INFORMATION AND CONTROL 17, 501–522 (1970)

A Review of "Perceptrons: An Introduction to Computational Geometry"

by Marvin Minsky and Seymour Papert.

The M.I.T. Press, Cambridge, Mass., 1969.
112 pages. Price: Hardcover $12.00; Paperback $4.95.

H. D. BLOCK

Department of Theoretical and Applied Mechanics, Cornell University, Ithaca, New York 14850

1. INTRODUCTION

The purpose of this book is to present a mathematical theory of the class of machines known as Perceptrons. The theory is carefully formulated and focuses on the theoretical capabilities and limitations of these machines.

It is a remarkable book. Not only do the authors formulate a new and fundamental conceptual framework, but they also fill in the details using strikingly ingenious mathematical techniques. They ask some novel questions and find some difficult answers. The most striking of these will be presented in Section 2.

The authors address the book to three classes of readers:

(1) Computer scientists, specializing in pattern recognition, learning machines, and threshold logic;

(2) Abstract mathematicians interested in the début of *Computational Geometry*;

(3) Those interested in a general theory of computation leading to decisions based on the weight of partial evidence. The authors hope that this class includes psychologists and biologists.

In Section 6 I shall give my estimate of the value of the book to each of these groups.

The conversational style and the childlike freehand sketches might mislead the casual reader into believing that this book makes light reading. For example, the review in *The American Mathematical Monthly* (1969) states that the prospective reader "requires little mathematics beyond the high school level." This, as we shall see, is somewhat sanguine. Another extreme opinion is due to Allen Newell (1969) who begins his review in *Science* with the statement "This is a great book.", and then proceeds to present in detail his reasons for this extraordinary judgement. My evaluation is somewhat more moderate. It is presented in Sections 5 and 6.

2. Highlights of the New Theory

In the interest of brevity we will not provide any of the proofs or introduce those concepts whose rôle is primarily technical. The reader who tries to provide his own proofs will, I believe, soon come to appreciate the mathematical virtuosity of the authors. We start with some definitions.

Let the plane rectangular region $\{0 \leqslant x \leqslant M,\ 0 \leqslant y \leqslant N\}$ be divided into squares $s_{ij} = \{i \leqslant x \leqslant i + 1,\ j \leqslant y \leqslant j + 1;\ i = 0, 1,..., M - 1, j = 0, 1,..., N - 1\}$. The set of all these squares is called[1] the *retina*, R. Thus R is a finite set of squares $\{s_{ij}\}$. A subset X of R is called an *image* (or *pattern*, or *figure*) in R. The number of squares (sometimes called *points*) in X is denoted by $|X|$.

A *predicate* P is a function of images, taking on the value "true" or "false." That is, $P(X)$ is a statement about the subset X which is either true or false. The authors introduce the helpful "partial bracket" notation $\ulcorner\ \urcorner$, where for any statement P, $\ulcorner P \urcorner = 1$ if P is true, $\ulcorner P \urcorner = 0$ if P is false. We now identify "1" with "true," and "0" with "false." Thus if P is a predicate so is $\ulcorner P \urcorner$. This "partial bracket" is often convenient for typographically displaying predicates.

The *support* of a predicate P is the intersection of all subsets S of R which satisfy the condition: $P(X) = P(X \cap S)$ for all subsets X of R. In intuitive terms, the support of P is the set of squares that P "depends on." The support of P is denoted by $S(P)$. Thus $|S(P)|$ is the number of squares in the support of P.

For a given family of predicates $\mathbf{\Phi} = \{\phi\}$, we say that a predicate ψ is a *linear threshold function* with respect to $\mathbf{\Phi}$ if there is a number θ and, for each ϕ in $\mathbf{\Phi}$, a number $\alpha(\phi)$, such that ψ has the representation

$$\psi(X) = \ulcorner \sum_{\phi \in \Phi} \alpha(\phi) \cdot \phi(X) > \theta \urcorner$$

[1] For simplicity and clarity of exposition in this review, I have slightly altered Minsky and Papert's definition, which initially takes R to be an arbitrary finite set of points. The special setting we use here will be more suitable for our exposition and involves no real loss in generality.

for all subsets X of R. The class of all predicates which are linear threshold functions with respect to $\mathbf{\Phi}$ is denoted by $L(\mathbf{\Phi})$. A *perceptron* is defined, by the authors, as a device capable of computing all the predicates in $L(\mathbf{\Phi})$ for some family $\mathbf{\Phi}$. In Sections 5 and 6 we discuss the relation of perceptrons so defined to the *Perceptrons* introduced by F. Rosenblatt.

The *order* of a predicate ψ is the least number k for which there exists a set of predicates $\mathbf{\Phi} = \{\phi\}$ such that $\psi \in L(\mathbf{\Phi})$ and $| S(\phi)| \leqslant k$ for all ϕ in $\mathbf{\Phi}$. (Note that the number of predicates in $\mathbf{\Phi}$ does not enter into the definition of *order*. This will be discussed in Section 6.)

We can now state two theorems.

THEOREM. *The "parity predicate"* $\ulcorner | X |$ *is odd*$\urcorner$ *has order* $| R |$.

THEOREM. *Let* $A_1, A_2, \ldots, A_m$ *be disjoint subsets of* R, *each containing* $4m^2$ *points. Then the order of the "one in a box predicate"* $\ulcorner \prod_i | X \cap A_i | > 0 \urcorner$ *is at least* m.

We need a few more definitions. A *transformation* g on R is a one-to-one mapping of R onto itself. A transformation g induces, in a natural way, the set function $g(X) = \bigcup_{s \in X} g(s)$. With the product of transformations defined in the usual way ($gh(s) = g(h(s))$), a set of transformations closed under product and inverse is a *group* of transformations. Translations are handled by the agreement that any part of a figure that is carried over the right boundary of the retina is brought back in at the corresponding place at the left, and vice versa. Similarly, parts carried over the top re-enter at the bottom and vice versa. That is, after a figure in R has been translated, the X and Y coordinates of its points are reduced to their residues modulo M and N, respectively. Thus, the retina is wrapped onto a torus, where the translations become a finite transformation group. If the figures and the translations are "small" relative to the retina, it is expected that the "local structure" of the toroidal retina will be in close agreement with that of the plane retina.

Given a group $G = \{g\}$ of transformations, a predicate ψ is *invariant* under G if $\psi(g(X)) = \psi(X)$ for every g in G and every subset X of R. A group G is *transitive* if for every pair (s, t) in $R \times R$ there is a g in G such that $g(s) = t$.

