PERCEPTRON-LIKE MACHINES

by

KRIS LANCE KRAFT
A.B., University of California (Berkeley), 1965

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

In the Department

of

MATHEMATICS

We accept this thesis as conforming to the
required standard

The University of British Columbia
April 1969

Department of Mathematics

The University of British Columbia
Vancouver 8, Canada

Date April 30, 1969

Supervisor: R. S. Rosenberg.

ABSTRACT

This paper investigates perceptron-like devices with an eye toward the recognition of simple predicates, such as parity and connectivity. First "multilayer" perceptrons are investigated to little avail. Then much more powerful perceptrons which have feedback are considered, yielding better results.

Examiners

TABLE OF CONTENTS

LIST OF FIGURES

INTRODUCTION

The purpose of this thesis is to further explore the capabilities of various kinds of "perceptron-like" devices. The main inspiration, spirit, approach and source of style will be a monograph of Minsky and Papert entitled Perceptrons and Pattern Recognition, (Sept. 1967). This monograph* is remarkable, not so much for impressive results as for its straight-forward unassuming approach which removes much of the mystery (and potential controversy) which has dogged earlier perceptron investigations.

The Minsky paper explores the complexity of various logical and geometric concepts with respect to a very simple kind of perceptron, a linear separation machine, where there is no "feed-back" or other communication among the "associator units".

Briefly these perceptrons operate in the following manner: They are presented with a rectangular array (called a retina) of squares which may be in the active or inactive state. The so-called "associator units" then look at small parts of the retina and decide on that basis to vote "yes" or "no". All votes are tallied and a particular weighted average taken of the votes. How the average is weighted, of course, varies from machine to machine. If the weighted average is above a certain threshold

* The monograph (plus some extended results) has just been published in book form. Perceptrons, M. Minsky and S. Papert, M.I.T. Press (1969). Some of the material in this paper has been anticipated though not expanded upon in Perceptrons on pages 228-232, a section which did not appear in 1967.

the machine is said to "accept" the particular activated subset of the retina, otherwise there is rejection.

The set of all accepted subsets is said to define a predicate or a boolean function on the retina. The machine is said to recognize the predicate so defined.

Perceptrons or closely related devices are often used in pattern recognition programs for computers because they are so simple in design (and easy to program). Rosenblatt did much work in the pattern recognition area with such machines "generated at random". Other "learning" programs, such as the Samuel Checker Player, also closely resemble perceptrons. Samuel, for example, has many so-called "board parameters" such as "number of pieces", and "number of possible moves". These parameters are his associators, for they grasp only a small part of the total information on the board. His parameters are combined in wieghted averages to help determine what is supposed to be the best checker move.

Minsky on the other hand is not concerned with the applications of perceptrons, but rather with the theoretical limit of their ability. To explore this question he defined a very restricted and specific kind of perceptron and investigates that.

Unfortunately these linear separation machines are not powerful enough to deal with many concepts (i.e., predicates) such as parity and connectivity which seem to the human at least, to be on the most elementary level. We are constantly surprised with the difficulty (high order) of some tasks as compared to the ease (low order) of others.

This paper will explore other designs for perceptrons along the lines of Minsky with an eye toward the difficulty of computation involved in various tasks.

Basically, the thesis will explore perceptrons with many associator "levels", and then perceptrons which allow feedback from machine to retina. The first alternative will not be especially fruitful. However, the second subject (feedback perceptrons) will be quite interesting and deserving of much more study than it can get here.

PERCEPTRON-LIKE MACHINES

We assume that we have a machine which is presented with a rectangular array R of squares which may be occupied (blacked in) or not. We wish to say certain things about the array (often called the <u>retina</u>).

<u>Definition</u>: A <u>predicate function</u> on R is a function $\varphi : P(R) \to \{0,1\}$, where $P(R)$ is the set of all subsets of R. A <u>predicate</u> is written as $[\varphi]$ where $[1] = T$ and $[0] = F$. For any $X \subseteq R$ we say $[\varphi(X)]$ <u>holds</u> or <u>is true</u> iff $\varphi(X) = 1$. The two terms will be used interchangeably henceforth.

<u>Proposition</u>: There are $2^{2^{|R|}}$ predicates on R.

<u>Proof</u>: There are clearly $2^{|R|}$ subsets $X \subseteq R$ in the domain of any predicate. Certain X's in dom (φ) are "accepted by φ" (i.e. $\varphi(X) = 1$) and certain are "rejected". Thus φ is uniquely described by the set of X's accepted i.e. by a subset of dom (φ). Since there are $2^{2^{|R|}}$ subsets of dom (φ) there are $2^{2^{|R|}}$ predicates φ.

A particularly simple and useful class of predicates are called <u>masks</u>.

<u>Definition</u>: A <u>mask</u> φ_A is a predicate such that either $\varphi(x) = 0$ for any $X \subseteq R$ or there is a set $A \subseteq R$ such that

$\varphi_A(X) = 1$ iff $X \supseteq A$ for all $X \subseteq R$. The zero mask will be denoted as φ_0; the one mask (logically) as φ_ϕ.

Definition: Suppose Φ is a set of masks. We say that ψ is realizable from Φ by a simple mask perceptron iff for all $X \subseteq R$

$$[\psi(X) = 1] \leftrightarrow \sum_{\varphi_i \in \Phi} \alpha_{\varphi_i} \varphi_i(X) \geq \theta \text{ for some reals } \alpha_{\varphi_i}, \theta .$$

The simple mask perceptron which can be thought of as a separate piece of "concrete" hardware (consisting of the threshold θ, the coefficients α_{φ_i} and the masks φ_i) is said to compute φ from Φ. We write $\varphi \in$ S.M.P. (Φ) in this case.

Remark: We could have assumed $\theta = 0$, α_{φ_i} to be rational or even integral in R, and we could also have written $>$ instead of $\geq$, obtaining an equivalent definition.

Proof: (i) If we make the assumption (which we will henceforth make) that $\varphi_\phi \in \Phi$ then

$$\sum_{\varphi_i \in \Phi} \alpha_{\varphi_i} \varphi_i(X) > \theta \text{ iff } \sum_{\substack{\varphi_i \in \Phi \\ \varphi_i \neq \varphi_\phi}} \alpha_{\varphi_i} \varphi_i(X) + (\alpha_{\varphi_\phi} - \theta)\varphi_\phi(X) > \theta$$

showing that θ can always be taken to be zero.

(ii) Only a finite number $(\leq 2^{|R|})$ of sums $S(X) = \sum_{\varphi_i \in \Phi} \alpha_{\varphi_i} \varphi_i(X)$ will occur. If h is the minimum nonzero pairwise difference of the finite set of sums $\{S(X)\}$ we then have $S(X) \geq 0$ iff $S(X) > 0 - {}^{h}/2$ for all $X \subseteq R$, showing that the $\geq$ could be taken as $>$.

(iii) If h is as before then we know that it is possible to pick $\alpha'_{\varphi_{i_0}} \in Q$ close enough to $\alpha_{\varphi_{i_0}}$ such that

$$\left| \sum_{\substack{\varphi_i \in \Phi \\ i \neq i_0}} \alpha_{\varphi_i} \varphi_i(X) + \alpha'_{\varphi_{i_0}} \varphi_{i_0}(X) - \Sigma\, \alpha_{\varphi_i} \varphi_i(X) \right| < \frac{h}{2|\Phi|}$$

for all $X \subseteq R$. If all such α_{φ_i}'s are so replaced we clearly have:

$$\Sigma\, \alpha'_{\varphi_i} \varphi_i(X) > \Sigma\, \alpha'_{\varphi_i} \varphi_i(Y) \quad \text{iff} \quad \Sigma\, \alpha_{\varphi_i} \varphi_i(X) > \Sigma\, \alpha_{\varphi_i} \varphi_i(Y)$$

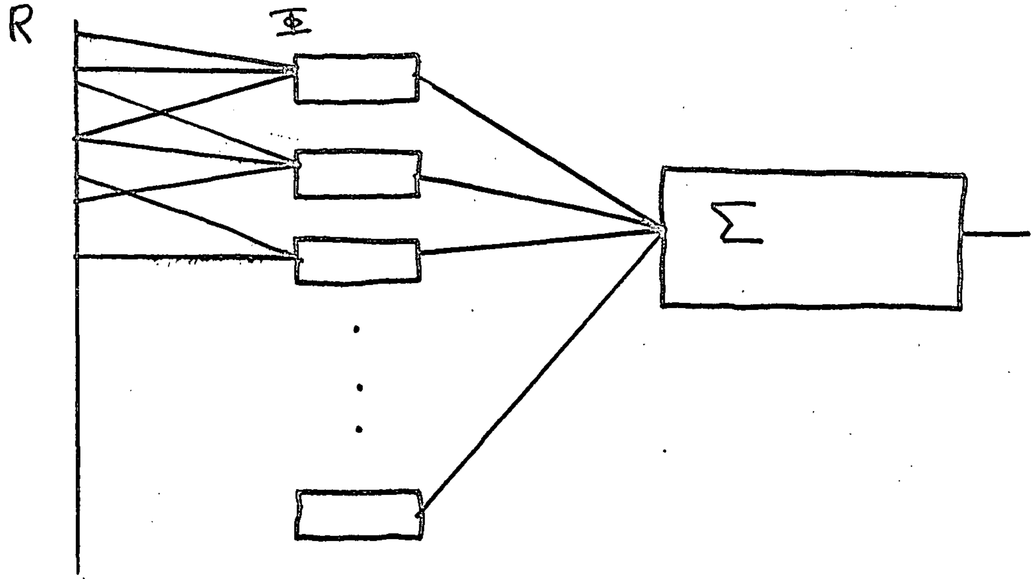
that is preservation of order for all of our finite number of sums, showing that α_{φ_i}'s can be taken to be rational. Multiplying through by the least common denominator shows the same thing for integers.

We now define a (formally) more complicated machine, which will turn out to be equivalent in computing power.

Definition: Suppose Φ is a set of predicates. We say that φ is realizable from Φ by a linear separation perceptron iff

$$[\varphi(X) = 1] \longleftrightarrow \sum_{\varphi_i \in \Phi} \alpha_{\varphi_i} \varphi_i(X) > \theta \quad \text{for some reals } \alpha_{\varphi_i}, \theta$$

and as before we write $\psi \in$ L.S.P. (Φ) .



Remark: As before we could have assumed $\theta = 0, \alpha_{\varphi_i} \in Z$ and $\leq$ to be $<$. We shall henceforth refrain from making this kind of remark explicit.

Proposition: If Φ = the set of all masks then any predicate $\psi \in$ S.M.P. (Φ) .

Proof: We write ψ in disjunctive normal form

$$[\psi(X)] \Leftrightarrow C_1(X) \vee C_2(X) \vee \ldots \vee C_n(X) \quad \text{where}$$

each $\quad C_i(X) = [z_{i_1}] \wedge [z_{i_2}] \wedge \ldots \wedge [z_{i_{mi}}]$ and

each $\quad [z_{ij}] = [X_k]$ or $[\bar{X}_k] \equiv [1-X_k]$ for some $X_k \in R$.

We then have

$$C_i(X) \Leftrightarrow z_{i_1} \cdot z_{i_2} \cdot \ldots \cdot z_{i_{mi}} = 1 .$$

We now replace each z_{ij} by X_k or $1-X_k$ obtaining a polynomial in the X_k's only. No $\bar{X}_k$'s appear, though some terms will be negative and all will therefore be definable by a mask φ or its negative $(-1)\varphi$. Therefore

$$C_i(X) \leftrightarrow \sum_{\varphi\in\Phi_i} \varphi(X) - \sum_{\tilde{\varphi}\in\tilde{\Phi}_i} \tilde{\varphi}(X) \geq 1 .$$

(Note that 0 and 1 are the only values attainable.) and

$$[\psi(X)] \leftrightarrow \sum_{\varphi\in\Phi} \varphi(X) - \sum_{\tilde{\varphi}\in\tilde{\Phi}} \tilde{\varphi}(X) > 0 \quad \text{since}$$

$\psi(X)$ will hold iff any one of the conjuctions C_i is true.

We now introduce the two concepts of <u>support</u> and <u>order</u>, which are to be some measure of how complicated a predicate is supposed to be.

<u>Definition</u>: The <u>support</u> $\underline{A}$ of a predicate φ is the smallest subset $A \subseteq R$ upon which φ depends, i.e. for which $\varphi(X \cap A) = \varphi(A)$ for all $X \subseteq R$. We write <u>$A = \text{supp}(\varphi)$</u>.

<u>Definition</u>: We say that <u>ψ is of order k</u> iff k is the smallest number such that

$$\psi(X) \leftrightarrow \sum \alpha_{\varphi_i} \varphi_i(X) > 0$$

where the φ_i's are all predicates with $|\text{supp}\ \varphi_i| \leq k$ and the α_{φ_i}'s are some real numbers.

<u>Theorem</u>: ψ is of order k iff k is the smallest number such

that $\varphi(X) \Leftrightarrow \sum_{p=1}^{m} \alpha_{\varphi_p} \varphi_p(X) > 0$ where the φ_p's are all masks with $|\text{supp } (\varphi_p)| \leq k$ for all p and the α_{φ_p}'s are reals.

Discussion: This theorem says in effect that the L.S.P.'s and the S.M.P.'s are equivalent in computing power. These machines are the center of study in the Minsky paper. Unfortunately the complexity of predicates with respect to these machines (and the order measure of complexity) does not correspond to the human notion of what kind of things are "easy" and what kind of things are "difficult". For example, we can say that the logical operation of $\neg$ (negation) is easy in that it does not increase the order of $\neg\psi$ over that of ψ. On the other hand the operations of $\vee$ "or" and $\wedge$ "and" can (in a way which will be made more precise later) not even be considered as being of finite order.

In the geometric realm, the counting of points and even the recognition of integers is "easy" as is the ability to recognize certain topological invariants. On the other hand, recognition of parity, and connectivity is not of finite order.

These disappointing results could be viewed as a defect of the simple kind of machines used, or of the order measure, or both. We could, for example, consider complexity as a function of both the support size of the φ_i's and the number of masks or other types of predicates used. Such a theory could be quite intractable, so we consider instead various methods of strengthening the machines. We must be careful though: a machine which is too strong will recognize all predicates in a small order and be quite uninteresting.

We will, of course, be sacrificing one of the "seductive" attributes of the perceptron, namely that it computes in a very straight-forward (and easily simulated way) by first computing a lot of easy to evaluate, simple information and then combining this partial information by a simple algorithm.

Proof: Let k be the least number such that

$$\psi(X) \leftrightarrow \Sigma\, \alpha_{\psi_p} \psi_p(X) > 0 \quad \text{where the } \psi_p\text{'s are}$$

predicates with $|\text{supp}\ (\psi_p)| \leq k$ for all p. We will replace each ψ_p by a sum of the form $\sum_{\varphi\in\Phi} \alpha_\varphi \varphi$ where the φ's are masks such that $\max_{\varphi\in\Phi} |\text{supp}\ \varphi| = |\text{supp}\ \varphi_p|$ the theorem will then follow.

Let $p \leqq m$. As before we will write

$$\psi_p(X) = C_1^p(X) \vee C_2^p(X) \vee \ldots \vee C_n^p(X) \quad \text{where}$$

$$C_i^p(X) = [z_{i_1}^p] \wedge [z_{i_2}^p] \wedge \ldots \wedge [z_{i_{mi}}^p] \quad \text{and}$$

$$[z_{i_j}^p] = [x] \quad \text{or} \quad [1-X] \ .$$

As before we obtain $C_i^p(X) \leftrightarrow \sum_{\varphi\in\Phi_i^p} \varphi(X) - \sum_{\tilde\varphi\in\tilde\Phi_i^p} \tilde\varphi(X) \geq 1$ where Φ_i^p and $\tilde\Phi_i^p$ are index sets of masks. The thing to notice is the identity $\alpha\bar{X}\beta = \alpha(1-X)\beta = (\alpha-\alpha X)\beta = \alpha\beta - \alpha X\beta$ where α and β are conjuctions. This identity shows that the maximum support of the φ's and the $\tilde{\varphi}$'s will be $\text{supp}\ C_i^p(X)$. And since at least one of the C_i^p will have maximal support = $\text{supp}\ \psi_p(X)$ we can write

$$[\psi(X)] \longleftrightarrow \sum_p \alpha_{\psi_p} \left(\sum_{\varphi \in \Phi^p} \varphi(X) - \sum_{\tilde{\varphi} \in \tilde{\Phi}^p} \tilde{\varphi}(X) \right) > 0$$

Remark: So far our predicates have depended upon the retina R. We wish to extend the definition of predicate to predicate schema, a term which will then be used interchangably with predicate.

