have an activation threshold and connections with associators and/or each other.

Parameters which must be specified to define the perceptron of Figure 1 are: The number of sensory elements, the number (or probability distribution) of excitatory and inhibitory connections and the geometrical constraints on them, the number of associators and the number of response units; also the thresholds, refractory periods (if any), summation intervals and transmission times. The reinforcement rule and the initial values of the connections must be specified.

For studying the behaviour of such a perceptron it is necessary also to specify the set of stimulus patterns; the order and times of their presentation and the observations to be made on the responses.

The specification of the above items may be given in terms of probability distributions.

These systems are defined in more precise terms in Reference 5. In the interest of brevity we shall not discuss here the consistency of the above model with the biological constraints. The reader interested in this question will find a thorough discussion in Reference 5.

The behaviour of such systems has been the object of considerable study by mathematical analysis, simulation on digital computers and by experiments with a simple hardware perceptron. The results to date are presented in Reference 5. A system of the generality described above presents formidable problems of analysis. For only a few special types is the analysis more or less complete. We describe some typical results in the next two sections.

Simple Perceptron

A simple perceptron is indicated in Figure 2. Let us denote typical sensory units by \( s_\sigma \), typical associators by \( a_\mu \), and typical stimuli by \( S_1 \). Let us represent the connection between \( s_\sigma \) and \( a_\mu \) by the real number \( C_{\sigma \mu} \); in particular the \( C_{\sigma \mu} \) might be random numbers having the possible values \(+1, -1, 0\). When the stimulus \( S_1 \) is applied to the retina, the signal

\[
\phi^{(1)} (\mu) = \sum_{s_\sigma \in S_1} s_\sigma \ C_{\sigma \mu}
\]

is transmitted instantly to the associator \( a_\mu \). If

\[
\phi^{(1)} (\mu) > \theta , \text{ where } \theta \text{ is an arbitrary but fixed real number, the associator } a_\mu \text{ is said to be active and instantly transmits a signal } v_\mu \text{ to the response unit. An inactive associator transmits no signal. The total signal arriving at the response unit is}
\]

\[
u^{(1)} = \sum_\mu v_\mu , \text{ where } \sum_\mu \text{ is taken over the associator units activated by } S_1. \text{ If } u^{(1)} > \Theta , \text{ where } \Theta \text{ is an
arbitrary but fixed non-negative number, the response output is +1. If
\[ u^{(1)} < -\theta \] the response output is -1. If \[ |u^{(1)}| \leq \theta \] the response output is 0.

Let us assign each stimulus \( S_i \) \( (i=1,2,\ldots,n) \) to one of two classes which we denote by +1 and -1. Say stimulus \( S_i \) is assigned to class \( \rho_i \) where \( \rho_i \) is +1 or -1. This dichotomization is then represented by \( \rho=(\rho_1, \rho_2,\ldots,\rho_n) \). The response of the perceptron to stimulus \( S_i \) is said to be correct if, when \( S_i \) is presented, the sign of the output is \( \rho_i \). We say that a solution exists to this discrimination problem for this perceptron if there exist numbers \( \gamma \) such that if \( v_{\mu}=\gamma \) then the perceptron will give the correct response to all the stimuli.

The 'error correction' reinforcement procedure is as follows. A stimulus \( S_i \) is shown, and the perceptron gives a response. If this response is correct then no reinforcement is made. If the response is incorrect then the \( v_\mu \) for active associators \( a_\mu \) is incremented by \( \rho_i \). The inactive associators are left alone. The initial values \( v_\mu \) are arbitrary. The stimuli are presented in any order \( S_1, S_2,\ldots \) with repetitions allowable. 1, 2

**Theorem.** Given a perceptron of the type described above and a dichotomy for which a solution exists, then there is a constant \( M \) such that, under the 'error correction' reinforcement procedure described above, the response of the perceptron will be incorrect at most \( M \) times. In particular if each stimulus recurs infinitely often then the perceptron, after a finite number of stimulus presentations, will thereafter identify all stimuli correctly.

In words: If a solution exists the perceptron will learn the dichotomy in a finite number of steps.

For proof, generalizations and analysis of other reinforcement rules see Reference 4 and 5. Although a few questions regarding the simple perceptron remain unanswered, the theory has reached the point where performance of many such systems can be closely predicted and, conversely, values of the parameters can be specified which will insure successful performance of prescribed discrimination tasks. These systems have a limited ability to generalize. Only in exceptional cases do they exhibit meaningful spontaneous classification of patterns. 5

**Four Layer Series-Coupled Perceptrons**

The simple perceptron of Figure 2 generalizes on the basis of overlap of the stimulus patterns on the sensory field. The perceptron of Figure 3 generalizes rather on the basis of temporal contiguity of the patterns.