THEOREM. *The "counting predicates"* $\ulcorner | X | < M \urcorner$ *and* $\ulcorner | X | > M \urcorner$ *(where* M *is some number) are of order one. These are the only first-order predicates invariant under a transitive group.*

Most of the predicates we shall meet, e.g., $\ulcorner X$ is a convex set$\urcorner$, $\ulcorner X$ is a connected set$\urcorner$, $\ulcorner X$ has a hole$\urcorner$, etc., can be described without explicit reference to the size and shape of the retina R. If a predicate is of order $\leqslant k$, regardless of the size and shape of the retina, then we say that the predicate is of *finite order*. (To be perfectly rigorous we should say that we have a class of retinas $\mathbf{R} = \{R\}$ and a class of predicates, one for each R. If the orders of all these predicates are bounded, then the *class* is of finite order. Since no confusion is likely to arise, the simpler locution is used.) For example, it follows from the theorems cited above that the parity predicate and the one-in-a-box predicate are not of finite order. The counting predicate is of finite order, in fact of order one.

The logical operators $\oplus$, $\equiv$, $\wedge$, $\vee$ are defined by

$$\begin{aligned}
\ulcorner P \oplus Q \urcorner &= \ulcorner P \urcorner (1 - \ulcorner Q \urcorner) + (1 - \ulcorner P \urcorner) \ulcorner Q \urcorner, \\
\ulcorner P \equiv Q \urcorner &= \ulcorner P \urcorner \ulcorner Q \urcorner + (1 - \ulcorner P \urcorner)(1 - \ulcorner Q \urcorner), \\
\ulcorner P \wedge Q \urcorner &= \ulcorner P \urcorner \ulcorner Q \urcorner, \\
\ulcorner P \vee Q \urcorner &= \ulcorner P \urcorner + \ulcorner Q \urcorner - \ulcorner P \urcorner \ulcorner Q \urcorner.
\end{aligned}$$

THEOREM. *If ψ_1 is of order r_1 and ψ_2 is of order r_2 then $\ulcorner \psi_1 \oplus \psi_2 \urcorner$ and $\ulcorner \psi_1 \equiv \psi_2 \urcorner$ are of order $\leqslant r_1 + r_2$.*

However, there is also the

"AND-OR" THEOREM. *There exist predicates ψ_1 and ψ_2, each of order one, such that $\ulcorner \psi_1 \wedge \psi_2 \urcorner$ and $\ulcorner \psi_1 \vee \psi_2 \urcorner$ are not of finite order.*

DEFINITION. Two squares of R are *adjacent* if they have a common edge. (Note that corner contacts are not counted.) An image X is *connected* if for every two squares s_1, s_K in X there is a connected path in X joining them (i.e., a sequence of squares: $s_1, s_2, \ldots, s_K$, where s_k, s_{k+1} are adjacent and in X).

THEOREM. *The predicate $\ulcorner X$ is connected$\urcorner$ is not of finite order. In fact it is of order $\geqslant C\sqrt{|R|}$.*

DEFINITION. A *component* of a figure is a maximal connected subset of the figure. A *hole* of a figure is a component of the complement of the figure; here however we allow corner connections[2]. Note that it is assumed that the figure is surrounded by an "outside" that does not count as a hole.

[2] It seems that this difficulty could be avoided by using hexagons instead of squares for the tessellation of the plane. Some of the theory would have to be checked to verify that no difficulties are thereby introduced.

Let the *Euler number* $E(X)$ equal the number of components of X minus the number of holes of X.

THEOREM. *Let M be any number. Then the order of* $\ulcorner E(X) < M \urcorner \leqslant 4$; *the order of* $\ulcorner E(X) = M \urcorner \leqslant 8$.

The authors define a predicate as *topologically invariant* if it is unchanged when the figure is distorted without changing connectedness or the inside–outside relations of its parts. (See Section 4 for some remarks about this definition.)

THEOREM. *Except for trivial predicates, such as* $\ulcorner$*X is (non-) empty*$\urcorner$, *the only topologically invariant predicates of finite order are functions of* $E(X)$.

DEFINITION. For a given figure X and any ordered pair of squares (s, t) in $X \times X$ let the *difference vector* $v(s, t)$ be the vector from the center of square s to the center of t. For each such vector $v(s, t) = \mathbf{v}$ let $n_{\mathbf{v}}(X)$ be the number of ordered pairs of squares having the difference vector $\mathbf{v}$.

THEOREM. *Any translation-invariant predicate of order two is of the form* $\ulcorner \sum_{\mathbf{v}} \alpha_{\mathbf{v}} n_{\mathbf{v}}(X) > \theta \urcorner$; *hence it is a function only of the "difference-vector spectrum"* $n_{\mathbf{v}}(X)$.

It follows, e.g., that any figure and the figure obtained by rotating it 180° (which have the same difference-vector spectrum) cannot be distinguished from each other by a translation-invariant second-order predicate.

THEOREM. *The following predicates are of order three*: $\ulcorner$*X is a solid rectangle*$\urcorner$; $\ulcorner$*X is a hollow rectangle*$\urcorner$; $\ulcorner$*X is a solid square*$\urcorner$; $\ulcorner$*X is a hollow square*$\urcorner$. *On the other hand, the "hollow square in context" predicate* $\ulcorner$*one component of X is a hollow square*$\urcorner$ *is not of finite order.*

THEOREM. *Let X_1 and X_2 be two figures which are not translationally equivalent. Then there exists a translationally-invariant predicate ψ of order $\leqslant 3$ which separates them, i.e., such that* $\psi(X_1) = 0 < \psi(X_2)$.

THEOREM. *Consider the "infinite linear" retina, consisting of the squares* $\{0 \leqslant y \leqslant 1,\ j \leqslant x \leqslant j+1;\ j = ..., -1, 0, 1, ...\}$. *Consider the class of finite figures* $\{X\}$. *The predicate* $\ulcorner$*X is symmetrical about some point*$\urcorner$ *is of order* $\leqslant 4$. *For any given figure* X_0, *the predicate* $\ulcorner$*X is a translate of* $X_0 \urcorner$ *is of order* $\leqslant 2$.

This theorem, concerning the "infinite linear" retina, might be compared with the third preceding theorem, which states that any second-order, translationally invariant predicate on a toroidal retina is a function only of the difference-vector spectrum. Thus, since the two patterns  and  have the same difference-vector spectrum, they are not distinguishable by such a predicate. But according to the last sentence of the present theorem they can be separated by a second-order, translationally invariant predicate on the infinite line.

THEOREM. *Consider the "doubly infinite" linear retina consisting of the squares* $\{i \leqslant y \leqslant i+1,\ j \leqslant x \leqslant j+1;\ i = 0, 1;\ j = ..., -1, 0, 1, ...\}$. *Let* X *be a finite figure composed of a part* $U(X)$ *in the upper row of squares and a part* $L(X)$ *in the lower row. The predicate* $\ulcorner L(X)$ is a translate of $U(X)\urcorner$ *is of order* $\leqslant 5$.

A somewhat similar theory, but with some differences, is obtained if one replaces the *order* limitation of a predicate by a *diameter* limitation. That is, instead of the *number of points* in $S(\phi)$ in the definition of *order*, we use the *diameter of the set* $S(\phi)$. The authors find that diameter-limited perceptrons *can* recognize the predicates $\ulcorner X = R\urcorner$, $\ulcorner | X | > M\urcorner$, $\ulcorner X$ is a triangle$\urcorner$, $\ulcorner X$ is a rectangle$\urcorner$, $\ulcorner X$ is a particular figure $X_0\urcorner$, but *cannot* recognize the predicate $\ulcorner M_1 < | X | < M_2\urcorner$ (which is second-order), or the infinite-order predicate $\ulcorner X$ is connected$\urcorner$; also, that the only nontrivial topological properties that can be recognized by a diameter-limited perceptron are the Eulerian predicates $\ulcorner E(X) > M\urcorner$, $\ulcorner E(X) < M\urcorner$.