Definition: A predicate schema is a rule which generates a particular predicate for a retina R of any given size.

Examples: (1) Denote the parity predicate by ψ_{PAR} which has the property that $\psi(X) = 1$ iff $|X|$ is even for every retina R and for all $X \subseteq R$.

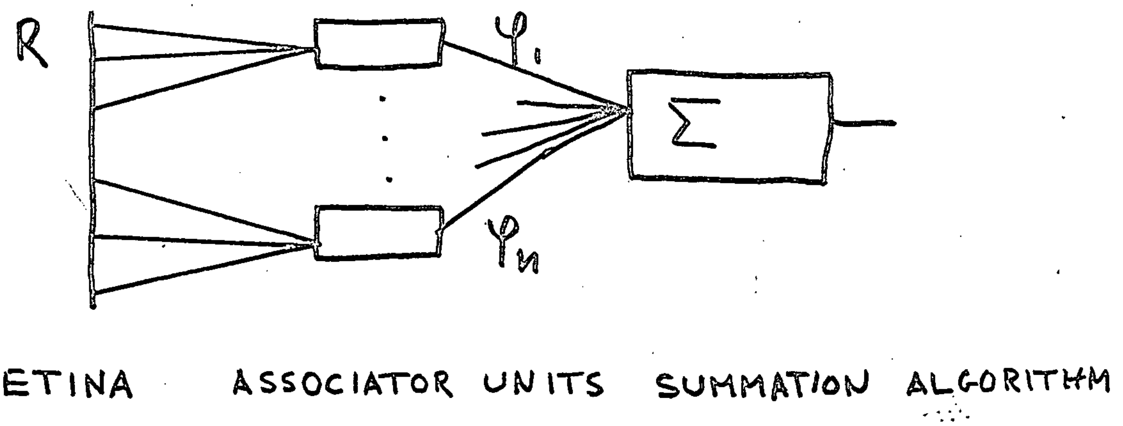
(2) Denote the connectivity predicate by ψ_{CONN} which has the property that $\psi_{CONN} = 1$ iff X is connected for every retina R and for all $X \subseteq R$.

We say that a figure $X \subseteq R$ is connected iff every point $x \in X$ can be connected to any other point $y \in X$ by a rectilinear path in R. We do not permit "diagonal connections" as they would permit two paths to "cross" without "touching".

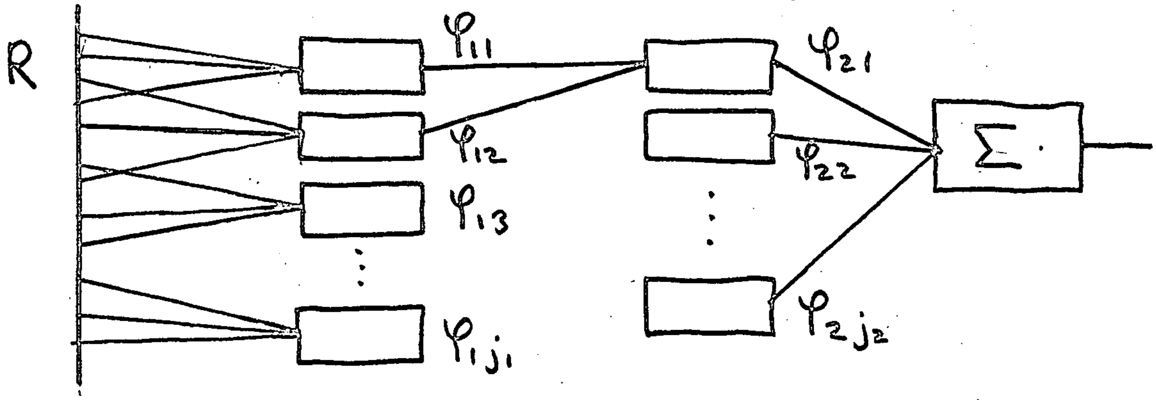
Notation: A rectilinear path γ in R going from x to y will be denoted as $\gamma = \{x = x_0, x_1, x_2, \ldots, x_{n-1}, x_n = y\}$ where x_i and x_{i+1} have a side in common for $i = 0,1,\ldots,n-1$.

<u>Definition</u>: A predicate (schema) ψ is <u>of finite order</u> iff the predicates generated by ψ are uniformly bounded. The minimum uniform bound M is called the <u>order of ψ</u>.

We now begin our search for more complicated kinds of machines which can cope with ψ_{PAR} and ψ_{CONN}. So far we have dealt with machines of the following design:



The associator units have been masks or other predicates the machine has been equipped with. So far they have all been on one "level" and cannot communicate among each other. We now define a 2-level perceptron with the following design.



We will make our definition in a form which can easily be generalized by induction.

Definition: A predicate ψ is realizable from a pair of sets of predicates Φ_1 and Φ_2 iff for all $X \subseteq R$

$$\psi(X) = 1 \Longleftrightarrow \sum_{\varphi_{2i} \epsilon \Phi_2} \alpha_{\varphi_{2i}} (X_1) \geq \theta \quad \text{for some reals } \alpha_{\varphi_{2i}}, \theta$$

where $X_1 = \{\varphi_{1j} \mid \varphi_{1j}(X) = 1\}$. We write

$$\psi \in 2LP\ (\Phi_1, \Phi_2)\ .$$

Remark: We could simplify our notation (and our thoughts) if we thought of our first level of associator units, Φ_1, as a new retina R_1, the output of our 2-level perceptron $P = \langle \theta, \{\alpha_{\varphi_{1i}} \varphi_{1i}\}_{i=1}^{M_1}, \{\alpha_{\varphi_{2i}} \varphi_{2i}\}_{i=1}^{M_2} \rangle$ at time $t = 0$ as $X \subseteq R$, the output at time $t = 1$ as $X_1 \subseteq R_1$ and the output at $t = 2$ as $X_2 \equiv \{\varphi_{2i} \mid \varphi_{2i}(X_1) = 1\} \subseteq R_2$ and the output time $t = 3$ as 0 or 1.

Definition: A predicate ψ is realizable from an n-tuple of sets of predicates $\{\Phi_i\}_{i=1}^{n}$ iff for all $X \subseteq R$

$$\psi(X) = 1 \Longleftrightarrow \sum_{\varphi_{ni} \epsilon \Phi_n} \alpha_{\varphi_{ni}} \varphi_{ni}(X_{n-1}) \geq \theta \quad \alpha_{\varphi_{ni}}, \theta \in R$$

where $X_0 \equiv X \subseteq R$

$$X_k = \{\varphi_{kj} \in \Phi_k \mid \varphi_{kj}(X_{k-1}) = 1\} \quad \text{for } k = 1 \ldots n\ .$$

We write $\psi \in nL.P.\ \{\Phi_i\}_{i=1}^{n}$.

Remark: As before we will think of the i^{th} level of associators Φ_i as a new retina R_i for the next level $i+1$. The output of the n level perceptron P at time t will be X_t for $t = 0,\ldots,n$ and 0 or 1 for $t = n+1$.

Definition: The order of ψ with respect to an n-level perceptron P will be the least number k such that $\psi \in$ nL.P. $\{\Phi_i\}_{i=1}^{n}$ and $\varphi_i \in \Phi_i \rightarrow |\text{supp}\ \varphi_i| \leq k$ for all i.

Theorem: If ψ has order k with respect to a 1-level perceptron (linear separation perceptron) then ψ has order $\leq [\sqrt[2]{k}] + 1$ with respect to a 2-level perceptron. N.B. [ ] denotes the "greatest integer function". e.g. $[5] = 5$, $[5\frac{1}{2}] = 5$, $[\sqrt{2}] = 1$.

Proof: Let $\psi(X) \Longleftrightarrow \Sigma\, \alpha_{\varphi_i}\varphi_i(X) > \theta \quad \alpha_{\varphi_i}, \theta \subseteq R$ and $|\text{supp}\ \varphi_i| \leq k$ for every i. By a previous theorem we might as well assume that the φ_i's are masks. We might further assume that the α_{φ_i}'s are either ± 1 for they may be assumed integral and hence ± 1 if we permit ourselves to copy a particular φ_i several times. We now have

$$\psi(X) \Longleftrightarrow \sum_{\varphi\in\Phi} \varphi(X) - \sum_{\varphi\in\tilde{\Phi}} \varphi(X) > \theta$$

with $|\text{supp}\ \varphi| \leq k$ for any $\varphi \in \Phi \cup \tilde{\Phi}$ where Φ and $\tilde{\Phi}$ are sets of masks.

We now design our 2-level perceptron. For each $\varphi \in \Phi$ (or $\tilde{\Phi}$) divide supp φ up into no more than $[\sqrt[2]{k}] + 1$ disjoint subsets S_i^{φ} containing no more than $[\sqrt[2]{k}\,]+1$ elements each. Such a division must be possible or we would have

$$k = |\mathrm{supp}\ \varphi| > ([\sqrt[2]{k}\,]+1)([\sqrt[2]{k}\,]+1) \geq \sqrt{k}\sqrt{k} = k\ .$$

For each i define a mask $(\psi_1\varphi_i)$ with

$$(\psi_1\varphi_i)(X) = 1 \Longleftrightarrow X \geq S_i^{\varphi} \quad \text{for all} \quad X \subseteq R\ .$$

Let $R_1 = \{(\psi_1\varphi_i) \mid \varphi \in \Phi \quad i \leq [\sqrt[2]{k}\,]+1\}$. Define $(\psi_2\varphi)$ as $(\psi_2\varphi)(X_1) = 1 \Longleftrightarrow (\psi_1\varphi_i) \in X_1$ for all $i \leq [\sqrt[2]{k}\,]+1$. Finally consider the predicate

$$\rho(X) \Longleftrightarrow \sum_{\varphi\in\Phi}(\psi_2\varphi)(X_1) - \sum_{\varphi\in\tilde{\Phi}}(\psi_2\varphi)(X_1) > \theta$$

where $X_1 = \{(\psi_1\varphi_i) \mid (\psi_1\varphi_i)(X) = 1\}$. Clearly ρ is realized from a pair of sets of predicates $\Phi_1 = R_1$ and $\Phi_2 = \{(\psi_2\varphi) \mid \varphi \in \Phi\}$. Clearly also by construction ρ is of order $\leq [\sqrt[2]{k}\,]+1$.

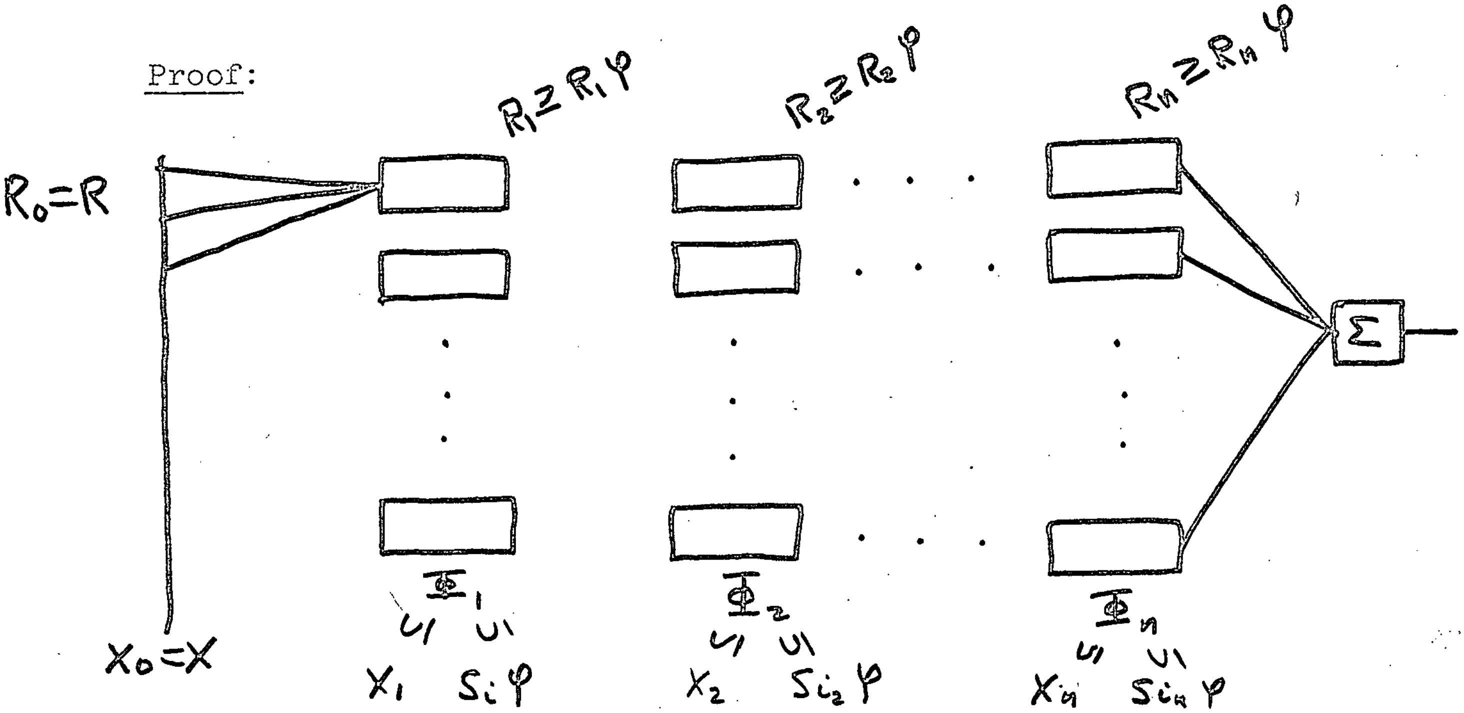
We wish to see $\rho(X) \Longleftrightarrow \psi(X)$. But $\psi(X) \Longleftrightarrow \sum_{\varphi\in\Phi}\varphi(X) - \sum_{\varphi\in\tilde{\Phi}}\varphi(X) > 0$ where $\Phi \cup \tilde{\Phi}$ contains only masks. And $\varphi(X) = 1 \Leftrightarrow X \supseteq S_i^{\varphi}$ for all $i \Leftrightarrow (\psi_1\varphi_i)(X) = 1$ for all $i \Leftrightarrow (\psi_1\varphi_i) \in X_1$ for all $i \Leftrightarrow (\psi_2\varphi)(X_1) = 1$. So our sum $\sum_{\varphi\in\Phi}\varphi(X) - \sum_{\varphi\in\tilde{\Phi}}\varphi(X) = \sum_{\varphi\in\Phi}(\psi_2\varphi)(X_1) - \sum_{\varphi\in\tilde{\Phi}}(\psi_2\varphi)X_1$.

$$\therefore\quad \psi(X) \Leftrightarrow \rho(X)\ .$$

<u>Theorem</u>: If ψ is of order k with respect to a 1-level perceptron then ψ has order $\leq [\sqrt[n]{k}\,]+1$ with respect to an

n-level perceptron.

Proof:



Our proof proceeds as in the previous theorem where we obtain $\psi(X) \Longleftrightarrow \sum_{\varphi\in\Phi} \varphi(X) - \sum_{\varphi\in\tilde{\Phi}} \varphi(X) > 0$ where Φ and $\tilde{\Phi}$ are sets of masks and $|\text{supp } \varphi| \leq k$ for all $\varphi \in \Phi \cup \tilde{\Phi}$.

Pick any $\varphi \in \Phi$ or $\tilde{\Phi}$ and divide supp φ into disjoint subsets S_i^φ containing at most $[\sqrt[n]{k}\,] + 1$ points .
$i \leq \dfrac{k}{[\sqrt[n]{k}\,] + 1}$.