The values of the \( S \) to \( A_I \) connections do not change with time. The \( A_I \) units are in one-to-one correspondence with the \( A_I \) units and have threshold \( \theta \). An active \( A_I \) unit \( a_I^{(\mu)} \) delivers a fixed signal of \( \theta \) to its corresponding \( A_I^{(\mu)} \) unit \( a_I^{(\mu)} \) and also a time dependent signal \( v_{\mu}^{(\mu)} \) to \( A_I^{(\mu)} \) \( (\nu=1,2,\ldots,N_a) \), where \( N_a \) is the number of \( A \) units. An inactive unit puts out no signal. The values \( v_{\mu}^{(\mu)} \) are initially zero and change with time as follows. Stimuli are presented at times \( 0, \Delta t, 2\Delta t, 3\Delta t, \ldots \). If \( a_I^{(\mu)} \) is active at time \( t \) and \( a_I^{(\nu)} \) is active at time \( t+\Delta t \) then \( v_{\mu}^{(\nu)} \) receives an increment \( (7.\Delta t) \); otherwise it does not receive this increment. At the same time each \( v_{\mu}^{(\mu)} \) is decremented by \( (7.\Delta t) v_{\mu}^{(\mu)} \). These two effects represent a facilitation of used pathways and a decay, respectively. The \( A_I^{(\mu)} \) to \( R \) connections have values which may be varied according to one of the standard rules of reinforcement as in the simple perceptron of Figure 2. For convenience we take these initial values to be zero, so there is no activity in the \( R \) units until the experimenter intervenes. Since we are interested in the self-organization of this system in the presence of an organized sequence of stimuli, this intervention of the experimenter will not occur, as will be seen below, until the experiment is almost over. There is no time delay of transmission of signals through the system.

The analysis of this system is...
given in detail in References 5 and 6. Here we shall describe some of the results of that analysis.

Let \( r^{(i)}(t) = \sum_{\mu} v_{\mu}(t) \), where \( \sum_{\mu} \) is taken over those \( \mu \) such that the associator \( a^{(i)}_{\mu} \) is activated by \( S_{i} \). That is, \( r^{(i)}(t) \) is the input signal from \( A^{I} \) units \( a^{I}_{\mu}(\mu \neq \nu) \) arriving at the \( A^{II} \) unit \( a^{II}_{\nu} \), at time \( t \). Let \( p^{(i)}_{\nu} \) be the input signal to \( a^{II}_{\nu} \) from \( a^{I}_{\nu} \) when stimulus \( S_{i} \) is presented. Let the probability of occurrence of stimulus \( S_{j} \) be \( P_{j} \) and the probability of transition from \( S_{i} \) to \( S_{j} \) be \( P_{ij} \). We assume these to be stationary. We also assume \( \Delta t < 1 \). Now it is shown in Reference 6 that for \( t \) sufficiently large the system response reaches a steady state. In this state the \( r^{(i)}_{\nu}(t) \) are the unique minimal solutions of the equations:

\[
\sum_{\nu} r^{(i)}_{\nu}(t) = \sum_{j} \sum_{k} n^{I}_{ij} P_{j} P_{jk} \phi(a^{I}_{\nu} - r^{(i)}_{\nu})
\]

where \( n^{I}_{ij} \) is the number of \( A^{I} \) associators activated by both \( S_{i} \) and \( S_{j} \); and \( \phi(x) = 1 \) for \( x > 0 \), \( \phi(x) = 0 \) for \( x < 0 \).

From this basic result conclusions can be drawn about the terminal state of many such systems. We illustrate with two training programs on such perceptrons.

In the first training program the stimuli \( S_{1}, S_{2}, \ldots, S_{n} \) are divided into two classes: \( [S_{1}, S_{2}, \ldots, S_{k}] \) is class \( X \), while \( [S_{k+1}, S_{k+2}, \ldots, S_{n}] \) is class \( Y \). There is assumed to be no appreciable difference in the retinal overlaps: \( n^{I}_{ij} = q + s \delta_{ij} \), where \( s > 0, q > 0 \). Thus the diagonal elements of the \( n^{I}_{ij} \) 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. The stimuli are presented at random, subject to the constraint that the probability of transition to a stimulus of the same class is \( p \), nearly one, while the probability of transition to a stimulus of the opposite class is \( (1 - p) \), nearly zero. Inside a class all members are equally likely.

Let \( A^{II}_{\nu}(S_{i}) \) be the subset of \( A^{II} \) units activated by stimulus \( S_{i} \) in the initial state and \( A^{II}_{\nu}(S_{j}) \) the subset activated in the terminal state. If the parameters satisfy the inequalities

\[
\frac{K^{2}}{s^{2} + qK} < \frac{q}{2q} \frac{K}{s^{2}} < \frac{1}{s(1-p) + qK}
\]

then if \( 1 \leq i \leq K \)

\[
A^{II}_{\nu}(S_{i}) = \bigcup_{1 \leq j \leq K} A^{II}_{\nu}(S_{j})
\]

while if \( K < i \leq n \) then

\[
A^{II}_{\nu}(S_{i}) = \bigcup_{K < j \leq n} A^{II}_{\nu}(S_{j})
\]

That is, in the terminal state all stimuli of class \( X \) activate precisely the same set of \( A^{II} \) units, namely the set consisting of all those units initially activated by any stimulus of class \( X \). Similarly each stimulus of class \( Y \) activates all those units initially activated by any stimulus of class \( Y \) and only these.