3. Summary of the Book

The book begins with Chapter 0, Introduction. Here Minsky and Papert present a clear, crisp, and masterful summary of the book as a whole. In addition they offer, in their ebullient style, their opinions and sentiments on the past, present, and future of Perceptrons.

Chapters 1–10 develop the new theory, the highlights of which we gave in Section 2, above.

Chapter 11 treats learning machines and gives several proofs of the well-known convergence theorem for perceptrons with error correction. Considerably more novel is the proof of the *Boundedness Theorem* in the nonseparable case. This theorem can be stated as follows:

Let V *be a vector space with an inner product. Let* $\mathbf{F} = \{\phi\}$ *be a finite set of* *vectors in V. Let* v_1 *be an arbitrary vector in V and define recursively* $v_{i+1} = v_i + \phi_i$, *where* ϕ_i *is any vector in* **F** *such that* $\phi_i \cdot v_i \leqslant 0$. *Then the vectors* v_i *stay bounded; in fact there is a number M depending on the set* **F**, *but not on* v_1, *such that* $\| v_i \| \leqslant \| v_1 \| + M$ *for* $i = 1, 2, \ldots$.

This theorem was conjectured by Nils Nilsson and, independently, by Terry Beyer. A proof was offered by Bradley Efron (1964), but it was rather difficult to follow and was never published in a standard journal. The literature has therefore lacked a clear and rigorous proof. Minsky and Papert's penetrating analysis goes a long way towards providing one. (But see the remark in Section 4, below).

Chapter 12 contains a general discussion of linear separation, learning, and heuristics. Also included are estimates of the storage capacity and computing time required by a variety of algorithms, such as maximum likelihood, isodata, template matching, and nearest neighbor. Layered machines, Samuel's checker player, neuronal models, hash coding, and incremental methods are briefly discussed.

In the final chapter, 13, the authors expound their general views on perceptrons and pattern recognition. They also recount the development of their ideas on the subject. The book ends with Bibliographic Notes, in which the authors comment briefly on some of the well-known papers in this field.

4. Detailed Notes for the Prospective Careful Reader of the Book

This section of the review is directed to those who will read the book carefully and is intended to help them over misprints, lacunae, ellipses, etc. Notation: An asterisk following the number of a line denotes that number of lines from the *bottom* of the page.

Page 26, line 14*, (i.e., 14 from bottom) for *that* read *than.*

Page 26, lines 5*, 6*, change *many predicates* to *every predicate.*

Page 31, line 9*, add *except for the mask of the empty set, which has order zero.*

Page 32, line 3*, the inequality should read $0 < M < | R |$, for the result is not true if $M = 0$ or $M = | R |$. The predicate then is, in fact, of order one.

Page 43, line 1, mentions the "group of all rotations about all points in the plane". This is not a group, since the product of two equal but opposite rotations about two distinct points is not a rotation about a point in the plane, but a translation. Adjoining the rotations about the point at infinity (i.e., the translations) will complete the group.

Page 58, line 3, add *except, possibly, the mask of the empty set.*

Page 60, line 2, $\frac{1}{4} | R |^{1/3}$ should be $(| R |/4)^{1/3}$. Furthermore, in the "one-in-a-box" theorem, page 59, line 1*, one can replace $4m^2$ by $(2m - 1)^2$. Also, it is easy to show that the order of ψ is equal to m.

Page 65, lines 8*, 2*, interchange the arguments in $f(\, , \,)$. The proof given applies only for N odd, but an easy proof can be made if N is even.

Page 68, line 3*, 2*, replace P by Q and *it* by P.

Page 71, line 1, change "*by the*" to "*by a topologically invariant*".

Page 74, line 6, after *path* add *in the figure.*

Page 75, lines 13, 14, replace X by Y.

Page 78, line 10, replace $\not\equiv \cdots \in$ by $\equiv \cdots \notin$

Page 78, last line, and p. 79, first line, interchange $\hat{R}$ with R.

Page 80, lines 11, 12, interchange "on R" with "on $\hat{R}$".

Page 80, line 17, for $2n$ read $(2n + 2)$.

Page 81, line 3, replace $(1/12) | R |$ by $(1/12) | R | - 1$.

Page 84, lines 2, 3, it seems that one can replace $5n$ by $3n$, and $2n$ by n (by running the first half about the left end and the second half about the right end).

Page 84, line 6, replace $n = 5$ by $n = 4$.

Let $P(X)$ denote the predicate $\ulcorner X$ is connected or X contains a hole$\urcorner$. Then the assertion on the bottom of page 85 that "$P(X)$ is of finite order," seems to be incorrect. Apparently the authors felt that the assertion followed from page 89, line 10, where it is shown that the predicate $\ulcorner E(X) = M \urcorner$ is of finite order. But these are not the same predicates for any value of M. Even though $(E(X) = 1)$ implies ($P(X)$ is true), the converse does not hold. In fact Theorem 5.9 (page 92), stating that "The only [nontrivial—*HDB*] topologically invariant predicates of finite order, are functions of the Euler number $E(X)$" contradicts the assertion. For we have $E(X) = 2$ if X has two components and no holes, or if X has three components and one hole. But in the first case $P(X)$ is false, while in the second $P(X)$ is true. Thus $P(X)$ is not a function of $E(X)$.

It might seem more natural to define a topological transformation as a *one-to-one mapping* (like the other transformations) which preserves adjacency and, perhaps, the "outsideness"; and to say that a predicate ψ is topologically invariant if it is invariant under topological transformations. Then however Theorem 5.9 (cited just above) would no longer be true, because, e.g., the counting predicate $\ulcorner | X | < M \urcorner$ is of order one and is invariant under any group of one-to-one transformations, but clearly it is not a function of the Euler number. Thus the authors have a reason for choosing their definitions as they do, but it would seem more natural for topological invariance to be less restrictive than invariance under the full permutation group, which allows discontinuities and tearing. This suggests the following question, which is not treated in the book. Which (one-to-one) transformations of a finite toroidal grid preserve adjacency and outsideness? Clearly we have the translations, reflections in horizontal or vertical lines or through a center, products of the preceding, and, for square grids, rotations through multiples of 90°. Are there any others? How can the whole class be characterized?