On this "first level" define $(\psi_{i1}\varphi)(X) = 1 \Longleftrightarrow X \supseteq S_i^\varphi$.

Now divide $R_1^\varphi = \{(\psi_{i1}\varphi) i_1 \leq \dfrac{k}{\sqrt[n]{k} + 1}\}$ into disjoint subsets S_{i2}^φ containing at most $[\sqrt[n]{k}\,] + 1$ points where $i_2 \leq \dfrac{k}{([\sqrt[n]{k}\,] + 1)^2}$. Define $(\psi_{i2}\varphi)(X_1) = 1 \Longleftrightarrow X_1 \supseteq S_{i2}\varphi$.

We now have $R_2\varphi = \{\psi_{i2}\varphi) \mid i_2 \leq \dfrac{k}{([\sqrt[n]{k}\,] + 1)^2}\}$.

Proceed inductively, dividing R^{φ}_{m-1} into disjoint subsets $S_{im}\varphi$ that contain at most $[\sqrt[n]{k}\,]+1$ points where $im \subseteq \dfrac{k}{([\sqrt[n]{k}\,]+1)^m}$, then defining on the m^{th} level

$$R_m\varphi = \{(\psi_{im}\varphi) \mid im \le \frac{k}{([\sqrt[n]{k}\,]+1)^m}\} \quad \text{and}$$

$(\psi_{im}\varphi)(X_{m-1}) = 1 \Leftrightarrow X_{m-1} \supseteq S_{im}\varphi$ until $m = n$. At this time

$$i_n \le \frac{k}{(\sqrt[n]{k}+1)^n} \le \frac{k}{(\sqrt[n]{k}\,)^n} = \frac{k}{k} = 1 .$$ That is at level n:

For any $\varphi \in \Phi$ (or $\tilde{\Phi}$) there is a unique $(\psi_{in}\varphi) \equiv \psi_n$.

Consider $\rho(X) \Leftrightarrow \sum_{\varphi\in\Phi} (\psi_n\varphi)(X_{n-1}) - \sum_{\varphi\in\tilde{\Phi}} (\psi_n\varphi)(X_{n-1}) > \theta$

where $X_o = X$.

$$X_{m+1} = \{(\psi_{im+1}\varphi) \in R_{m+1} \mid \varphi \in \Phi \cup \tilde{\Phi} \text{ and } (\psi_{im+1}\varphi)(X_m) = 1\}$$

for $m = 0,\ldots,n-1$.

Clearly (?) ρ is realised by an nL.P. from $\{R_i\}_{i=1}^n$ and is of order $\le [\sqrt[n]{k}\,]+1$.

But $\varphi(X) = 1 \Leftrightarrow X \supseteq \text{supp}\,\varphi$

iff $X \supseteq S_{i_1}\varphi$ $\quad i_1 \le \dfrac{k}{([\sqrt[n]{k}\,]+1)}$

iff $(\psi_{i_1}\varphi)(X) = 1$ $\quad$ "

iff $(\psi_{i_1}\varphi) \in X_1$ $\quad$ "

iff $R_1\varphi \subseteq X_1$

iff $X_1 \supseteq S_{i_2}\varphi$ $\quad i_2 \leq \frac{k}{([\sqrt[n]{k}\,] + 1)^2}$

iff $(\psi_{i_2}\varphi)(X_1) = 1$ $\quad$ "

iff $(\psi_{i_2}\varphi) \in X_2$ $\quad$ "

iff $R_2\varphi \subseteq X_2$

.

.

.

iff $X_{n-1} \supseteq S_{i_n}\varphi$

iff $(\psi_{i_n}\varphi)(X_{n-1}) = 1$

i.e. $(\psi_n\varphi)(X_{n-1}) = 1$

$\therefore \quad \sum_{\varphi\in\Phi} (\psi_n\varphi)(X_{n-1}) - \sum_{\varphi\in\tilde{\Phi}} (\psi_n\varphi)(X_{n-1}) > \theta$

iff $\sum_{\varphi\in\Phi} \varphi(X) - \sum_{\varphi\in\tilde{\Phi}} \varphi(X) > 0$

$\therefore \quad \rho(X) \Leftrightarrow \psi(X)$

Corollary I: If φ is any predicate then there is N large enough so that the order of $\varphi \leq 2$ with respect to an N-level perceptron.

Proof: Let R be a retina and φ a predicate on R. Clearly φ is of order $k \leq |R|$ with respect to a 1-level perceptron. We conclude then that order $\psi \leq [\sqrt[n]{k}\,] + 1$ with respect to an n level perceptron. Since $\lim_{n\to\infty} [\sqrt[n]{k}\,] + 1 = 2$ and $[\sqrt[n]{k}\,] + 1$ is integer valued, we conclude that there is an N such that $[\sqrt[N]{k}\,] + 1 = 2$. Therefore order $\varphi \leq 2$ with respect to an N level perceptron. Most complicated predicates will be of order 2, however some may be of order 1, for example, the order 1 predicates with respect to a 1-level perceptron.

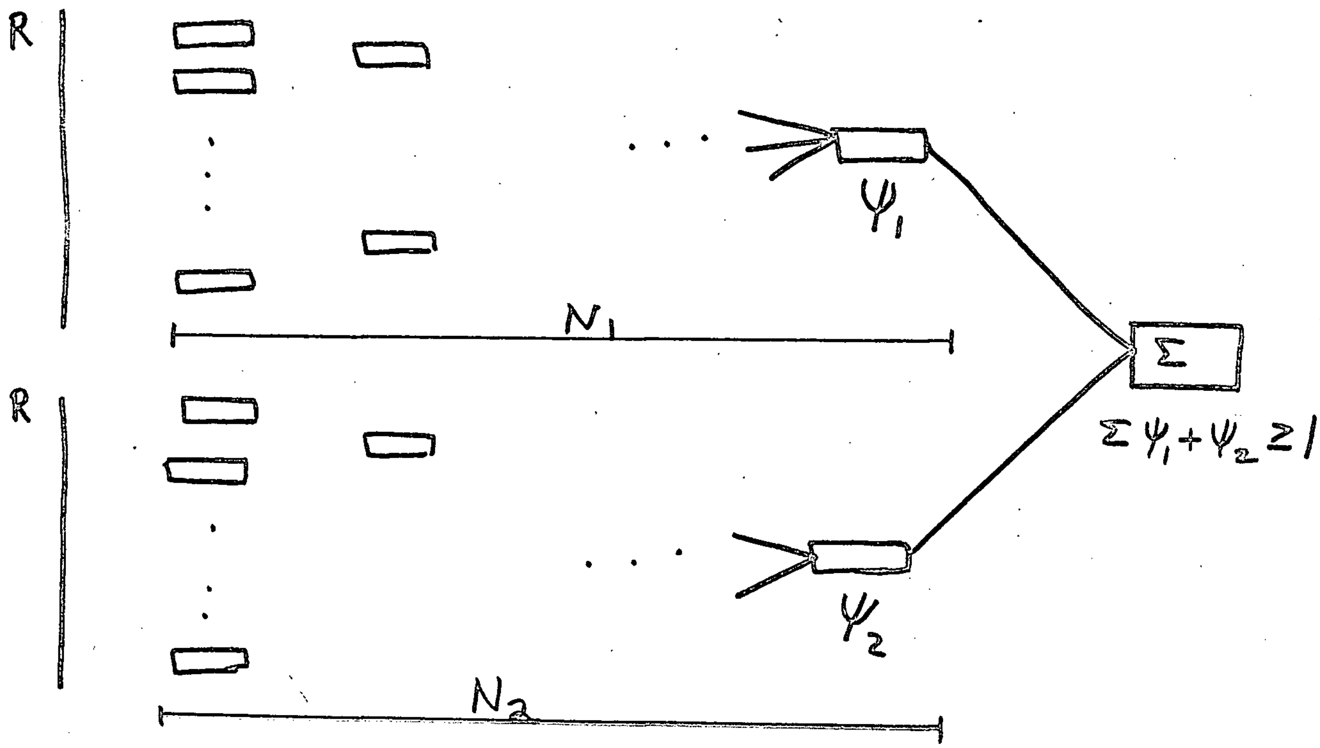
We might just as well mention that there is a direct proof which may be sketched as follows. (Incidently this proof would be typical of other machines endowed with too much power.)

Let R be a retina. Form the set S of all predicates ψ such that ψ is of order 2 with respect to an n-level perceptron for some n. We show that S = the set of all predicates on R by showing that

(1) S contains the masks of support 1 (trivial).

(2) $\psi \in S \Rightarrow \neg\psi \in S$. This may be proved by reversing inequalities or changing the sign of the coefficients.

(3) $\psi_1 \in S$ and $\psi_2 \in S \Rightarrow \psi_1 \vee \psi_2 \in S$. This fact may be proved considering the N_1 and N_2-level perceptrons P_1 and P_2 which recognize ψ_1 and ψ_2 respectively and constructing a new N_3 perceptron with $N_3 = \max\{N_1, N_2\} + 1$ in a manner suggested by the picture:

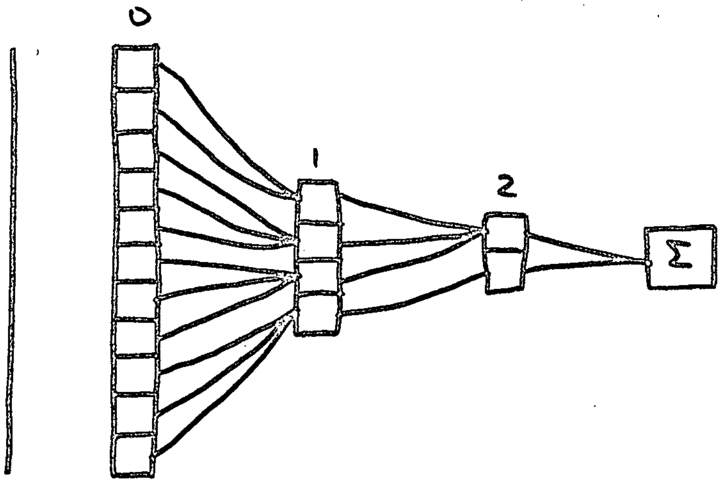


We will not go into details.

Corollary II: Any predicate schema Ψ is of order ≤ 2 with respect to the class of all finite level perceptrons.

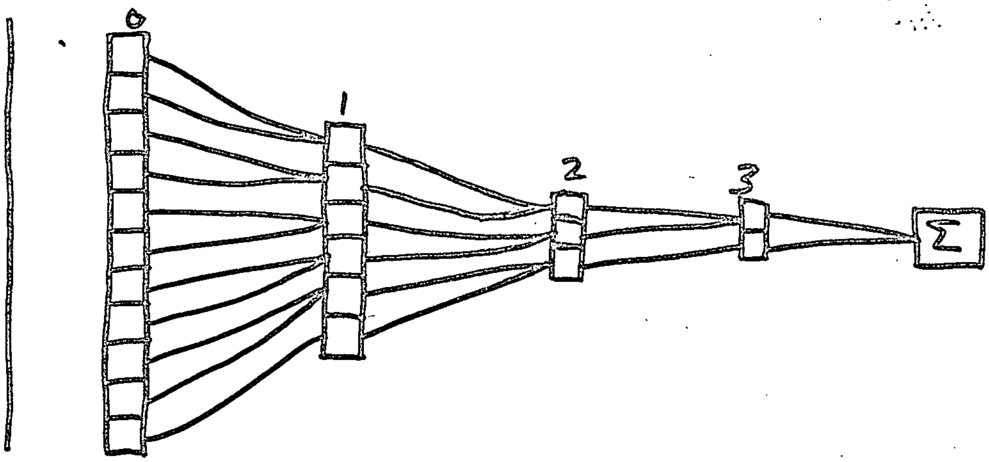
Proof: This statement is obvious given the definition of predicate schema (and the unstated definition of "order with respect to the class of all finite level perceptrons") since all ψ generated by Ψ are uniformly bounded in order by 2. We mention the "result" only to clear up lingering ambiguities.

Discussion: The proofs of the last two theorems seem formidable, mostly because of our strict adherance (for the time being) to the formalism and our liberal use of subscripts. Actually the idea behind the theorems is simple (or trivial). We merely convert our perceptrons into mask perceptrons and then break down the big masks by "pyramiding" smaller ones in the manner of the suggestive picture.



which converts a mask "of order" [incorrect usage] 11 into a "mask net" of order 3 and level 2. Our theorem says with 2 levels the worst order we could have expected was $[\sqrt[2]{11}\,] + 1 = [3.31] + 1 = 4$.

To obtain a mask net of order 2 we must use 3 levels.



The theorem says that for 3 levels our order will be $\leq [\sqrt[3]{n}\,] + 1 = [1.82] + 1 = 2$.

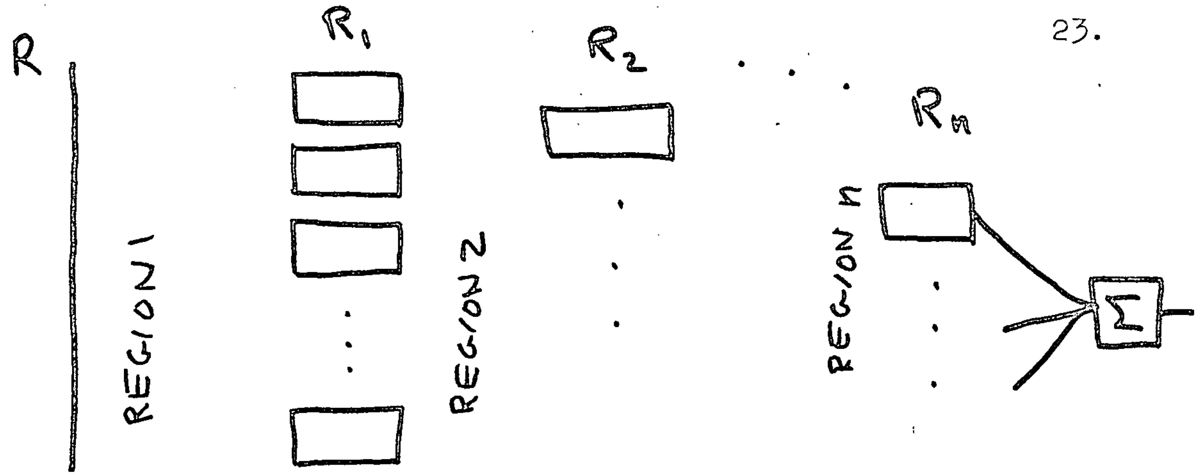
<u>Theorem</u>: If ψ is of order k with respect to an n- level perceptron then ψ has order $\leq k^n$ with respect to a 1-level perceptron,

Proof: (Sketch) Informally, for any $\varphi \in R_n$ we look at "supp φ" in each of $R_{n-1}, R_{n-1}, \ldots, R_o$. We can easily see that these supports will be $\leq k, k^2, \ldots, k^n$ in the various retina. To construct our 1-level perceptron we merely take for any $\varphi \in R_n$ a new $\tilde{\varphi}$ which is a predicate on R_o and depends on only the (possibly) k^n X_i's in R_o.

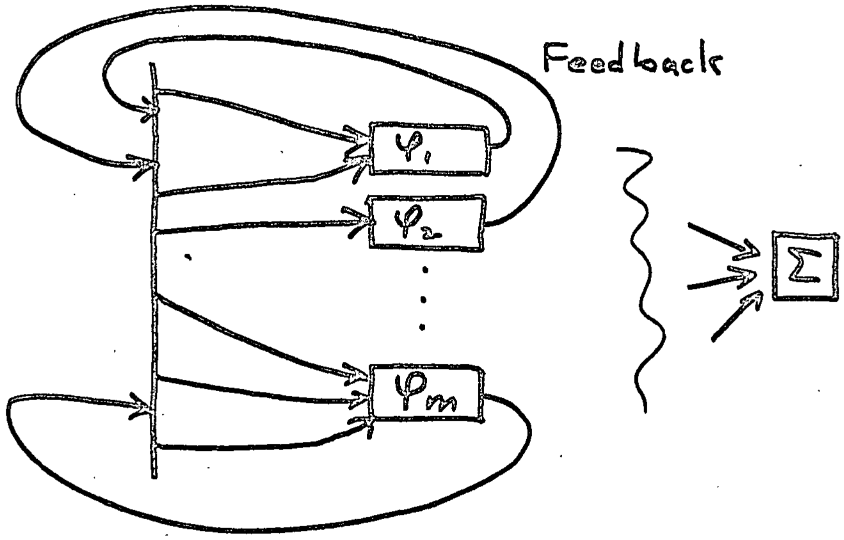
Remark: Our theorems have made the order problem rather uninteresting for the perceptrons with uniformly support restricted higher order associator units as well as casting serious doubt on the order measure itself for this class of machines. Specifically, the complexity of the predicate seems measured more by the interconnections of the associators than the size of associator support.