Thus the machine has dichotomized the classes, its codification being in terms of this intrinsic code on the \( A^{II} \) units. The experimenter has had no part in this experiment up to this point, except for his presenting the stimuli with the specified transition probabilities. If he now intervenes to give a single corrective reinforcement to the \( R \) units for one stimulus of each class the perceptron will then yield the correct response for all the stimuli. A similar analysis holds for more than two classes of stimuli.

The above result can be restated in the following, possibly more descriptive, terms. Here we take non-zero values on the \( A^{II} \) to \( R \) connections. 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 and 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.

For the second training program consider the stimuli $S_1$, $S_2$, ..., $S_K$ and their transforms $S_{K+1} = T(S_1)$, $S_{K+2} = T(S_2)$, ..., $S_{2K} = T(S_K)$ under some one-to-one transformation $T$ of the retinal points. For example $S_1$, ..., $S_K$ 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. $S_y = T(S_x)$, ($y=2K+2$). Let us assume $S_x$ intersects $S_1$, ..., $S_L$ ($L < K$), to a larger extent than it does the others. Specifically (cf. Figure 4)

$$
\begin{align*}
I_{xj} &= \left\{ (q+s_j) \right\} \\
I_{yj} &= \left\{ (q+r) \right\}
\end{align*}
$$

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

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

If the parameters satisfy the inequalities

$$
\frac{q(K+1)}{K} < \frac{2K\delta_0}{q+r}
$$

then it is shown in Reference 6 that

$$
A_{\infty}(S_x) = A_{\infty}(S_y) \bigcup_{1 \leq L} A_0(T(S_j))
$$

and

$$
A_{\infty}(S_y) = A_{\infty}(S_y)
$$

From this it follows first that 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).

A more important result is this: 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, and has since been demonstrated in digital simulation experiments.

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 another stimulus $S_y$, automatically gives the same response to $T(S_x)$ as it does to $S_x$ and the same response to $T(S_y)$ as to $S_y$.

This result can be worded in perhaps more descriptive language as follows. The first stimulus is $S_1$, a random blob.
located in the left half of the sensory field. The second stimulus is T(S₁), the same blob moved rigidly to the right half of the field by a translation, T, of a fixed number of retinal spaces. The third stimulus is another random blob S₂ in the left half of the field; the fourth is T(S₂), its transform into the right half. The sequence is continued with random blobs in the left half of the field immediately followed by their transform to the right half, under the fixed transformation T. Then the system ultimately reaches a steady state in which, if the parameters are suitably chosen, the following behaviour will be exhibited:

In the terminal state, the machine is shown an A on the left and taught (for example by the error correction reinforcement procedure, described in connection with Figure 2, applied to the A₁₁ to R connections) to give the response Rₐ. Then it is shown a B on the left and taught to give the response R₉. Now, when it is shown an A on the right (a stimulus it has never seen before), it gives the response Rₐ. When shown a B on the right (a stimulus it has not seen before), it gives the response R₉. Thus it has learned to 'identify as equivalent' (give the same response to) two patterns which are equivalent under the transformation T (and similarly T⁻¹); it has learned the transformation. The selection of T as a horizontal translation was for purposes of illustration; the result remains true for any one-to-one transformation. The use of blobs in the training sequence, rather than completely random pepper and salt patterns, is essential however, since with the pepper and salt patterns the same effect would be much more difficult to produce. This was already observed earlier.

Other training procedures, in particular the symmetrical one in which either the stimulus or its transform can be the initial stimulus of the pair, are discussed in Reference 5.

Cross-Coupled Systems

The analysis of the general perceptron of Figure 1 presents several additional complications.

a) Closed loop reverberations are possible. Activity can go on independently of the stimuli presented and even if all stimuli are removed. The question of whether these reverberations die out, stabilize, or spread to activate all units, is crucial.

b) The set of Aᵢ units activated by a given stimulus is no longer constant in time. This complicates the analysis. Moreover, they depend on the sequence of stimuli preceding the current one, rather than on the current stimulus alone. However, with suitable modifications, an analysis analogous to that used on the four layer system has been carried through. The application of these results to specific situations is being studied further.

One further result of the above analysis which might be of interest to the engineer, is that the performance of these systems is relatively insensitive to malfunction or extirpation of a considerable fraction of the components. This means that as the system gets larger, the common (and often impossible) requirement for increased reliability of all components, is not encountered here.

References


Organization of a Perceptron

Figure 1

Simple Perceptron

Figure 2

From the collection of the Computer History Museum (www.computerhistory.org)
Organization of Four Layer Series-Coupled Perceptron

Figure 3

Stimulus Patterns on Sensory Field

Figure 4