In the last nine lines of p. 98 the authors discuss the "tolerance difficulties" (caused by the finite mesh size) when one tries to deal with rotations other than multiples of 90° or with affine contractions. However, their definition of convexity "($a \in X$ and $b \in X$) implies (midpoint $[a, b] \in X$)" (p. 103) ignores this difficulty. Moreover, they are using "lattice points" as the points in R, in which case the midpoint of two lattice points is not always another lattice point. One can partially get round this difficulty by using closed squares, as we have, and defining a set X as convex if every line segment joining the centers of two squares in X has its midpoint in X. Another definition is that there exists a convex subset of the plane which meets each square in X and no others; a detailed study can be found in Sklansky (1970) or Montanari (1970). Minsky and Papert again discuss the tolerance difficulties on page 134, lines 5–19, and examine convexity further in Section 9.3.

Page 108, line 9, replace *orders* by *sizes of the supports.*

Page 109, line 7*, replace Φ^2 by $\bigcup_i \Phi_i^2$.

Page 109, line 3*, replace Φ^2 by Φ_i^2.

On page 111 the authors define a predicate *in context,* by

$$\psi_{\text{context}}(X) = \ulcorner\psi(Y) \text{ for some component } Y \text{ of } X\urcorner.$$

(I have added the second Y and deleted the word *connected—HDB.*) This definition is not completely satisfactory. For example, take a disconnected figure, such as the character "i". Let the predicate $\psi^{``i"}(X) = \ulcorner X$ is a translate of "$i$"$\urcorner$. Then $\psi^{``i"}_{\text{context}}$ is always false. Similarly if $\psi^{``w"}(X) = \ulcorner X$ is a translate of "$w$"$\urcorner$ then a figure consisting of "$w$" with one extra square connected to it does not satisfy the predicate $\psi^{``w"}_{\text{context}}$. The authors do admit to some doubts about their definition of a "predicate in context", but these doubts do not inhibit them from using their definition as the basis for some rather strong statements on page 112 (lines 3–11) and page 113 (last 10 lines).

On page 118, line 10*, the authors take $B_j = d + 1$ and then make the parenthetical remark "(We resisted the temptation to write $B_j = (1/2)\, d$.)". I found this remark disconcerting, since I computed $B_j = (1/2)(d + 1)$, for d odd, and $(1/2)\, d$ for d even, so one can use $B_j = (1/2)(d + 1)$ in any case. I still haven't discerned the source of the authors' temptation or their reason for resisting so strongly.

Page 119, lines 15, 16, replace i by j and j by k.

Page 136, last line, replace *path* by *pair*.

The definition on page 137, line 10, "We choose x_{i*} to be the boundary point to one's right when standing on x_i and facing the complement of X," is ambiguous if, as in the figure (top of page 138), the figure has parts that are only one square thick. The dotted lines in Figure 9.1 confused me; they do not seem to agree with the algorithm in the text.

Page 144, line 12, change 2^m to $2^{m-1} + 1$, and modify subsequent estimates.

Page 151, line 22, change $3\alpha_3$ to $3\alpha_2$.

In a book dedicated to mathematically precise exposition, it is something of a shock to meet the statement (page 160, lines 9*–7*): "Probably this means that Theorem 10.4.1 is not strictly true, but we do not think the exceptions are important."

In Chapter 10, Minsky and Papert show that extremely large weights are required for certain predicates. It should be noted that the weights could be reduced if additional layers were used.

Page 166, line 10, replace *vector* by *direction*, since any positive multiple of a solution vector is again a solution vector.

Page 170, line 3, replace *dialectrical* by *dialectical*. In the caption of Figure 11.3 replace A_1 by A_n.

The proof of the Convergence Theorem (pages 164–175) seems excessively labored. The briefest proof, based on one by W. C. Ridgeway, is only a few lines long and may be found in Block and Levin (1970).

Page 175, lines 12–14, insert a minus sign on the right side of the displayed equation.

Page 175, line 17*, the phrase "after a finite number of transfers" is incorrect. In fact it is easy to see that for $c < 1$, a solution is never reached, since *after* each step, $A \cdot \Phi$ is still negative. The algorithm of Kameda (1967), which converges rapidly to a minimal length solution, deserves to be cited here.

The proof of the *Boundedness Theorem* (pages 182–187) is, as has been mentioned above, a welcome addition to the theory, filling a long-standing need. The presentation in the text is somewhat disorganized and unclear. There is one serious flaw, where the authors draw the inference "... so that $\| C \| < M_{n-1}$;" (page 187, line 17), apparently based on the inequality

$\| B_1 + C \| < \| B_1 \| + M_{n-1}$. This hiatus can be repaired, but it is not trivial. The reader will find in Block and Levin (1970) a clearer proof, which is based on the ideas introduced by Minsky and Papert in this section of the book.

In the discussion of the Bayes procedure (page 193, lines 3*–1*, and in Section 12.4.2) there appears to be some confusion regarding the relationship between (a) the hypothesis that the ϕ's are statistically independent, and (b) the linearity of the discriminant function. The following elementary derivation will, I believe, clarify the situation.

Bayesian Decision Procedure, Statistical Independence, and Linear Discriminant Functions

Consider K urns labeled 1, 2,..., K. Each urn contains a collection of patterns $\{X\}_k$. (It may happen that the same pattern occurs in more than one urn.) The contents of each urn are known. One urn is chosen at random but its label is not revealed. From that urn a pattern X is chosen and revealed. From which urn did it come?

We must decide on a number k from the set (1, 2,..., K), knowing that a certain pattern X has been chosen. There are three well-known approaches.

1. The *maximum likelihood method* decides on the k which maximizes $P(X \mid k)$, the probability that pattern X is drawn, under the hypothesis that urn k was selected.

2. The *maximum a posteriori method* decides on the k which maximizes the conditional probability of k, given that X is observed: $P(k \mid X) = [P(X \mid k)\, P(k)/P(X)]$; or, equivalently, the k which maximizes $P(X \mid k)\, P(k)$. In the special case in which the a priori probabilities $P(k)$ are equal, method (2) reduces to method (1)

3. Assume that a loss $\lambda(k \mid j)$ is suffered if we decide on urn k when in fact j was the urn form which X came. If, when we observe X, we decide on urn k, then the *expected loss* is $L(k \mid X) = \Sigma_j\, \lambda(k \mid j)\, P(j \mid X) = \Sigma_j\, \lambda(k \mid j)\, [P(X \mid j)\, P(j)/P(X)]$. The *minimal expected loss method*, or the *Bayes Procedure*, decides on the k which minimizes $L(k \mid X)$ (or, equivalently, $\Sigma_j\, \lambda(k \mid j)\, P(X \mid j)\, P(j)$).

If $\lambda(k \mid j) = 1 - \delta_{kj}$ then method (3) reduces to method (2); also $L(k \mid X)$ is then the probability of an error if k is decided on when X is observed.

In general then, having observed X, we choose k so as to minimize

$$g(k \mid X) = \sum_{j=1}^{K} \lambda(k \mid j)\, P(X \mid j)\, P(j).$$

For each subset S of R let ϕ_S be the predicate $\phi_S(X) = \ulcorner X = S \urcorner$. Then

$$\begin{aligned} g(k \mid X) &= \sum_{j=1}^{K} \lambda(k \mid j)\, P(j) \left(\sum_S \phi_S(X) P(S \mid j) \right) \\ &= \sum_S \alpha_S(k)\, \phi_S(X). \end{aligned}$$

Hence we have a discriminant function *linear* in the ϕ's, with *no assumption of independence*. (Minsky and Papert seem to suggest that linearity requires the assumption of independence.) In typical applications, most of the coefficients $P(S \mid j)$ will be zero. Therefore the last summation may involve considerably fewer than all $2^{|R|}$ possible predicates.