FEEDBACK PERCEPTRONS

Discussion: We now continue our search for stronger perceptrons, this time obtaining a machine which looks more like an automaton than the combinatorial nets we had before. We are motivated by a desire to minimize the amount of "wiring" in the machine so that order will play a larger role. Our idea is to eliminate wiring between the many levels by using the first level over and over again in a "feedback" arrangement. So whereas before we had a situation like this:



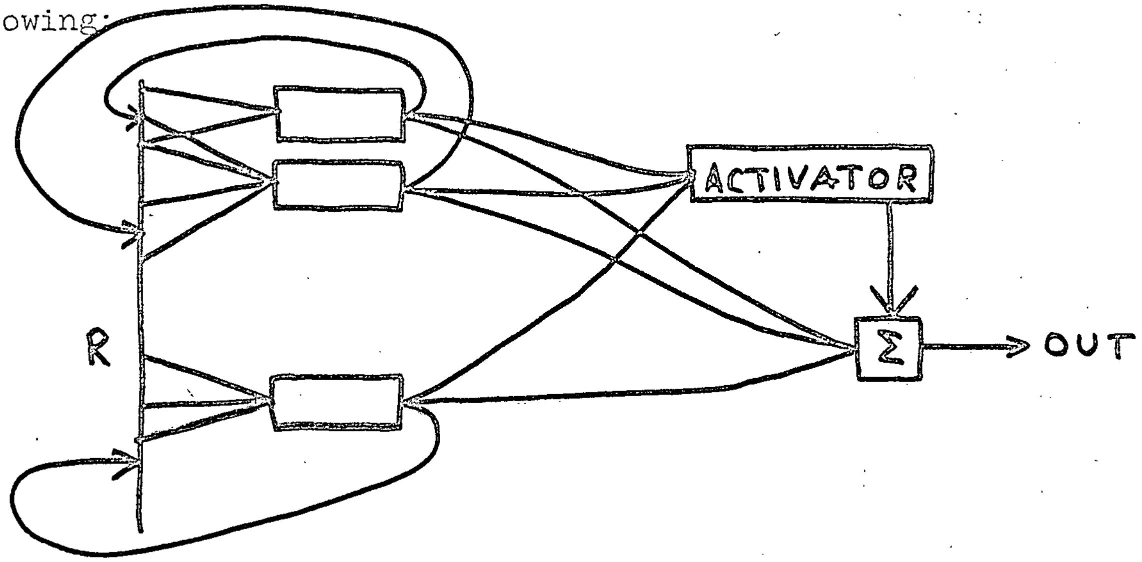
with potentially complicated wiring situations in regions 1 through n as marked, we now wire region 1 once and for all. We might alternatively think of this situation as requiring all regions 1 through n to be wired identically.



We could think of the state of the machine at $t = 0$ as $ST(0) = X \subseteq R$, at $t = 1$ as $ST(1) = \{\varphi(X) \mid \varphi(X) = 1\}$, at $t = 2$ $ST(2) = \{x \in R \mid \varphi \in ST(1)$ and x is at the end of a feedback arrow$\}$.

We have one problem left: We must tell the summation operator when to operate otherwise the machine will cycle endlessly, never giving an output. For this purpose we allow all $\varphi \in R_1$ to communicate into a common channel a "0" or a "1".

A "1" into the channel from φ_i means that φ_i is ready to proceed to the summation state; a zero means it is not. Only when there is unanimous consent does the summation take place. The machine then shuts down. Diagramatically then we have the following:



Definition: A feedback perceptron (F.P.)P is a triple $\langle R, R_1, A\rangle$ where the (rectangular) retina $R = \{x_i\}_{i \in I}$ is a set of points, indexed by a finite set I. R_1 is the set of associator units $\{\varphi_j\}_{j \in J}$ indexed by a finite set J, and $A = \{\alpha_j\}_{j \in J} \cup \{\theta\}$ is the set of coefficients α_j together with the threshold θ. The associator units φ_j are functions from $P(R)$ to $\{X, O\} \times \{0,1\} \times \{0,1\}$ where $X \in R$. (Think of this triple as indicating the following.

$\langle$YES;NO/reactivation of retina, YES;NO/proceed to summation,
YES;NO subpredicate true$\rangle$

Remark: Any F.P. P is a finite automaton. We define the initial state of the machine as $S_0 \leq R$. Inductively we define $S_{i+1} = \{x \in R \mid \exists\ \varphi_j \in R_1$ and $\varphi_j(S_i) = \langle x, 0$ or 1, 0 or $1\rangle\}$.

Define the <u>output</u> at time n by

$$\mathrm{OUT}(n) = 1 \quad \text{iff} \quad \varphi_j(S_n) = \langle x \text{ or } 0,1,0 \text{ or } 1\rangle \text{ for all } j \in J\ .$$

$$\text{and} \quad \sum_{j\in J} \alpha_j\, \hat{\varphi}_j(S_n) \geq \theta \quad \text{where } \wedge = \text{projection onto last co-ordinate.}$$

$$\mathrm{OUT}(n) = 0 \quad \text{iff} \quad \varphi_j(S_n) = \langle x \text{ or } 0,1,\ 0 \text{ or } 1\rangle \quad \text{for all} \quad j \in J$$

$$\text{and} \quad \sum_{j\in J} \alpha_j\, \hat{\varphi}(S_n) < \theta$$

$$\mathrm{OUT}(N) = \tfrac{1}{2} \quad \text{otherwise.}$$

<u>Definition</u>: We say that X is <u>accepted by the F.P.P</u> iff $\mathrm{OUT}(n) = 1$ where n is the least integer such that $\mathrm{OUT}(n) \neq \frac{1}{2}$ <u>and</u> the state of P (i.e. the activated subset of R) remains the same after $t = n$.

<u>Remark</u>: We might as well assume that our associators φ can reactivate whole subsets of R since we may add as many "redundant "φ's" as needed, connecting each feedback arrow with a different x_i on our subset.

ORDER OF PARITY PREDICATE FOR FEEDBACK PERCEPTRONS

Theorem: Parity is of order 2 with respect to a F.P.

Proof: See page 11 for a definition of ψ_{PAR}, the parity predicate. Enumerate R as $\{x_i\}$ $i = 1,2,\ldots,2n$. If I is odd a similar argument works. Enumerate R_1 as φ_j $j = 1,2,\ldots,n$ and define $\varphi_j(X) = \langle a,b,c \rangle$ where

$a = j/2$ if j even and x_{2_j} or $x_{2_{j-1}} \in X$ but not both.

$= j/2 + \frac{1}{2}$ if j odd and x_{2_j} or $x_{2_{j-1}} \in X$ but not both.

$= 0$ otherwise.

$b = 0$ iff $a = 0$ unless $j = 1$ then set $b = 0$ iff $x_1=1 \wedge x_2=1$

$c = 0$ iff only one of x_{2_j} and $x_{2_{j-1}} \in X$

$c = 1$ iff both or neither x_{2_j} and $x_{2_{j-1}} \in X$.

Let $A = \{\alpha_j\}$ where $\alpha_j = 1$ if $j = 1$ and $\alpha_j = 0$ otherwise. then

$$[X \text{ is even}] \Leftrightarrow \Sigma\, \alpha_j\, \varphi_j \geq 1 .$$

Clearly this machine operates by continually "trimming X down" (preserving parity) until only one or zero points are left. If one point is left, $|X|$ is odd, if none, $|X|$ is even. An example of this machine is simulated and the sequence of steps shown in Figs. 1, 2, 3 and 4.

To complete our proof we should also show that no "order

1" F.P. can recognize parity. Namely, the two parts of the proof would show that order (ψ_{PAR}) on F.P.s is (1) ≤ 2 and (2) > 1 . We show this fact after the next theorem in a section entitled "parity proof continued".

CANNONICAL OPERATIONS FOR FEEDBACK PERCEPTRONS

Remark: The last machine operated by reducing a figure to a more tractable form which it could then deal with. In mathematics we usually call this kind of process normalization or reduction to cannonical form. We now show that all (F.P)s operate in such a way.

Definition: A (F.P) P computes ψ by reducing figures to cannonical form iff (1) $\psi(X) = 1 \Longleftrightarrow$ the machine accepts X and (2) the sequence of states $S_o, S_1, \ldots, S_n$ before acceptance or rejection of any $X \subseteq R$ are such that $[\psi(S_i) = 1]$ for all i or $[\psi(S_i) = 0]$ for all i .

Theorem: All (F.P)'s operate by reducing figures to cannonical form.

Proof: Assume P is a (F.P) and does not so operate. Then there is a predicate ψ such that P accepts ψ and there exists $X \subseteq R$ such that the sequence of states before acceptance $S_o, \ldots, S_n$ contains S_k, $k < n$, with $\psi(S_k) = 0$. If we let

$S_k = R_o$, a new initial state, the sequence $R_o, R_1, \ldots,$ will coincide with $S_k, S_{k+1}, \ldots, S_n$ with S_n accepted, which means $\psi(S_k) = 1$ contradiction.

Discussion: Clearly any state of a (F.P) P on R can be described as leading to acceptance, rejection, or endless cycling. The initial state merely starts the path of arrows induced by the next state function which (in the absence of additional input) would normally lead to an endless cycle. The only event that can prevent such a cycling situation in this case is to hit on a state which activates the summation operation and tends to acceptance or rejection.

The problem then is to partition the set of states ($= p(R)$) of the finite automata P into $\mathfrak{A}$ and $\mathfrak{B}$ with $X \in \mathfrak{A} \Leftrightarrow \psi(X) = 1$ and $X \in \mathfrak{B} \Leftrightarrow \psi(X) = 0$. Normally such a problem is easy, but in this situation it is not since the next state function (and hence the arrow paths) are severly constrained in S by the order of P .

A full and satisfactory theory on this subject would require relating order constraints to the state graph of the automata and then to the class of predicates, a weighty task indeed! We will content ourselves with the problem of "programming" (F.P)'s to handle various predicates such as connectivity.

Parity Proof Continued: Consider any order 1 F.P. which recognizes ψ_{PAR} . A consequence of Minsky's group invariance theorem is that the only predicates of order 1 with respect to a 1-level perceptron are of the form $\varphi(X) \Leftrightarrow |X| \leq M$ or $\varphi(X) \Leftrightarrow |X| > M$. We conclude that the blackened subsets of R (i.e. the sequence of states of MP) must tend in cardinality to an odd number $> M$ for some $M \leq |R|$ for rejection and an even number $\leq M$ for acceptance in any order 1 parity recognizer. Of course, the states could also tend to any odd number $\leq M$ for rejection and any even $> M$ for acceptance.

We now consider the following small example. Let R be a 1x2 retina. We examine the behaviour of all possible associator units to show the impossibility of recognizing the parity predicate with an order 1 F.P.

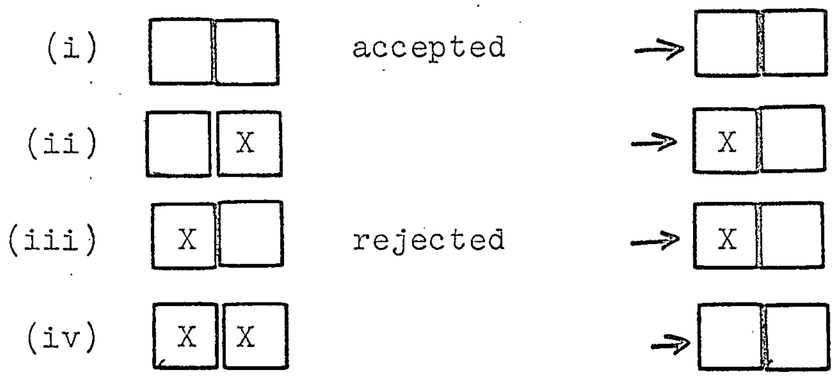
$$R = \boxed{\varphi_1}\boxed{\varphi_2}$$

φ_1 is the associator on the left square x_1 of R .
φ_2 is the associator on the fight square x_2 of R .

We break the proof down into cases where the cannonical form for "accept" is $\boxed{}\boxed{}$ (Case I) or $\boxed{X}\boxed{X}$ (Case II). We further break things down depending upon whether the cannonical form for "reject is $\boxed{X}\boxed{}$ (Case a) or $\boxed{}\boxed{X}$ (Case b) or both (Case c) .

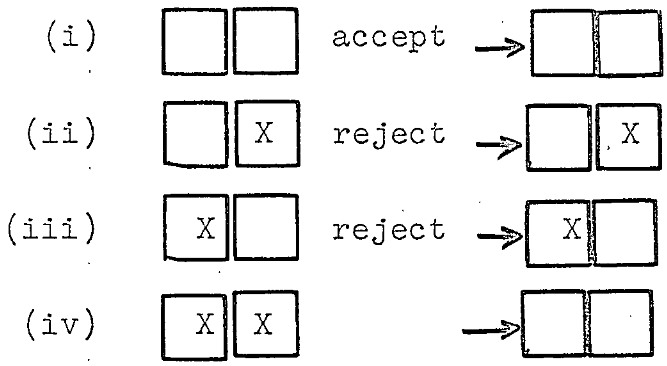
Case Ia: if " $\rightarrow$ " is read as "leads to the state" we can symbolize the required transformations in this case in the following way. (Remember: we must prevent endless cycling on non-

(accept or reject) states.



Now (i) shows that $\varphi_1(\bar{X}_1) = \langle\emptyset, 1 \text{ or } 0, 1 \text{ or } 0\rangle$. That is upon being presented with a blank φ_1 reactivates nothing. Clearly φ_2 has the same property. Next (iv) shows φ_1 and φ_2 reactivate nothing upon "seeing" an active square. Therefore φ_1 and φ_2 cannot reactivate anything so parts (ii) and (iii) are impossible, contradicting the assumption that Case Ia is possible. We briefly run through the other cases (omitting Ib and IIb for reasons of symmetry) showing that they are all impossible.

<u>Case Ic</u>: The required transformations are:



As before (i) and (iv) show φ_1 and φ_2 are inert, showing (ii) and (iii) impossible.

<u>Case IIa</u>: The required transformations are:

$$
\begin{array}{llll}
\text{(i)} & \boxed{}\boxed{} & & \longrightarrow \boxed{X}\boxed{X} \\
\text{(ii)} & \boxed{}\boxed{X} & & \longrightarrow \boxed{X}\boxed{} \\
\text{(iii)} & \boxed{X}\boxed{} & \text{reject} & \longrightarrow \boxed{X}\boxed{} \\
\text{(iv)} & \boxed{X}\boxed{X} & \text{accept} & \longrightarrow \boxed{X}\boxed{X}
\end{array}
$$

Items (ii) or (iii) show that neither φ_1 or φ_2 will reactivate both X_1 and X_2 together. And (i) further shows that both must activate at least one of X_1 or X_2, given a blank. Therefore one of the following possibilities obtains:

(1) $\varphi_1(0) = <X_1,-,->$ $\qquad$ $\varphi_2(0) = <X_1,-,->$

(2) $\varphi_1(0) = <X_1,-,->$ $\qquad$ $\varphi_2(0) = <X_2,-,->$

(3) $\varphi_1(0) = <X_2,-,->$ $\qquad$ $\varphi_2(0) = <X_1,-,->$

(4) $\varphi_1(0) = <X_2,-,->$ $\qquad$ $\varphi_2(0) = <X_2,-,->$

Numbers (1) and (4) can be rejected out of hand as the run counter to (i). Number (2) is contrary to (iii) and (3) is contrary to (ii).

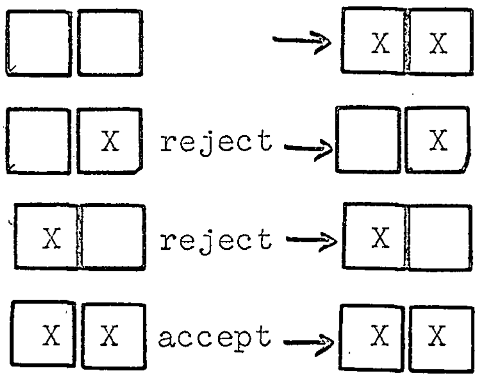
Case IIc:

(i)

(ii)

(iii)

(iv)



Item (i) shows that $\varphi_1(0) = <-,0,->$ or $\varphi_2(0) = <-,0,->$. Thus either (ii) or (iii) are not reject states. Contradiction.

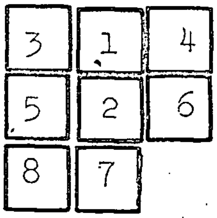
Aside: It is interesting to note that alternative (3) in Case IIa is a consistant assignment for Case IIc. Also: If the last clause in the definition of "accept" (i.e. the state stays the same) were not required we would NOT have been able to derive a contradiction in this 1×2 example. Furthermore even for larger examples, I do not know whether any contradiction would exist. The number of possibilities for associator action is very large.

Having eliminated all possible cases, we have shown that an order 1 parity recognizer is not possible for F.P.'s

ORDER OF CONNECTIVITY PREDICATE FOR FEEDBACK PERCEPTRONS

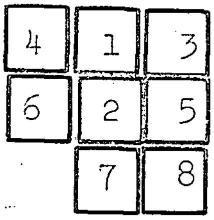
Preview: We devote the next section to showing that ψ_{CONN} is of order ≤ 8. That is we find F.P., M_c which will recognize ψ_{CONN} with no associator of M_c having support > 8. M_c is simulated, as is M_p at the end of this paper. (See p. 11 for definition of ψ_{CONN}).

Description of M_c: M_c will operate on an $m \times n$ retina R and will be equipped with six kinds of associator units, the Right Dribbler, the Left Dribbler, the Surrounder, the Sidewinder, the Backslider and the Scalawag as pictured below.



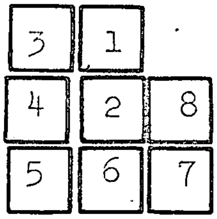
Right Dribbler

(RD)



Left Dribbler

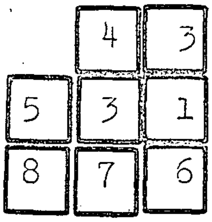
(LD)



Surrounder

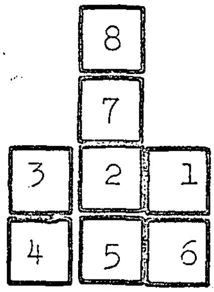
(SR)

(DR = RD or LD)



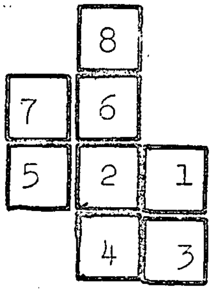
Sidewinder

(SW)



Backslider

(BS)



Scalawag

(SC)

The arrows show which way active squares tend to propogate (as shown below) under the various kinds of units.

At each square $x_i \in R$, one of each kind of unit (six in all) is placed upright in such a way that square 2, the <u>centre of the unit</u> coincides with x_i on the retina. The associators may overlap the edge of the retina in which case such overlapped squares are always registered as "blank" or "inactive".

We must now describe the output (i.e. the triple) of each unit as well as the summation operator Σ.

<u>First Component:</u> Each unit will have only two outputs, the <u>"growth" or "activating"</u> output and the <u>"standpat" or "recopying"</u> output. The standpat output merely copies the existing input with no additions or subtractions so that no information will be lost during the time interval. The growth output copies the existing input and adds to it, if possible, square 2. Roughly speaking, most outputs will tend to standpat; only a few are growth outputs. We give for each kind of unit the necessary and sufficient conditions that the output be a growth output.

<u>Both Dribblers:</u> All four conditions must be met.

(1) Square 1 is active. Square 2 is not. i.e. $1^+, 2^0$

(2) Square 5 is active implies Square 3 is. i.e. $5^+ \rightarrow 3^+$

(3) $6^+ \rightarrow 4^+$

(4) $7^+ \rightarrow (3^+ \text{ and } 5^+ \text{ and } 8^+)$ or more simply $7^+ \rightarrow (3,5,8)^+$

Surrounder: $(1,3,4,5,6,7,8)^{+}\ 2^{0}$

Sidewinder: $(3,1,6,7,8)^{+}\ 2^{0}$

Backslider: $(3,4,5,6,1)^{+}\ (2,7,8)^{0}$

Scalawag: Both conditions must be met:

(1) $(8,6,7,5,2)^{0}\ 1^{+}$

(2) $4^{+} \rightarrow 3^{+}$

Second Component: For all units the second component (which says whether or not the unit is ready to sum) is a 0 iff the unit's first component was a growth output and (of course) a 1 iff the unit's first component was a standpat output. Therefore M_c proceeds to summation only after all units standpat, which, of course, implies that the machine state has stabilized.

Third Component: We define the third component (which says whether or not the unit accepts the small bit it sees) in a way which will become clear only after a proof. The right dribbler puts out a 0 iff

(1) $1^{+}\ 2^{0}$ OR (2) $1^{+}\ 3^{0}$.

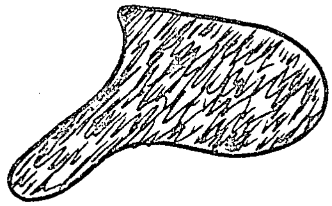
All other units always put out a one.

Summation Operator: M_c accepts X iff M_c is ready to sum upon being given X and $\Sigma\ (1 - \varphi(X)) \leq 0$; φ an associator unit of M_c.

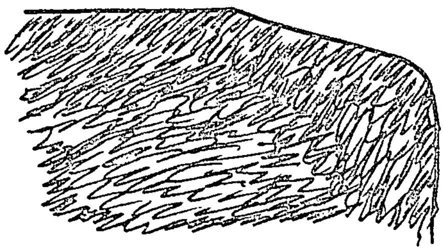
That is, acceptance or rejection is on a "blackball" system, X being rejected iff any one of the right dribblers puts out an 0 in its third component.

Definition: A square $x \in R$ is said to be <u>in the scope of</u> <u>$x_o \in R$</u> (written $x \in$ scope (x_1)) iff $x = x_o$ OR x is below x_o in x_o's column OR x is to the left of x_o in x_o's row OR x is down and to the left of x_o .

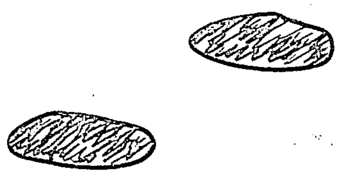
Remark: Informally the way M_c operates is to take any $x \in X$ of a connected figure and fill in its scope, if it has not been done already. Thus a figure starting out like this:



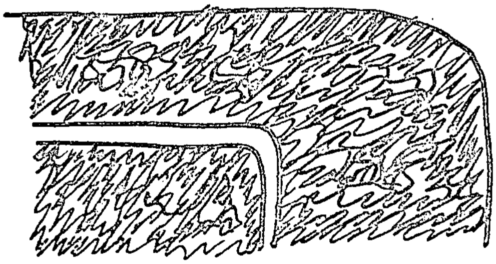
would end up looking like this at the time of acceptance.



Really then the third component of the RD is merely checking to see if all scopes have been filled in on a processed and (supposedly connected) figure. Disconnected figures like this



will end up like this at the time of rejection.



This fact is very clearly illustrated after the text in the figures.

Notation: Let a rectilinear path γ in R going from x to y be denoted as $\gamma = \{x = x_o, x_1, x_2, \ldots, x_{n-1}, x_n = y\}$ where x_i and x_{i+1} have a side in common for $i = 0, 1, \ldots, n-1$.

Lemma 1: If $x_o \in X$ and $y_o \in X$ are not connected to each other then no single associator φ will connect them directly. (i.e. in one time step.)

Proof: Suppose x_o and y_o are not connected. In order that φ connect x_o to y_o directly, y_o must be at one of the squares marked 1,2,3,4,5,6,7 or 8 in relation to x_o .

	11	2	10	
12	3	9	1	9
4	b	x_o	d	8
13	5	c	7	14
		6		

We eliminate all possibilities, first eliminating 5,6,7 and 8 by symmetry. (For example: If y_o is at 7 the picture looks the same as if it were at 3 , etc.)

Case I: y_o is at 4 and b must have been activated. We look at the six associators centered at b . A right or left dribbler at b would imply a previous connection from 4 to x_o via $\{4, 12, 3, a, x_o\}$ to activate b . A backslider would connect 4 to x_o via $\{4, 13, 5, c, x_o\}$ if it were active as would a sidewinder and a surrounder. A scalawag could NOT activate b under any circumstances.

Case II: y_o is at 3 . Either a or b could have been activated.

Sub-case IIa: y is at 3 and a was activated. We examine the associators centered at a . A dribbler (right or left) activating a would imply a previous connection from x_o to 3 via $\{3, b, x_o\}$ or $\{x_o, d, 1, 10, 2, 11, 3\}$. A backslider at a implies a previous connection via $\{3, b, x_o\}$ as does a sidewinder and a surrounder. A scalawag cannot activate a .

Sub-case IIb: y is at 3 and b was activated. Either DR implies a connection via $\{3, a, x_o\}$. A BS could not do the job nor could a SC. A SW implies a connection via $\{3, a, x_o\}$. A SR connects via $\{3, 12, 4, 13, 5, c, x_o\}$.

Case III: y is at 2 and a was activated. Active dribblers at a would imply a connection $\{2, 11, 3, b, x_o\}$ as would a SR or $\{2, 10, 1, d, x_o\}$. Neither a BS or a SC would work. An active SW at a implies a connection $\{2, 10, 1, d, x_o\}$.

Case IV: y is at 1 and either a or d was activated.

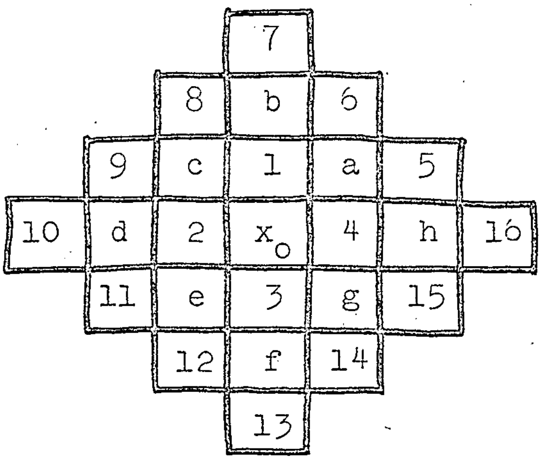
Sub-case IVa: DR's give a path $\{x_o, b, 3, 11, 2, 10, 1\}$ or $\{x_o, d, 1\}$. A BS, SW, SR, or SC yields the path $\{x_o, d, 1\}$.

Sub-case IVd: A DR activating d connects via $\{x_o, a, 1\}$ as does a SR. A SC or BS would not do the job. A SW connects via $\{x_o, 6, 7, 14, 8, 9, 1\}$.

All possibilities having being eliminated, the lemma is shown.

Lemma 2: Suppose x_o and y_o are not connected. Then no two associators φ_1 and φ_2 will connect them directly.

Proof: Suppose x_o and y_o are not connected. In order that φ_1 and φ_2 connect x_o to y_o directly, y_o must be at one of the squares marked 5,6,7,8,9,10,11,12,13,14,15 or 16, in relation to x_o. For if y_o were at 1,2,3, or 4 we would have a connection.



If y_o were outside the assigned area two associators could not possibly connect them directly. If y_o were at a,b,c,d,e,f,g, or h and were connected via $\{x_o, x_1, x_2, x_3 = y_o\}$ to x_o either x_1 or x_2 would have to be at 1,2,3, or 4, forming a direct connection using only one associator's activated square, contrary to Lemma 1.

As before we eliminate all possibilites, eliminating first 11,12,13,14,15 and 16 by symmetry.

<u>Case I</u>: y_o is at 5. There are 3 direct paths from 5 to x_o.

(i) $\{5, a, 1, x_o\}$ (ii) $\{5, a, 4, x_o\}$

(iii) $\{5, h, 4, x_o\}$

<u>Case II</u>: y_o is at 6. There are 3 direct paths from 6 to x_o.

(i) $\{6, b, 1, x_o\}$ (ii) $\{6, a, 1, x_o\}$

(iii) $\{6, a, 4, x_o\}$.

<u>Case III</u>: y_o is at 7. There is only one direct path $\{7, b, 1, x_o\}$ from 7 to x_o.

<u>Case IV</u>: y_o is at 8. There are 3 direct paths from 8 to x_o.

(i) $\{8, b, 1, x_o\}$ (ii) $\{8, c, 1, x_o\}$

(iii) $\{8, c, 2, x_o\}$.

<u>Case V</u>: y_o is at 9. There are 3 direct paths from 9 to x_o.

(i) $\{9, c, 1, x_o\}$ (ii) $\{9, c, 2, x_o\}$

(iii) $\{9, d, 2, x_o\}$.

Case VI: y_o is at 10. There is only one direct path $\{10, d, 2, x_o\}$ from 10 to x_o .

Note: We will shorten this long, tedious proof somewhat by the following observations and notational conveniences.

(i) When we argue a case we might as well assume that both squares in the path in question were initially blank, and hence needed activation. Take, for example, Case I(iii). We argue in this way: y_o is at 5 and was connected in one time step to x_o via the path $\{5, h, 4, x_o\}$. Assume to the contrary that either h or 4 was previously active. We derive contradictions in both cases. Case h: h was previously active. But then h was not connected to x_o since 5 wasn't. So h and x_o (two disconnected points) were connected in one time step by two associators via the square 4 alone. Therefore h and x_o were connected in one time step by one associator. This statement violates Lemma 1. Case 4 uses the same idea.

(ii) To avoid needless verbiage, our case arguments will merely list the type of associator alledged to have activated a path square, followed by a path or a statement of impossibility. This path will represent a pre-existing connection between x_o and y_o whose existence is implied by the active associator. For example, in Case I(iii) again we will write (in part)

BS at 4 → x_o, 3, g, 15, h, 5

SC at 4 → DNW

These cryptic lines mean that "in order for a BS to activate square 4 a path must have existed before connecting x_o to 5 via $\{x_o, 3, g, 15, h, 5\}$; contrary to assumption". Also we have "A SC Does Not Work, i.e. it would not activate 4 in this case". First x_o inhibits the SC. Second h is blank also inhibiting the SC. Note that we could also have written BS at 4 → DNW since h is blank. We will not be fussy about where our contradictions come from.

I(i):

	7		
8	b	6	*
c	1	a	5
2	x_o	4	h
	3	g	15

LD, BS, SW, SR or SC at 1 → x_o, 4, a, 5 . So we assume 1 was activated by an RD. In this case BS, SW or SR at a → x_o,4,h,5 SC at a → DNW since b was active. DR at a → x_o,2,c,8,b,6,*,5. Therefore 1 was previously active, contradiction to Lemma 1.

I(ii),(iii): SW, BS or SR at 4 → x_o, 3, g, 15, h, 5 . SC at 4 → DNW . DR at 4 → x_o,1,a,5.

II(i)(ii): LD, BS, SC, SW or SR at 1 → x_o, 4, a, 6 . RD at 1 → x_o, 2, c, 8, b, 6.

II(iii): SR, DR at 4 → x_o, 1, a, 6 . BS, SC at 4 → DNW . So a SW at 4 is the only possibility. Assume a SW at 4 . DR, SW at a → x_o, 3, g, 15, h, 5, *, 6 . SC, BS at a → DNW . SR at a → x_o, 1, b, 6 .