Denoting the retina R by $(s_1, s_2, \ldots, s_{|R|})$, we can represent X in the form $(x_1, x_2, \ldots, x_{|R|})$, where x_i is the predicate $x_i(X) = \ulcorner s_i \in X \urcorner$. Let us denote the corresponding random variables by ξ_i. Then if the ξ_i are independent under each hypothesis j,

$$g(k \mid X) = \sum_{j=1}^{K} \lambda(k \mid j)\, P(j) \prod_{i=1}^{|R|} P(\xi_i = x_i \mid j).$$

If $\lambda(k \mid j) = 1 - \delta_{kj}$ then we can take logarithms and choose k so as to maximize

$$\begin{aligned} h(k \mid X) &= \log P(k) + \sum_i \log P(\xi_i = x_i \mid k) = \log P(k) + \sum_i \{x_i \log P(\xi_i = 1 \mid k) \\ &\quad + (1 - x_i) \log P(\xi_i = 0 \mid k)\} \\ &= \sum_i \alpha_i(k) x_i + \theta(k), \end{aligned}$$

which is a linear discriminant function in the one-point predicates x_i.

The analysis in Minsky and Papert follows from the above, if we consider their ϕ's as if they were x_i's in a second level. The assumption of independence can often be validated in practice by suitably randomizing the retinal connections, as Rosenblatt did.

On pages 194–199, the nearest-neighbor method is discussed. The authors would have done well to cite the elegant results of T. Cover (1967b), showing the effectiveness of this method.

Page 205, line 2*, a factor p_j should be inserted on the right side of the equation.

The figure on page 206 is somewhat misleading; it should be noted that the input to $\prod_j$ is $\prod_i (p_{ij}/q_{ij})\, \phi_i$.

In Section 12.6 reference should be made to the remarkable finding of T. Cover (1967a) that, for a large number of patterns in general position, the number of patterns that can be separated is approximately twice the number of adjustable weights.

Page 244, line 19, refers to Theorem 11.6. There doesn't seem to be any theorem with this number. What is probably intended is the Theorem in Section 11.9. Similarly on line 23, the reference to Section 11.6 should probably read Section 11.10.

Page 248, line 15*, change *proceduce* to *procedure*.

On page 248 (lines 6*–4*) it is stated that "the [perceptron convergence] theorem would have been instantly obvious had the cyberneticists interested in perceptrons known about Agmon's work." Since there is nothing in "Agmon's work" (1954) about termination of the process after a finite number of steps, this aspect of the theorem at least does not seem to be "instantly obvious". Furthermore, it is not clear who "the cyberneticists" are; but presumably the authors do not include themselves in this category. One might wonder why the rebuke does not apply to all those interested in the perceptron, e.g., Papert (1961), or Minsky and Selfridge (1961), rather than just to "the cyberneticists". In this connection one may also wonder about the remark on page 4 (lines 5*–3*): "We feel, in fact, that the solemn experts who most complained about the 'exaggerated claims' of the cybernetic enthusiasts were, in the balance, much more in the wrong." Are the authors to be counted among the "solemn experts"? (cf. Minsky (1961)).

5. Comparison with Rosenblatt's Perceptrons

Let us compare the perceptrons studied in this book with the *Perceptrons* introduced by F. Rosenblatt (1957) and investigated extensively by him and others over the past fifteen years. (We denote Rosenblatt's version by italics, as indicated.)

A perceptron, as defined by Minsky and Papert, is slightly more general than what Rosenblatt called a *simple Perceptron*. (While Minsky and Papert allow all predicates ϕ to participate in the linear threshold function, Rosenblatt allowed only "neurons"; not every predicate can be realized by a "neuron.") On the other hand, the *simple Perceptron* (which consists of a set of inputs, one layer of neurons, and a single output, with no feedback or cross coupling) is not at all what a *Perceptron* enthusiast would consider a *typical Perceptron*. He would be more interested in *Perceptrons* with several layers, feedback, and cross coupling. Let us take a few moments to explain why this is so.

By 1930 it had become generally accepted that the mind resides in the brain and the brain is packed with neurons. The threshold response of neurons and their electrochemical pulses were also known, as was the general nature of the synapses. Scientists therefore began to look forward to an explanation of brain functions, such as memory, perception, or reasoning, in terms of brain structures, such as neurons, synapses, and thresholds. The classic paper of McCulloch and Pitts (1943) showed that all logical functions could be effected by simple mathematical abstractions of neurons. (Incidentally, Kleene (1956), in clarifying the results of McCulloch and Pitts, introduced the connection between "regular expressions" and "finite-state machines," thus initiating an important part of the field of computer science;

but these developments turned away from the problem of brain modelling.) There followed a good deal of discussion on "neuroeconomics": If the neurons in the brain were of this simple type, would there be enough of them to account for brain functions? It was also asked whether the specific configurations used by McCulloch and Pitts actually occur in the brain. The answer to both questions appeared to be negative, or at least not encouraging for brain models of this type. A model proposed by Hebb (1949), using reverberating cell assemblies, appeared to overcome some of these difficulties; but Hebb's model was vague in many details.

One purpose of Rosenblatt's *Perceptrons* was to define the Hebb model more precisely, so that its performance could be analyzed mathematically. One of the crucial features of Rosenblatt's machines was the provision for *change* in the neural net as the result of its activity. Only by including a mechanism for change in the net can one hope to achieve a model for memory and learning. The analysis of complicated networks of this type appeared to be very difficult. In order to see if a form of learning occurs, even in the most primitive case, the *simple Perceptron* was studied first, and for it the "*Perceptron* convergence theorem" was proved. This was encouraging, not because the *simple Perceptron* is itself a reasonable brain model (which it certainly is not; no existing *Perceptron* can even begin to compete with a mouse!), but because it showed that adaptive neural nets, in their simplest forms, could, in principle, improve. This suggested that more complicated networks might exhibit more interesting behavior. Minsky and Papert view the rôle of the *simple Perceptron* differently: p. 247 (lines 15–17) "... a key part of the process leading to the convergence theorem was the molding of the concept of the machine to the appropriate form." Thus, what the *Perceptronists* took to be a temporary handhold, Minsky and Papert interpret as the final structure.