III:

⁑	7	*
8	b	6
c	1	a
2	x_o	4

DR at 1 → x_o, 2, c, 8, b, 7 or x_o, 4, a, 6, b, 7 . BS, SC, SR at 1 → DNW . So SW at 1 is the only possibility. Assume an SW at 1 . SC, SW, BS, SR at b → DNW . DR at b → x_o, 4, a, 6, *, 7 .

III(i)(ii):

	⁑	7	ǂ
*	8	b	6
9	c	1	a
d	2	x_o	4
11	e	3	

DR at 1 → 8,c,2,x_o, or 8,b,6,a,4,x_o . SR, BS at 1 → 8,c,2,x_o SC at 1 → DNW . The only alternatice is a SW . Assume a SW at 1 . We must now examine the units at b and at c . DR at b → x_o,4,a,6,ǂ,7,⁑,8. BS, SW at b → x_o,1,c,8 . SC at b → DNW . SR at b → x_o,1,c,8 . Now for c . DR at c → 8,b,1,x_o or 8,*,9,d,2,x_o . BS, SC or SR at c → DNW . SW at c → x_o,1,a,6,b,8. Contradiction in both cases.

IV(iii): SC, BS at 2 → DNW. SR at 2 → x_o,3,e,11,d,9,c,8 . DR at 2 → 8,c,1,x_o . So SW at 2 is the only possibility. Assume SW at 2 . DR, SW at c → x_o,1,b,8 . BS, SC, SR at c → DNW .

V(ii)(iii):

	*	8	b	6
⁑	9	c	1	a
10	d	2	x_o	4
ǂ	11	e	3	

DR at 2 → 9,c,1,x_o . SR, BS at 2 → x_o,3,e,11,d,9 . SC at 2 → DNW . So SW at 2 is assumed. We examine d and c . DR at c → 9,*,8,b,1,x_o . SW, BS at c → x_o,2,d,9 . SC SR at c → DNW . Now for d : DR at d → x_o,3,e,11,ǂ,10,⁑,9 or x_o,2,c,9 ∴SW at d → x_o,2,c,9 . SR, BS, SC at d → DNW .

V(i): SR, BS, SW at c → 9,d,2,x_o. SC at c → DNW . The only possibility then is a DR at c , meaning that 8 and * are active. We examine square 1 . DR at 1 → x_o,4,a,6,b,8,*,9 or x_o,2,c,9 . SC, SR at 1 → DNW . BS at 1 → x_o,2,c,9 . The only possibility is an SW at 1 meaning that 2 was originally active when the DR at c was applied. This yields x_o,2,d,9 or x_o,1,b,8,*,9 .

VI:

*	9	c	1
10	d	2	x_o
⁑	11	e	3

BS, SW at d → 10,⁑,11,e,2,x_o . SC, DR at d → DNW . The only possibility left is a DR ; hence $*$ and 9 were previously active. We examine square 2 . SC, SR at 2 → DNW . DR at 2 → x_o,1,c,9,$*$,10. BS at 2 → x_o,3,e,11,d,10 . So an SW at 2 is the only possibility. Hence 11 was previously active when the DR at d was applied. This yields x_o,3,e,11,$\overset{*}{*}$,10 or x_o,2,c,9,$*$,10 .

Q.E.D.

Remark: This proof was very long and boring! It was put in for completeness and to ease nagging doubts. The proof might have been shortened somewhat by ingenious assumptions, however it is very easy to get fooled! The brute force method though boring is at least clear.

Lemma 3: The state transitions of M_c go only from connected states to connected states or from disconnected states to disconnected states. They never switch back and forth.

Proof: (i) Let X be a connected figure in R and let $X^1 \supseteq X$ be the figure formed from X by an application of the state transition function of M_c . We show that X^1 is connected. Let $x^1, y^1 \in X^1$. We show x^1 is connected to y^1 . If both x^1 and $y^1 \in X$ we are done. If either x^1 or $y^1 \in X^1 \sim X$ it must have been produced by one of the six kinds of associator units with x^1 or y^1 in position 2 . But in such a case x^1 or y^1 (or both) would be connected to associator position 1, denoted by x and y respectively, which is in X . A path from x^1 and y^1 could then be of the form

$\gamma = \{x^1 = x_o, x_1, x_2, x_3, \ldots, x_{n-1}, x_n = y^1\}$ where $x_1, \ldots, x_{n-1} \in X$ and either $x_1 = x$ or $x_{n-1} = y$.

(ii) Let X be a disconnected figure and let $X^1 \supseteq X$ be the figure obtained from X by an application of the state transition of M_c . We show X^1 is disconnected by assuming it is connected and deriving a contradiction. Suppose X^1 is connected and X is disconnected. Let x and y be in separate components A and B respectively of X . There is a path $\gamma = \{x^1 = x_o, \ldots, x_n = y^1\}$ connecting x to y in X^1 . Every $x_i \in \gamma$ is classed either as (1) an element of A or (2) is derived from some element of A by an associator unit centered at x_i or (3) is an element of some other component than A (say C) or X or (4) is derived only from elements of other components by associators centered at x_i . Let the first two alternatives be of type I and the second two of type II . Since the path γ starts in type I and ends in type II there must be an i_o such that x_{i_o} is the last $x_i \in \gamma$ of type I. If x_{i_o} is of class (1) x_{i_o+1} cannot be of class (3) (C and A are disconnected) or of class (4) without violating Lemma 1. If x_{i_o} is of class (2) x_{i_o+1} cannot be of class (3) or (4) without violating Lemmas 1 and 2 respectively.

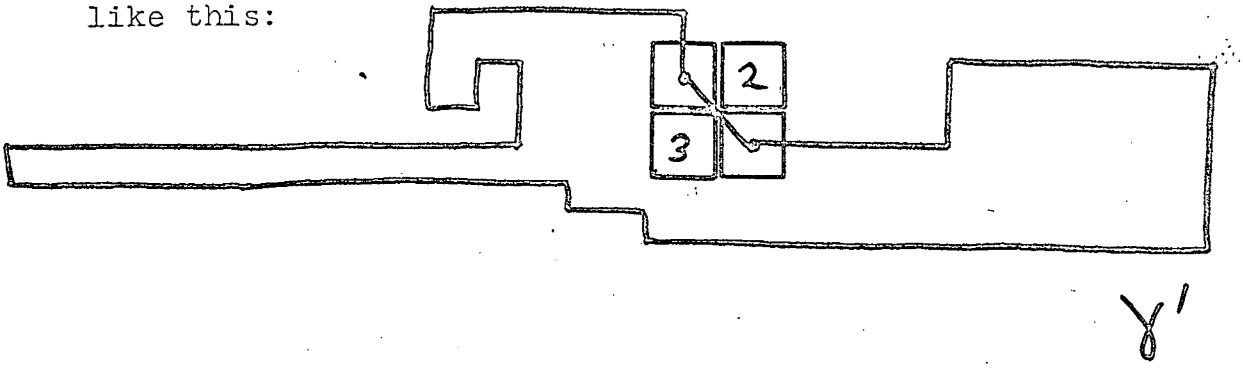
<u>Lemma 4</u>: If R is a retina which contains a connected figure and if M_c is ready for summation then the following configuration cannot occur on R

X	
	X

Proof: Assume $X \subseteq R$ is connected, that M_c is ready for summation (i.e. no associators will be active) and that

1	2
3	4

occurs on R where 1 and 4 are active, 2 and 3 are blank. There is a simple path γ connecting 1 to 4 not passing through 2 or 3. $\gamma = \{1 = x_0, x_1, x_2, \ldots, x_{n-1}, x_n = 4\}$. Form the unquantized simple closed curve γ^1 by connecting the centre of 1 to centre (2) and centre (x_i) to centre (x_{i+1}) for $i = 0,\ldots,n-1$ with straight line segments. Our picture looks like this:

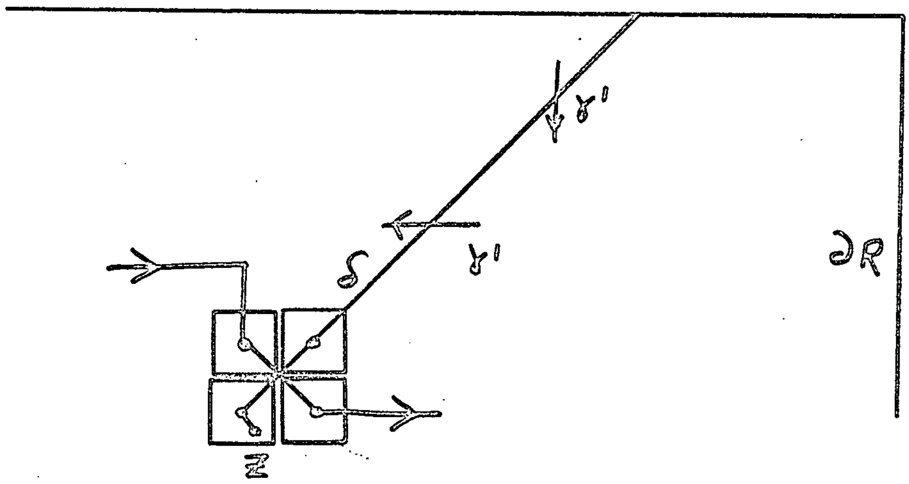


Definition: We say <u>γ semi-encloses a square x</u> iff γ is of the form given above and γ^1 encloses x in the usual sense of the Jordan Curve Theorem.

Claim: If 2 and 3 are blank then either γ semi-encloses 2 or γ semi-encloses 3 but not both.

Proof: Let z be a point (not a square) inside 3. Connect

z to centre 3 by a line segment and then draw a 45° ray from centre (3) upward out to ∂R. This new path δ will intersect γ^1 either an odd or an even number of times. If odd, 3 is semi-enclosed by γ; if even 3 is not semi-enclosed by γ.



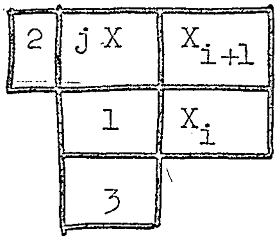
Clearly also since γ^1 passes through δ at the "centre" we have: γ semi-encloses $3 \rightarrow \gamma$ does not semi-enclose 2 and γ semi-encloses $2 \rightarrow \gamma$ does not semi-enclose 3.

We resume the proof of Lemma 1. Pick a γ connecting 1 and 4 semi-enclosing <u>minimal</u> area. Pick a highest row in γ. There will then be a local area represented as

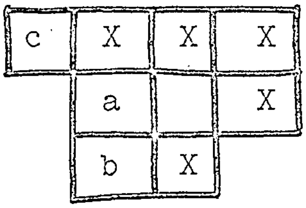
X_{i+1}	highest row .
X_i	

The path may then turn either to the left or to the right.

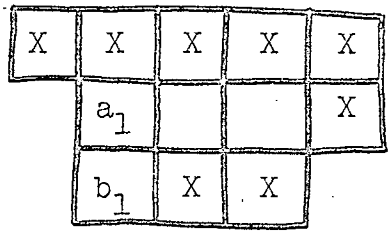
<u>Case L</u>: γ turned to the left in this particular spot.



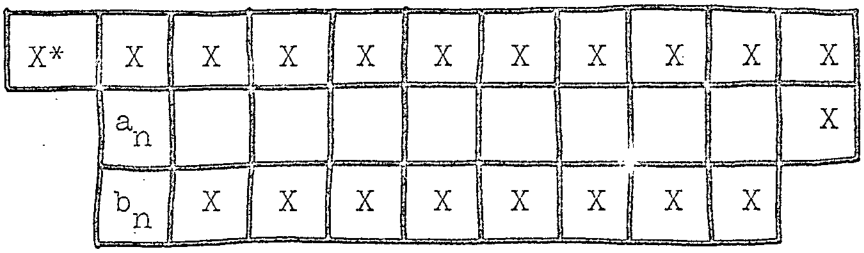
Now 1 must be blank since 1 is semi-enclosed by γ and γ is a minimal semi-encloser. If 1 is active, the path η obtained by substituting 1 for X_{i+1} in γ will semi-enclose less area than γ while still connecting our original two squares. So 2 is active, avoiding a dead end. And 3 is active by the dribbler rule and the fact that 1 is blank. We have



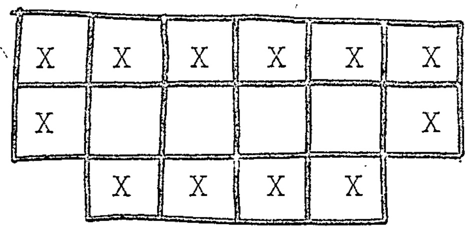
We investigate the various possibilities for a and b . a and b cannot both be blank because c would be active and the DR rule would be violated. a and b cannot both be active. To have such a situation would again violate the DR rule. We are left with two cases. First a is blank and b is active. Second b is blank and a is active. In the first case we have:

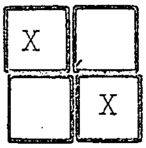


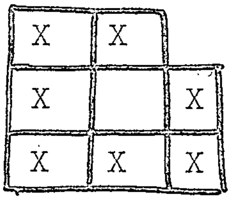
where again the only possible assignments for a_1 and b_1 are (a_1 blank, b_1 active) or (a_1 active, b_1 blank). Clearly this first alternative could not continue indefinitely to the exclusion of the second,



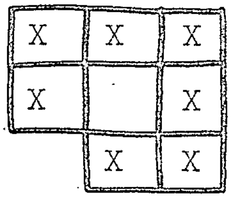
because X* is not one of our end squares. The only possibility left is that at some time the second alternative must occur yielding:



which gives us our original configuration - only now we must be able to connect  with a new single path γ_1, totally semi-enclosed by γ (or on γ). In this way we obtain a sequence of paths $\gamma_1, \gamma_2, \ldots, \gamma_m$ connecting our figure and semi-enclosing a <u>strictly decreasing amount</u> of area. Eventually we get to the point where the absolute minimum area is semi-enclosed which means γ looks like



or



whence the blank in the middle is filled in by a SR or a LD .

Contradiction.

Case R: The path turned to the right. Here we have as before:

X	X	X
X		a
c	X	b

where c blank yields a new path γ_1, totally enclosed in γ. If c is active we get a contradiction with the RD rule.

Lemma 5: If R is a retina which contains a connected figure and if M_c is ready for summation, then the configuration

	X
X	

cannot occur on R .

Proof: For this lemma, the proof goes as before except that we end up with minimal area semi-enclosing configurations like

	X	X
X		X
X	X	X

or

X	X	X
X		X
X	X	

violating the SW and RD rules.

Lemma 6: If R is a retina which contains a connected figure and if M_c is ready for summation then

X
X

cannot occur on R .

<u>Proof</u>:

3	X	6
1		2
4	X	5

It is easy to see that 1 and 2 must be blank in this case. We assume that 2 (for example) is active as the argument for 1 is symmetric. If 2 is active then we have either 6 blank or 5 blank or (3 blank and 1 active) all situations leading to the forbidden patterns of the last lemmas.

We argue now as before picking the highest row of a minimal area semi-enclosing path γ connecting our original figure. (note: we have abused the definition slightly.) We obtain the figures

X	X_1	X	top row
X		a	
c	X_2	b	

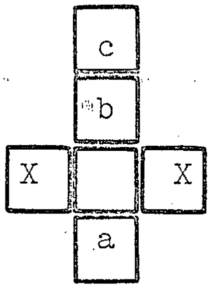
and

X	X_1	X	top row
a		X	
b	X_2	c	

depending upon whether γ "turned right or left". In both cases there is no consistent choice for c . Note that our first remark in the proof was necessary to prevent the path from ending or beginning at x_1 or x_2 .

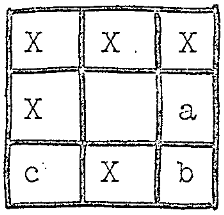
<u>Lemma 7</u>: If R is a retina containing a connected figure and if M_c is ready for summation, then $\boxed{X}\boxed{}\boxed{X}$ cannot occur on R .

Proof: (sketch)

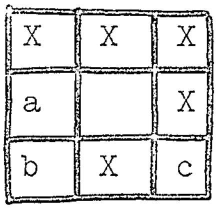


We can show by process of elimination that a, b and c must be blank. Then we plan to use the same type of argument as before. But we must be careful. This time the path γ may have as highest row only the end points. In such a case we cannot build such a strong configuration around our highest row as before.