Another difference seems to stem from a venerable misunderstanding as to why Rosenblatt used randomized connections. Switching theorists look for some particular virtue in this arrangment. Finding none, they condemn the system as inefficient. Except for certain small blessings of independence mentioned earlier (at the end of the passage in Section 4 on "Bayesian Decision Procedures,...") there is *no* particular virtue in randomization. This is precisely the reason it is used! If a randomized net can learn, then certainly so can a net with carefully specified connections. The postulation of a highly specific connection scheme obliges the modeller to find such a network in the brain. He would also have to face such problems as explaining how functions of damaged parts of the brain are taken over by other parts; or how one learns to see with inverting eyeglasses. So in the sense of a "worst case" or "minimal constraint" feasibility study, one randomizes the connections. This approach goes back at least to Craik (1943):

"Models of the brain–on the pattern of a telephone exchange–would be much more convincing if they did not postulate any particular connections. Such constancy of connections is very unlikely in view of individual variations in micro-anatomy.

"It is possible that a brain consisting of randomly connected impressionable synapses would assume the required degree of orderliness as a result of experience, just as a randomly connected telephone exchange might become usable if any pair of people could lower the resistance of their line and so get into audible communication, if they tried often enough."

It also appears in McCulloch (1951):

"...[Von Neumann] has to be very careful to specify in detail which relays are to be connected to a given relay to trip it. That is not the case in human brains. Wiener has calculated that the maximum amount of information our chromosomes can convey would fill one volume of the *Encyclopedia Britannica*, which could specify all the connections of ten thousand neurons if that was all it had to do. As we have 10^{10} neurons, we can inherit only the general scheme of the structure of our brains. The rest must be left to chance. Chance includes experience which engenders learning. Ramon y Cajal suggested that learning was the growing of new connections."

The idea is also implicit in Ashby's (1956) "Law of Requisite Variety," where the valuable configurations would become strengthened and useless ones atrophied.

In recent days, even the autonomic nervous system has been suspected of learning [Miller (1969), DiCara (1970)]. This belief has been held for some time in the East.

In some *Perceptrons*, Rosenblatt (1962) does use specific-connection models, patterned after the arrangements found by Hubel and Wiesel (1959) in the cat and Lettvin et al. (1959) in the frog. These do perform better, but in a certain sense evade the deeper problem of the mechanism of learning.

One may conceive models in two fundamentally different ways. Given several competing theories, models may be employed to decide which theory is closest to the observed facts. On the other hand, given an observed function, one might seek to construct a model which, without regard to accuracy in detail, exhibits that function. The latter was Rosenblatt's approach. Since there exists no model which even remotely imitates brain function, Rosenblatt's aim was to devise some model, of no greater complexity than biological neurons, which might be capable of approximating at least some elementary brain functions; cf. Block (1962):

"...Admittedly the model represents an enormous simplification of even the known brain structure; but if it does not violate the biological constraints (such as the number of units, the organization of connections, the reliability of components, the mechanism of signal transmission, the speed of response, the stability of the performance with respect to component malfunction or extirpation, the capacity for information storage, etc.) and if it exhibits even rudimentary brain functions, then, even if it does not in fact operate in the same manner as the brain does, it still provides at least a possible explanation of how the brain structure, as we know it at this time, *might* be organized to perform these functions."

The opposite view of the rôle of models can be seen in Minsky and Papert p. 211, line 17:

"Thus the simple anatomy, combined with the membrane becoming permeable briefly following a nerve impulse, could give a quantity that is an estimator of the appropriate probability.

"How could this representation of probability be translated into a useful neuronal mechanism? One could image all sorts of schemes: ionic concentrations— or rather, their logarithms!—could become membrane potentials, or conductivities, or even probabilities of occurrences of other chemical events. The 'anatomy' and 'physiology' of our model could easily be modified to obtain likelihood ratios. Indeed, it is so easy to imagine variants—*the idea is so insensitive to details* [my italics—HDB]—that we don't propose it seriously, except as a family of simple yet intriguing models that a neural theorist should know about."

"Insensitivity to details" is a fault in the eyes of someone trying to decide which among several theories is true; but it is a virtue to someone trying to find some model, not inconsistent with the known facts, that will function in a specific way. Both approaches are, of course, legitimate, but a lack of understanding of the differing perspectives of the parties can lead to much fruitless debate.

To exhibit the difference in still another way: Suppose that we randomly scrambled 10^{10} mathematical neurons, furnished inputs, specified reinforcement rules, and observed outputs. Suppose further that we then found that the mass functioned like a brain; perceiving, thinking, remembering, deciding, guessing, controlling purposeful behavior, etc. This might be very exciting; but it really wouldn't prove anything about how any actual brain works. It is incorrect in all details; it is too complicated to analyze; it can't tell us anything about the function of the hippocampus, or where memories are stored or how they are called up. It does suggest how a brain *might* be organized, but has nothing to contribute to the question of how any living brain is *in fact* organized. A *Perceptron* is not a description of a brain; it is rather a direction and a hope.

Work on the four-layer *Perceptrons* has been difficult; but the results suggest that such systems may be rich in behavioral possibilities, once the mathematical tools become available to analyze them [cf. Rosenblatt (1960), (1964), Block, Knight, and Rosenblatt (1962), Konheim (1963)]. Even more suggestive are the multilayer machines with feedback (the *C*-systems and *F*-systems of Rosenblatt (1967)). The models studied extensively by Grossberg (1967–1969), although differing from the *Perceptron* in several respects (continuous variables, instead of discrete; linear, instead of a step-thresholding function, etc.) are nevertheless much closer to the spirit of Rosenblatt's *Perceptron* than is the book under review. The same can be said of other brain models, such as those of Kabrisky (1966) or Baron (1970a), (1970b). From this point of view, the potential capabilities of *Perceptrons* are still mostly unexplored.

Another indication of this difference of perspective is Minsky and Papert's concern with such predicates as *parity* and *connectedness*. Human beings cannot perceive the parity of large sets (is the number of dots in a newspaper photograph *even* or *odd*?), nor connectedness (on the cover of Minsky and Papert's book are two patterns; one is connected, one is not. It is virtually impossible to determine by visual examination which is which). Rosenblatt would be content to begin to approach human capabilities, and in fact would tend to regard unfavorably a machine which went beyond them, since it is human perception he is trying to approximate. Recognition of commonly occurring shapes, familiar faces, partially obscured objects, the detection of significant features, etc., would seem to provide more relevant tests than parity or connectedness. Rosenblatt's approach would call for quantitative studies of more natural recognition problems.

While Minsky and Papert enumerate at length the difficulties that such predicates as parity and connectedness cause for their perceptrons, they neglect to mention the remarkable ability of *Perceptrons* to continue to function reliably even after many of their components have been destroyed. This capability is inherent in the organization of *Perceptrons* and does not require special arrangements. "Reliability of the system in spite of malfunction of the components" is important to the Rosenblatt viewpoint because it is common in biological systems but rare in computers, where the malfunction of a single element generally results in a nonsensical output.

Thus, although the authors state (p. 4, lines 12–14) "we have agreed to use the name 'perceptron' in recognition of the pioneer work of Frank Rosenblatt.", they study a severely limited class of machines from a viewpoint quite alien to Rosenblatt's. As a result, the title of the book, although generous in intent, is seriously misleading to the naïve reader who wants to find out something about the general class of *Perceptrons*.