Case A: γ extends above the end points: Here as before we get



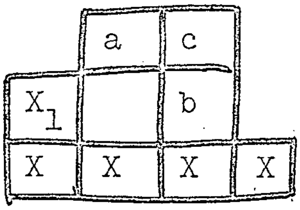
or



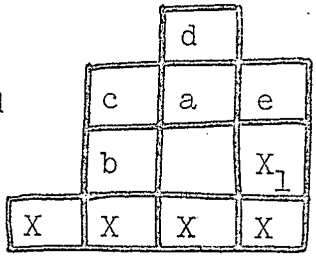
leading to contradictions.

Case B: γ does not extend above the endpoints: Here we must look at the lowest row for the γ semi-enclosing least area.

moved right

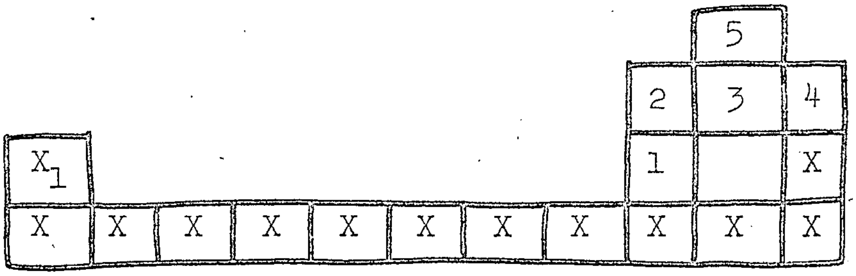


moved left



By Lemma 6 a is blank. By our comment at the beginning of the proof b is blank, hence c is too. If γ has moved to the left we could further conclude that e is blank and d is active. (Using the SW and SC rules respectively.) But then the RD and LD at a would be active contradiction.

If γ had moved to the right, it would eventually have to turn back up again to meet an end point,



duplicating the situation of the left moving path.

We conclude that there can be no path γ of the required type connecting the original figure. Contradiction.

Theorem: Upon receiving a connected figure X on an $m \times n$ retina R, M_c will continue to add retina cells to X until a "solid" figure X^o is produced. This figure $X^o = \bigcup_{x \in X} \text{scope}(x)$ After X^o is attained, no state change will take place.

Proof: Suppose X is a connected figure and has been inserted into M_c. It is clear from the form of the six associator units that the state at time t, $X_{(t)}$ satisfies

$$X \subseteq X_{(t)} \subseteq \bigcup_{x \in X} \text{scope }(x) .$$

It is also clear that if there is a time t_o with $X_{(t_o)} = X^o$ no state change will take place after t_o. Our task then is to show that if M_c sums at some time t_1, then we have $X_{(t_1)} = X^o = \bigcup_{x \in X} \text{scope}(x)$. We do so by contradiction.

Assume that at time $t = t_1$ M_c sums and that $X_{(t_1)} \not\supseteq \bigcup_{x \in X} \text{scope}(x)$. Therefore there is an $x_1 \in X$ such that $y \in \text{scope}(x_o)$ and that $y \notin X_{(t_1)}$ for some $y \in R$.

Adopt the following <u>scanning procedure</u> in scope x_o. Start at x_o and scan down the last column. Then scan the second to last column from top to bottom and so on until all columns have been scanned. Pick the <u>first</u> such y in the scan and call it y_o.

Pictorially then we have one of the following three pictures, depending upon where y_o is in scope x_o in relation to x_o. Remember y_o is "blank" since $y_o \in X_{(t_1)}$.

<u>Case I</u>:

$$\begin{array}{cccc}
 & & & x_o \\
X & X & & \\
X & X & & \\
y_o & X & & \\
 & X & & \\
 & X & &
\end{array}$$

<u>Case II</u>:

$$\begin{array}{cccc}
 & & & x_o \\
y_o & X & X & \\
 & X & X & \\
 & X & X &
\end{array}$$

Case III:

X_o

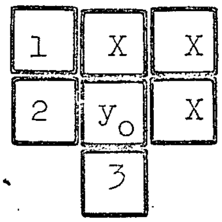
X

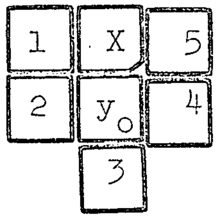
X

y_o

We show that none of these cases are possible. That is $y_o \in X_{(t_1)}$ contradiction.

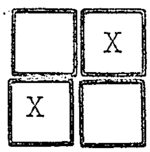
Cases I and III: By choice of y_o we have one of the following figures with y_o blank



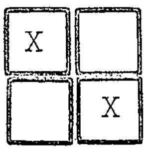


In order that y_o be blank and not have been activated by a DR we must have either 3 active or (2 active and 1 blank) or (4 active and 5 blank). These cases lead to the forbidden configurations.

or

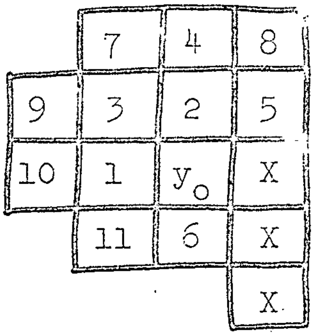


or



.

Case II: By choice of y_o we have



In order to avoid activation of y_o by a SC we must have either 1, 2, 3, or 4 active.

<u>Sub-case (1)</u>: 1 active yields $\boxed{X}\boxed{}\boxed{X}$ contradiction!

<u>Sub-case (2)</u>: 2 active implies also that 6 is active or (1 is active and 3 blank) or 5 is blank, all contradictions.

<u>Sub-case (4)</u>: 4 is active. We must avoid having 2 active by sub-case (2). The only way to avoid having a DR fill 2 is by having y_o active (false) or (3 active and 7 blank) or (5 active and 8 blank) both contradictions.

<u>Sub-case (3)</u>: 3 is active. We must avoid 1 being active. So we have 11 active or 10 active or (10 active and 9 blank) or y_o active.

In no case could we avoid the forbidden patterns. Therefore $X_{(t_1)} = X^o = \bigcup_{x \in X}$ scope (x) and $t_1 = t_o$. The theorem is shown.

<u>Corollary 1</u>: M_c recognizes ψ_{CONN}.

<u>Proof</u>: Let X be connected. When M_c sums it will do so on

the figure $X^o = \bigcup_{x \in X} \text{scope } x$, causing every associator to output a 1 , causing acceptance.

Let X be disconnected. When M_c sums (which it surely will do in a finite amount of time) it will do so on a figure Y . Certainly $Y \subseteq \bigcup_{y \in Y} \text{scope } (y)$. (Just look at the associators.) Since $Y \neq \bigcup_{y \in Y} \text{scope } y$, a connected set, we must have $y_o \in Y$ with some $x_o \in \text{scope } y_o \sim Y$. Pick any northeasterlymost such x_o . Then either

$$\boxed{x_o}\,\boxed{x} \quad \text{or} \quad \begin{array}{c}\boxed{x}\\ \boxed{x_o}\end{array}$$

with $(x_o \in Y , x \in Y)$ will occur leading to rejection because of the third component RD rule.

Therefore M_c accepts X iff X is connected. Therefore M_c recognizes ψ_{CONN} .

<u>Corollary 2</u>: ψ_{CONN} is of order ≤ 8 .

<u>Proof</u>: Since M_c recongizes ψ_{CONN} , the proof consists only of noting that all associators units of M_c have eight or less squares in them.

<u>Topics for further study</u>: Of course, one might always try to get better bounds on order (ψ_{CONN}) . The order 1 or 2 cases might be possible to rule out via brute force, but for order 3, 4, 5, 6 or 7 the problem seems to amount to either a hopelessly

large case elimination problem or a hopelessly difficult F.P. programming problem.

Other predicates might be investigated. $\psi_{\text{mod } n}$ would be easy. I have thought briefly about $\psi_{\text{SIMPLY CONN}}$ or $\psi(X) \Leftrightarrow X$ has exactly n components.

As Minsky notes, geometric properties involving straight lines and circles etc. involve tolerance topologies which seem to be more of a problem than the F.P.'s. Other algebraic predicates might be investigated.

Various training sequences could be investigated on the F.P.'s using computers. Unfortunately I would expect "learning" to be slower for these machines than for regular perceptrons if only because there are so many more modes of action for the F.P.

The problem of formulating a general theory of F.P.'s seems almost hopeless. In fact many of the minimal state problems for finite automata and iterative arrays are very much like our problems with connectivity. Only rough bounds are computed for specific tasks. No general theory is even hoped for.

We could vary the F.P.'s in such a way as to simplify tasks (i.e. lower orders). For example, we might permit the retina to be activated by n colours rather than 1.

We might note in passing that the design of feedback perceptrons is quite similar to that of the <u>cellular structures</u>

treated in von Neumann's Theory of Self Reproducing Automata. The main differences are in the summation algorithm of the end and in the fact that logical operations can be broken down over many cell neighbourhoods in feedback perceptrons. Perhaps a closer study of the relation of feedback perceptrons to cellular structures would lead to stronger methods for the perceptrons as well as giving them wider appeal.

The problem that really started us off was the computational complexity problem. As previously mentioned, the unlimited number of associators we allow undercuts the order measure as a true measure of complexity. Maybe a theory which counts total computation steps of any kind is accessible.

BIBLIOGRAPHY

Arbib, A., Brains Machines and Mathematics, McGraw-Hill, N.Y., 1965.

Feigenbaum, Edward A. and Julian Feldman, Computers and Thought, McGraw-Hill, N.Y., 1963. (Read especially the Samuel Checker Player).

Hebb, D. O., Organization of Behaviour, Wiley, N.Y., 1949.

Minsky, Computation: Finite and Infinite Machines, Prentice Hall, Inc., Englewood Cliffs, N.J., 1967.

Minsky, Marvin L. and Seymour Papert, Perceptrons and Pattern Recognition, (a monograph), M.I.T. Press, Cambridge, Mass. 1967.

Minsky, Marvin L. and Seymour Papert, Perceptrons, M.I.T. Press, Cambridge, Mass., 1969.

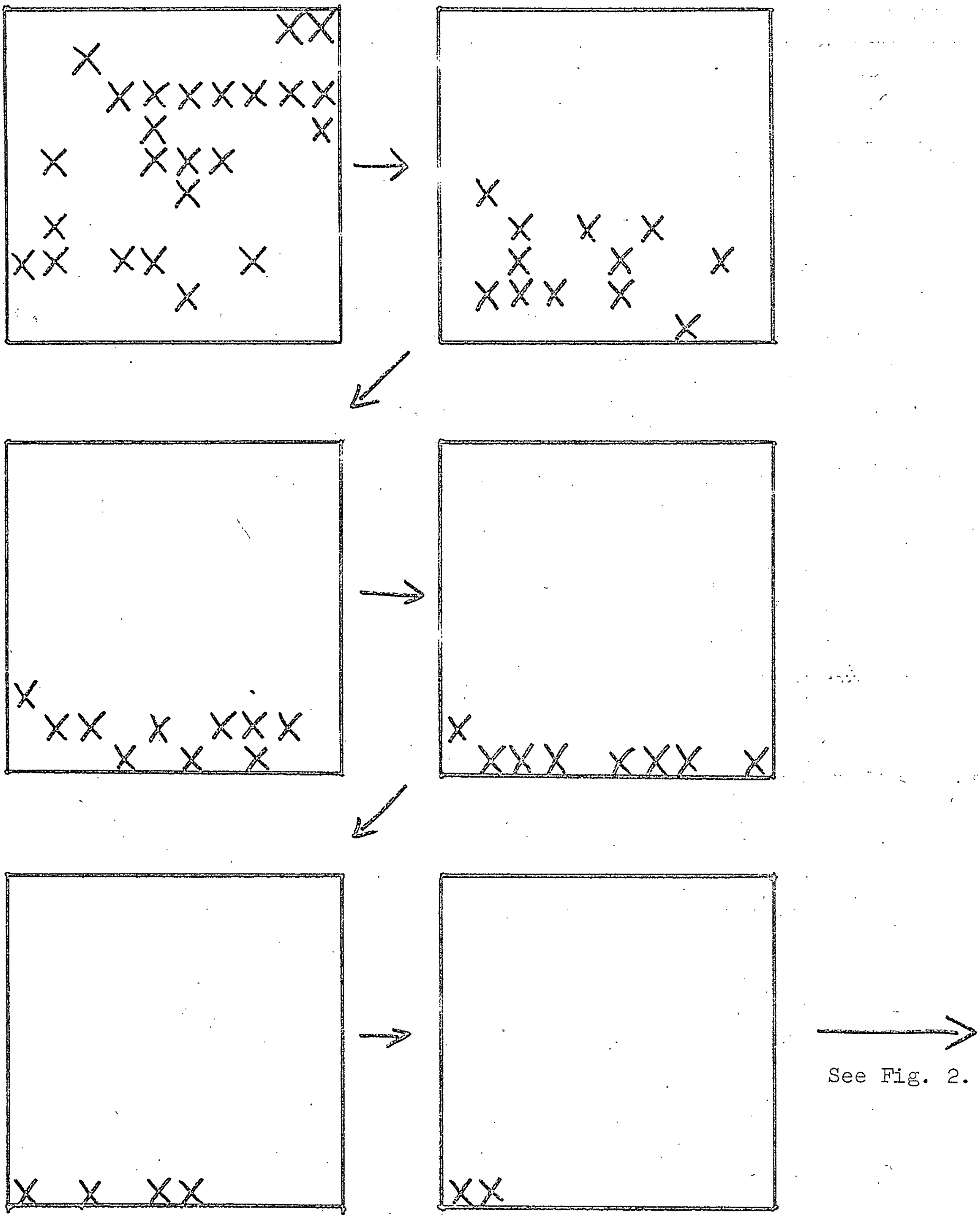
Nilsson, N.J., Learning Machines: Foundations of Trainable Pattern Classifying Systems, McGraw-Hill, N.Y., 1965.

Rosenblatt, F., Principles of Neurodynamics, Spartan Books, 1962.

Uhr, Leonard and James G. Miller, Pattern Recognition, Wiley, N.Y., 1966.

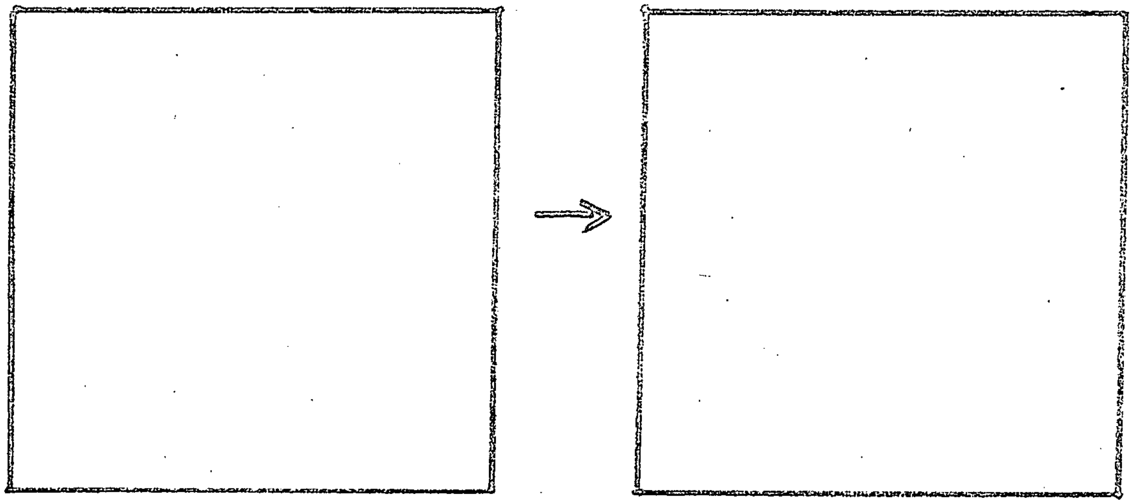
von Neumann and Burks, Theory of Self-Reproducing Automata, University of Illinois Press, 1966.

24 SQUARES



See Fig. 2.