In summary then, Minsky and Papert use the word perceptron to denote a restricted subset of the general class of *Perceptrons*. They show that these simple machines are limited in their capabilities. This approach is reminiscent of the *möhel* who throws the baby into the furnace, hands the father the foreskin and says, "Here it is; but it will never amount to much."

6. Conclusions

We might address the following comments to the three classes of readers for whom Minsky and Papert have intended their book.

1. Specialists in "pattern recognition" who are interested in the practical recognition of visual patterns by computers will find the book of limited value. Such readers might more usefully consult the books of Kolers and Eden (1968), Rosenfeld (1969), Grasselli (1970), Cheng et al. (1970), or Uhr (1966). Although some of Minsky and Papert's theorems might prove useful, there seem to be more promising avenues to practical pattern recognition; as Minsky and Papert indicate, one might be Guzman's (1968); others would be Alan Shaw's (1969), (1970), and David Noton's (1969), (1970).

"Learning Machine" enthusiasts will find, unfortunately for their purposes, that most of the book concerns *fixed* networks. However, the chapters on "learning machines" are sprinkled with provocative comments, which many readers may find stimulating and instructive.

Specialists in "threshold logic" may find the mathematical techniques useful. Unless they are concerned with visual pattern recognition however, they may find much of the theory to be outside their domain of interest.

2. I had thought that abstract mathematicians might find intriguing the idea of a computational geometry. Those with whom I have tried to discuss the subject were not captivated. They objected to the concept of order as not taking into account the *number* of predicates used. Thus, for example, they felt that *convexity*, although of order three, involved so many predicates that it really should be of infinite order. In the same way, they felt that the *And-Or Theorem*, far from revealing a deep fact about the nature of geometry, was merely the consequence of a poor choice of the basic definitions. If a relation as primitive (and as easy to implement technically) as "and" or "or" can, when it operates on two *first-order* predicates, produce a predicate of *infinite* order, then some doubt is cast on the appropriateness of "order" as a measure of complexity. Also, it is easy to prove that there exists a set of predicates $\mathbf{\Phi} = \{\phi\}$, each having only one point of support,

such that any predicate $\psi \in L(L(\Phi))$. Does this mean that with two layers, all predicates are of order one? If not, how does one define *order* for multilayer machines? It seems that there should be a trade-off between logical depth (number of layers) and complexity in a given layer. This basic relationship still remains to be formulated and explored.

I don't take too seriously these criticisms offered by pure mathematicians, recalling the reception accorded Heaviside and Dirac; but the notion of *Computational Geometry* did stir up a lot of apathy.

3. For psychologists and biologists, the level of mathematical maturity demanded will, I believe, make the book somewhat difficult to read. Moreover, since the types of neural nets that Minsky and Papert study are very restricted, it seems unlikely that theorems about their *limitations* can be of much relevance to psychologists or biologists. It is like demonstrating to a surgeon that if he wears boxing gloves he cannot possibly operate effectively.

The absence of Exercises or Problems might limit the usefulness of the book as a classroom text.

As to future research in the new theory: I would expect that the charisma of the authors will attract many able young workers. I would also expect that any results which have eluded the mathematical inventiveness of Minsky and Papert will turn out to be very difficult indeed to establish. One has the impression that if there were anything further of interest in this direction, Minsky and Papert would probably have found it! Thus, one would expect that the next phase of this research will turn toward more elaborate systems: models having several layers, heirarchical organization, feature detectors, feedback, etc.

In sum then, Minsky and Papert's formulation of their theory of perceptrons is precise and elegant. Their mathematical analysis is brilliant. Their exposition is lively, often bombastic, and, occasionally, snide (p. 242, lines 14*–11*, "We were pleased and encouraged by the enthusiastic reception by many colleagues at the A.M.S. meeting and no less so by the doleful reception of a similar presentation at a Bionics meeting."). Two questions remain:

Will the new subject of "Computational Geometry" grow into an active field of mathematics; or will it peter out in a miscellany of dead ends?

Will the formulations or methods developed in the book have a serious influence on future research in pattern recognition, threshold logic, psychology, or biology; or will this book prove to be only a monument to the mathematical virtuosity of Minsky and Papert?

We shall have to wait for a few years to find out.

Acknowledgments

I am grateful to several colleagues for their criticisms of earlier drafts of this review. In particular, I would like to thank Professor Simon Levin for many cogent technical suggestions and Professor Edward Morris for helping me improve the presentation.

References

Agmon, Shmuel (1954), The relaxation method for linear inequalities, *Canad. J. Math.* **6**, No. 3, 382–392.

Amer. Math. Monthly **76**, No. 7 (Aug.–Sept. 1969), p. 856.

Ashby, W. Ross (1952), "Design for a Brain," Chapman and Hall, Ltd., London.

Ashby, W. Ross (1956), "An Introduction to Cybernetics," Chapman and Hall, Ltd., London.

Baron, R. J. (1970a), A model for cortical memory, *J. Math. Psych.* **7**, 37–59.

Baron, R. J. (1970b), A model for the elementary visual networks of the human brain, *Internat. J. Man-Machine Studies* **2**, to appear.

Block, H. D. (1962), The perceptron: A model for brain functioning, I, *Rev. Mod. Phys.* **34**, 123–135 (citation from p. 126, col. 2).

Block, H. D., Knight, B. W., and Rosenblatt, F. (1962), Analysis of a four layer series-coupled perceptron, *Rev. Mod. Phys.* **34**, 135–142.

Block, H. D., and Levin, S. A. (1970), On the boundedness of an iterative procedure for solving a system of linear inequalities, *Proc. Amer. Math. Soc.*, **26**, No. **2**, 229-235.

Cheng, George C., Ledley, Robert S., Pollack, Donald K., and Rosenfeld, Azriel (Eds.) (1968), "Pictorial Pattern Recognition," Thompson Book Co., Washington, D. C.

Cover, Thomas M. (1967a), Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition, *IEEE Trans. Electron. Comput.* **EC-14**, No. 3.

Cover, T. M., and Hart, P. E. (1967b), Nearest neighbor pattern classification, *IEEE Trans. Information Theory* **IT-13**, 21–27.

Cover, Thomas M. (1969), Learning in pattern recognition, *in* "Methodologies in Pattern Recognition" (M. Watanabe, Ed.), Academic Press, New York, 1969.

Craik, Kenneth (1943), "The Nature of Explanation," Cambridge University Press, Cambridge, England (citation from p. 115, Paperback ed., 1967).

DiCara, Leo V. (1970), Learning in the autonomic nervous system, *Sci. Amer.* **222**, No. 1, 30–39.

Efron, Bradley (1964), "The Perceptron Correction Procedure in Nonseparable Situations," Rome Air Development Center Technical Documentary Report RADC-TDR-63-533.

Grasselli, A. (Ed.) (1970), "Automatic Interpretation and Classification of Figures," Academic Press, New York.

Grossberg, Stephen (1967), Nonlinear difference-differential equations in prediction and learning theory, *Proc. Nat. Acad. Sci. U.S.A.* **58**, 1329–1334.

Grossberg, Stephen (1968), Some nonlinear networks capable of learning a spatial pattern of arbitrary complexity, *Proc. Nat. Acad. Sci. U.S.A.* **59**, 368–372.

GROSSBERG, STEPHEN (1969a), Some networks that can learn, remember, and reproduce any number of complicated space-time patterns, Part I, *J. Math. Mech.* **19**, 53–91.

GROSSBERG, STEPHEN (1969b), Some networks that can learn, remember, and reproduce any number of complicated space-time patterns, part II, *SIAM J. Appl. Math.*, in press.

GROSSBERG, STEPHEN (1969c), On the serial learning of lists, *Math. Biosci.* **4**, 201–253.

GROSSBERG, STEPHEN (1969d), On the production and release of chemical transmitters and related topics in cellular control, *J. Theor. Biol.* **22**, 325–364.

GUZMAN, ADOLFO (1968), Decomposition of a visual scene into bodies, "Proceedings of Fall Joint Computer Conference, 1968."

HEBB, DONALD O. (1949), "The Organization of Behavior," John Wiley and Sons, New York, 1949.

HOUSEHOLDER, A. S., AND LANDAHL, H. D. (1945), "Mathematical Biophysics of the Central Nervous System," Math. Biophysics Monograph Series, No. 1, Principia Press, Bloomington, Ind.

HUBEL, D. H., AND WIESEL, T. N. (1959), Receptive fields of single neurons in the cat's striate cortex, *J. Physiol. (London)* **148**, 574–591.

KABRISKY, MATTHEW (1966), "A Proposed Model for Visual Information Processing in the Human Brain." University of Illinois Press, Urbana, Ill.

KAMEDA, T. (1967), "A New Relaxation Algorithm of Adaptation which Finds the Solution with Minimum Magnitude," Technical Note No. 3, Digital Systems Lab., Princeton University, Princeton, N. J. See also *Comput. Rev.* (March 1968), p. 140, Review No. 13,796.

KANAL, LAVEEN N. (Ed.) (1968), "Pattern Recognition," Thompson Books, Washington, D. C.

KLEENE, STEPHEN C. (1956), Representation of events in nerve nets and finite automata, *in* "Automata Studies" (C. E. Shannon and J. McCarthy, Eds.), Annals of Mathematics Studies No. 34, Princeton Univ. Press, Princeton, N. J.

KOLERS, PAUL A., AND EDEN, MURRAY (1968), "Recognizing Patterns," The M.I.T. Press, Cambridge, Mass.

KONHEIM, ALAN G. (1963), A geometric convergence theorem for the perceptron, *J. Soc. Indust. Appl. Math.* **11**, 1–14.

LETTVIN, JEROME Y., MATURANA, H., MCCULLOCH, W. S., AND PITTS, W. (1959), What the frog's eye tells the frog's brain, *Proc. IRE* **47**, 1940–1951.

LEWIS, P. M. (1966), A lower bound on the number of corrections required for convergence of the single threshold gate adaptive procedure, *IEEE Trans. Electron. Comput.* **EC-15**, 933–935.

MCCULLOCH, WARREN S., AND PITTS, WALTER H. (1943), A logical calculus of the ideas immanent in nervous activity, *Bull. Math. Biophys.* **5**, 115–133.

MCCULLOCH, WARREN S. (1951), Why the mind is in the head, *in* "Cerebral Mechanisms in Behavior," The Hixon Symposium (Jeffress, Lloyd A., Ed.), John Wiley and Sons, New York (citation from p. 55).

MONTANARI, U. (1970), A note on minimal length polygonal approximation to a digitized contour, *Commun. ACM* **13**, 41–47.

MILLER, NEAL E. (1969), Learning of visceral and glandular responses, *Science* **163**, 434–445.

MINSKY, M. L. (1961), Steps toward artificial intelligence, *Proc. IRE* **49**, 8–30.

MINSKY, MARVIN, AND SELFRIDGE, OLIVER G. (1961), Learning in neural nets, "Proc.

of the Fourth London Symposium on Information Theory" (C. Cherry, Ed.), Academic Press, New York.

NEWELL, ALLEN (1969), *Science* **165**, 780–782.

NILSSON, NILS J. (1965), "Learning Machines," McGraw-Hill Book Co., New York.

NOTON, DAVID (1969), A proposal for serial, archetype-directed pattern recognition, "Record of the 1969 IEEE System Science and Cybernetics Conference, Philadelphia, Pa., Oct. 22–24, 1969," IEEE Catalog No. 69C-37-SSC.

NOTON, DAVID (1970), A theory of visual pattern recognition, *IEEE Trans. System Sci. Cybernetics*, to appear.

PAPERT, SEYMOUR (1961), Some mathematical models of learning, "Proc. of the Fourth London Symposium on Information Theory" (C. Cherry, Ed.), Academic Press, New York.

PITTS, W. R., AND MCCULLOCH, W. S. (1947), How we know universals, *Bull. Math. Biophys.* **9**, 124–147.

ROSENBLATT, FRANK (1957), "The perceptron, a perceiving and recognizing automaton (Project PARA)," Cornell Aeronautical Laboratory Report No. 85-460-1, Buffalo, New York.

ROSENBLATT, FRANK (1960), Perceptual generalization over transformation groups, *in* "Self-Organizing Systems" (Yovits and Cameron, Eds.), pp. 63–96, Pergamon Press, New York.

ROSENBLATT, FRANK (1962), A comparison of several perceptron models, *in* "Self-Organizing Systems" (Yovits, Jacobi, and Goldstein, Eds.), Spartan Books, Washington, D. C.

ROSENBLATT, FRANK (1962), "Principles of Neurodynamics," Spartan Books, Washington, D. C.

ROSENBLATT, FRANK (1964), A model for experiential storage in neural networks, *in* "Computer and Information Sciences" (J. T. Tou and R. H. Wilcox, Eds.), pp. 16–66, Spartan Books, Washington, D. C.

ROSENBLATT, FRANK (1967), Recent work on theoretical models of biological memory, *in* "Computer and Information Sciences—II" (Julius Tou, Ed.), pp. 33–56, Academic Press, New York.

ROSENFELD, AZRIEL (1969), "Picture Processing by Computer," Academic Press, New York.

SHAW, ALAN C. (1969), A formal picture description scheme as a basis for picture processing systems, *Information and Control* **14**, 9–52.

SHAW, ALAN C. (1970), Parsing of graph-representable pictures, *J. Assoc. Comput. Mach.*, to appear.

SHIMBEL, A., AND RAPOPORT, A (1948), A statistical approach to the theory of the central nervous system, *Bull. Math. Biophys.* **10**, 41–55.

SKLANSKY, J. (1970), Recognition of convex blobs, "Pattern Recognition" **2**, No. 1, to appear.

UHR, LEONARD (Ed.) (1966), "Pattern Recognition," John Wiley and Sons, New York.

WATANABE, SATOSI (Ed.) (1969), "Methodologies of Pattern Recognition," Academic Press, New York.