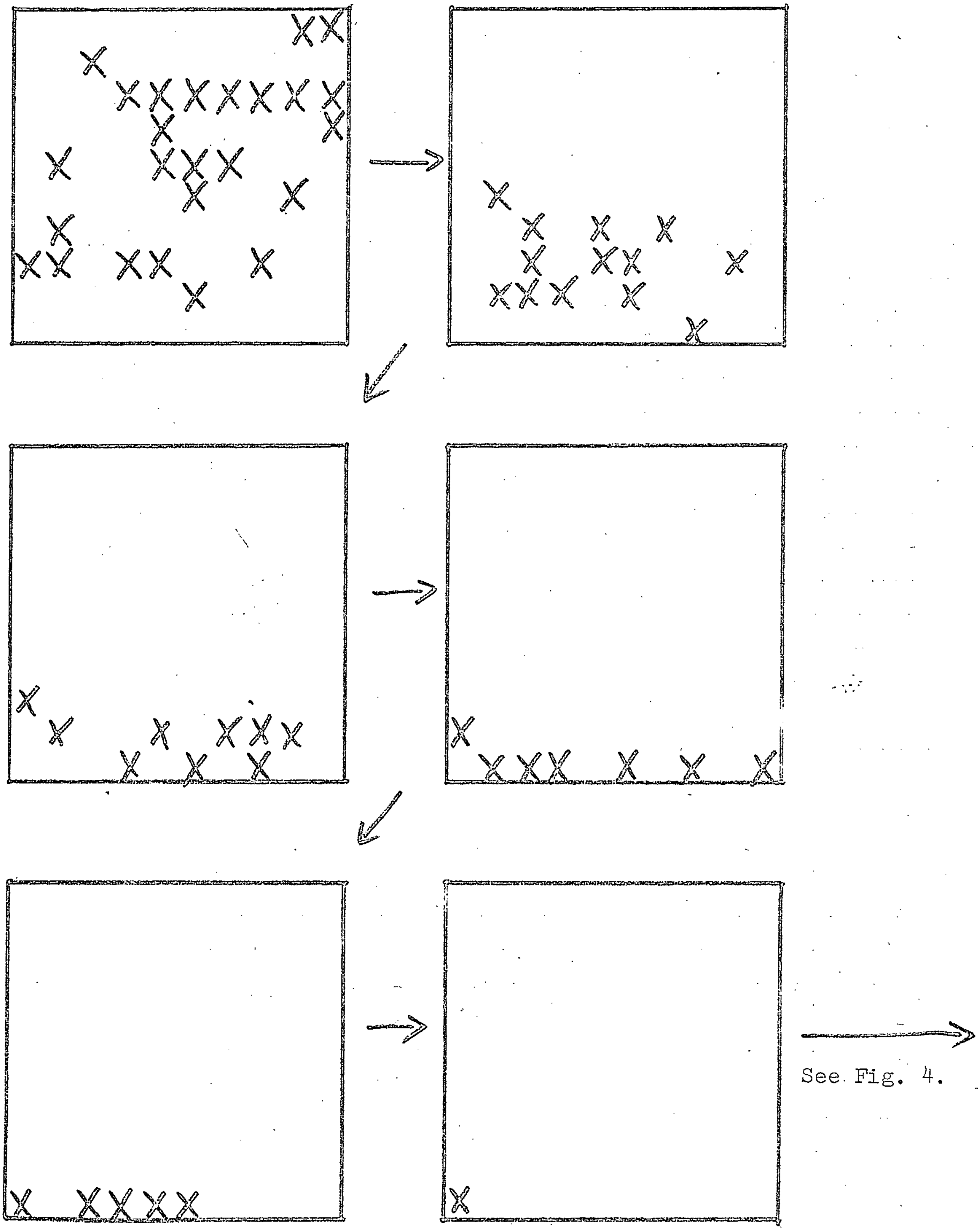
FIGURE 1



ACCEPT (SAME STATE)

FIGURE 2

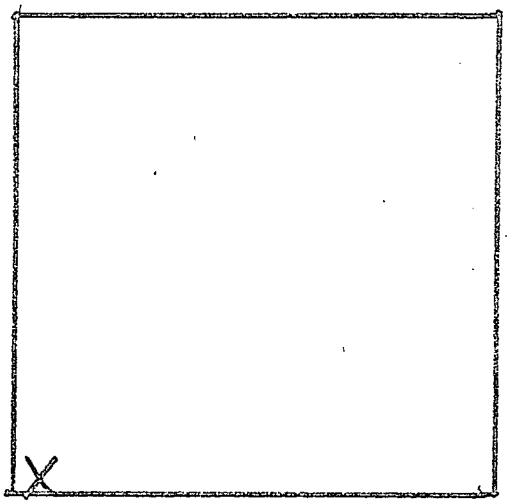
25 SQUARES



See Fig. 4.

REJECT

FIGURE 3



(SAME STATE)

FIGURE 4

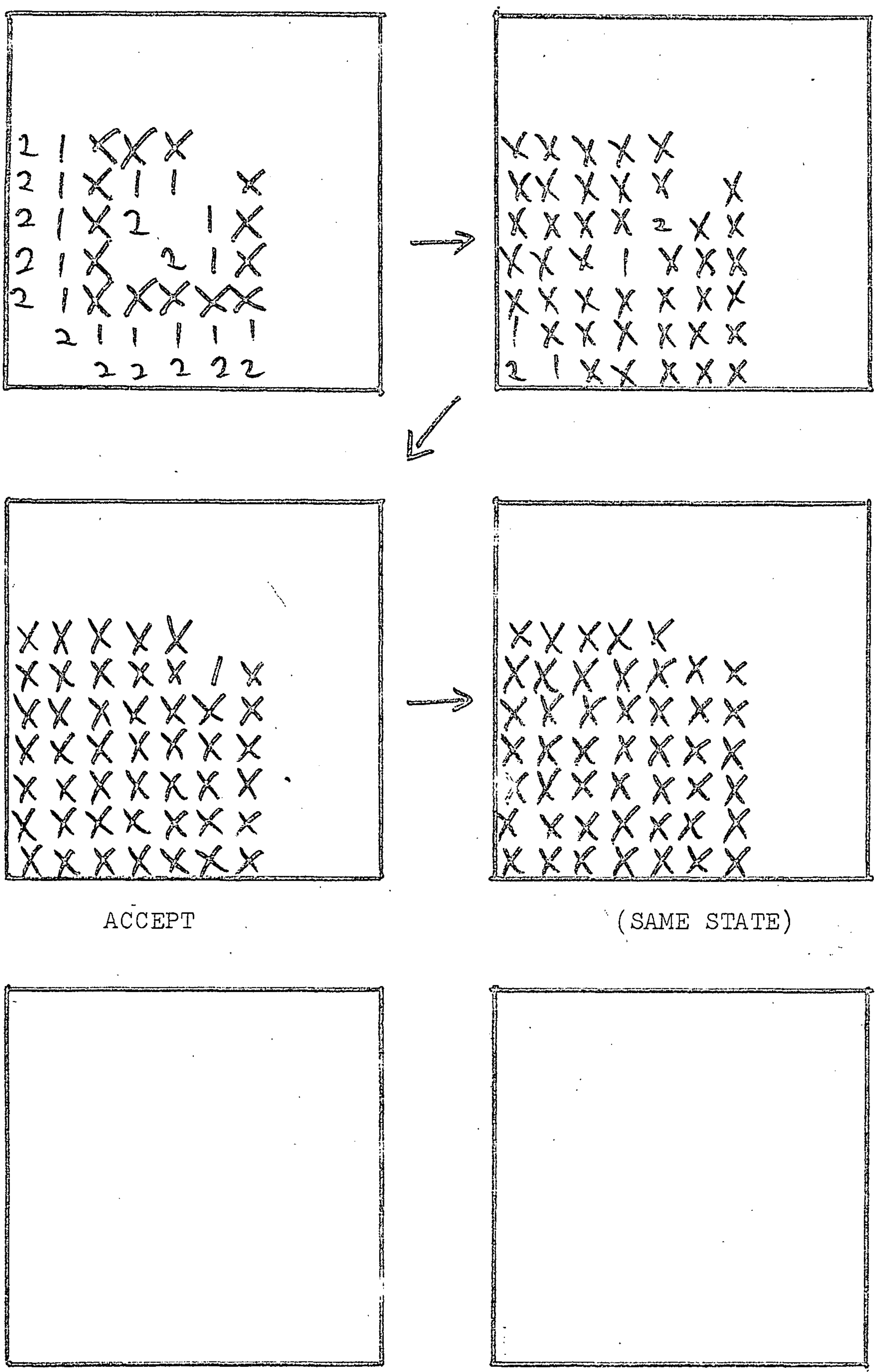


FIGURE 5

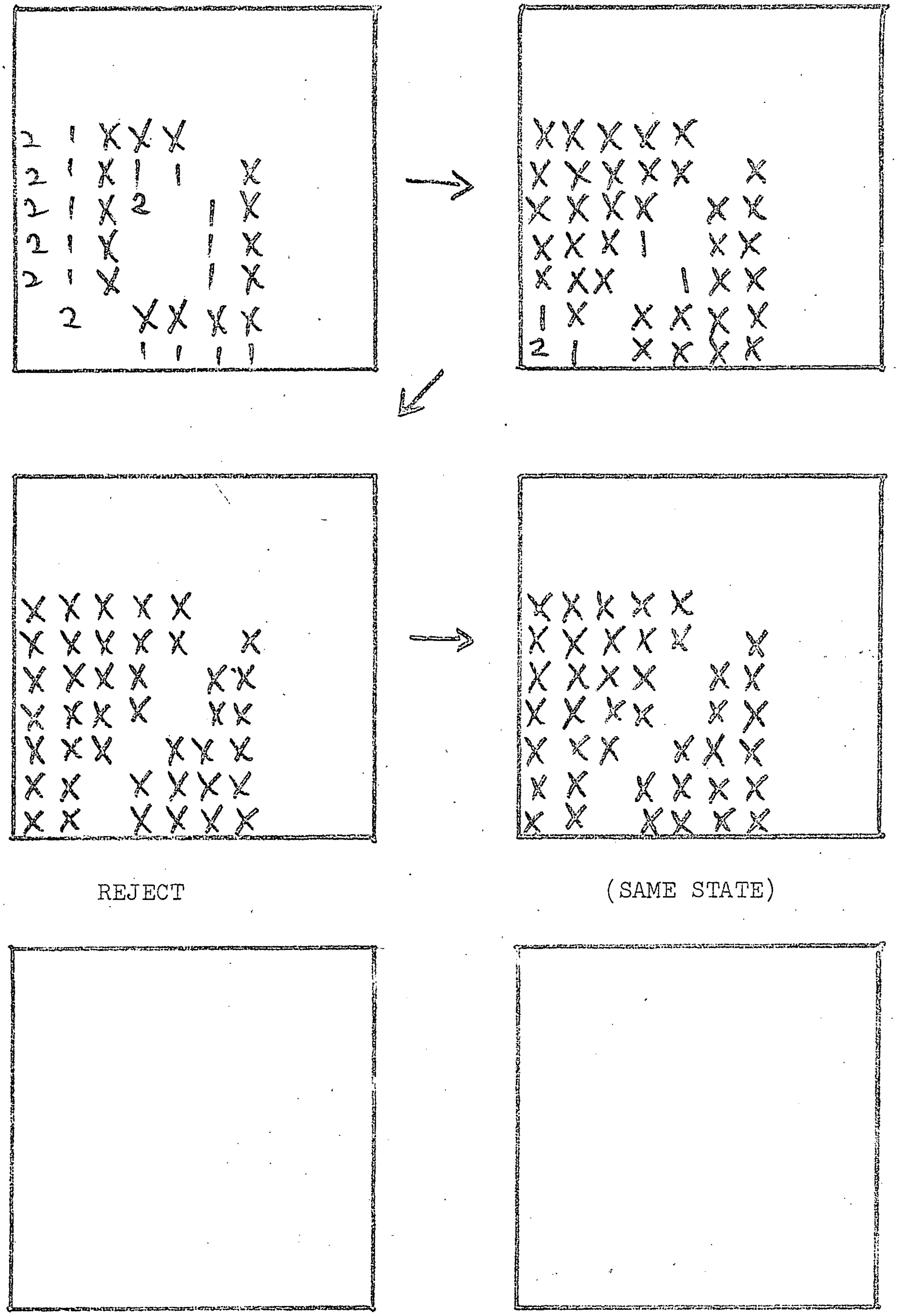
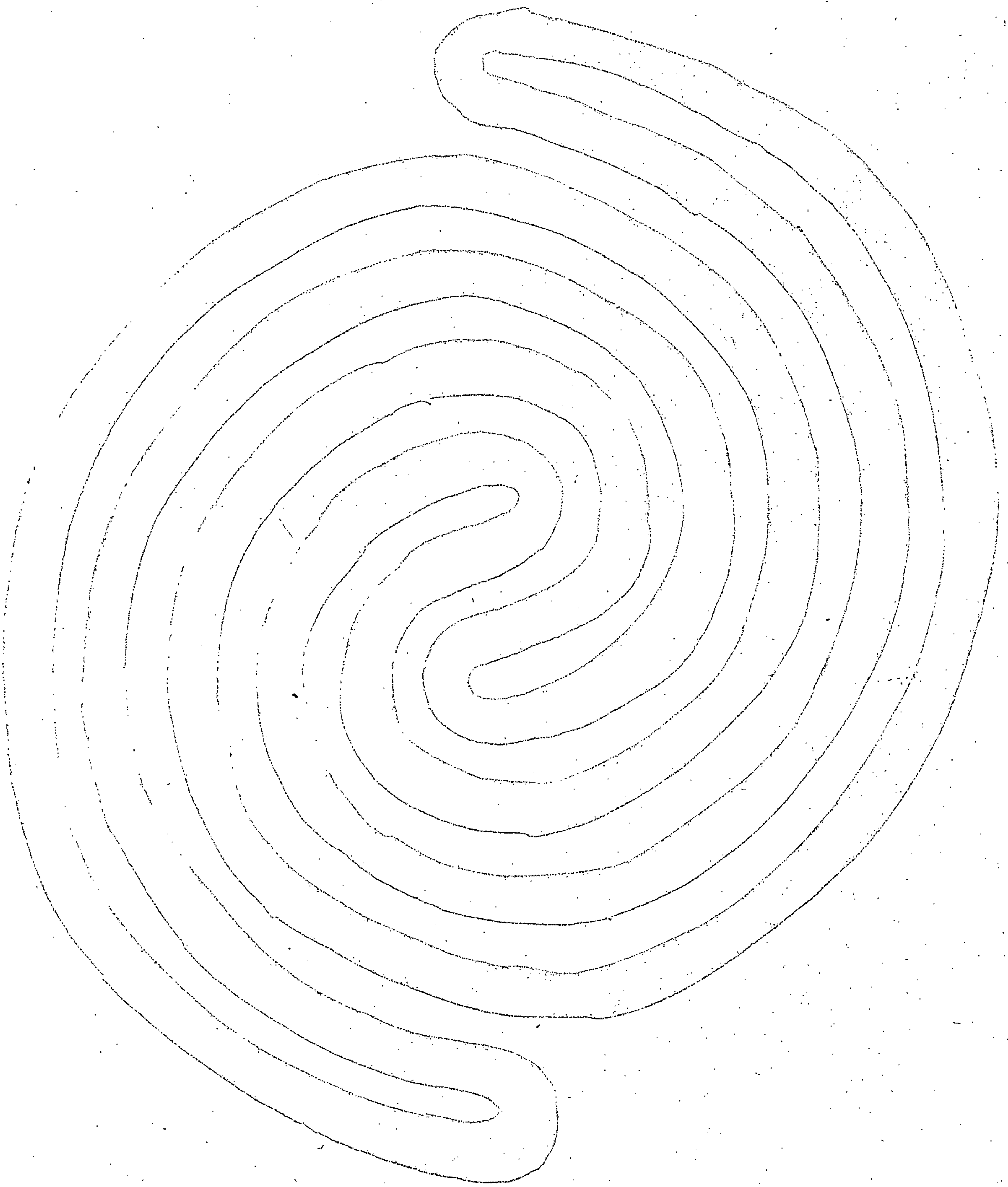


FIGURE 6



CONNECTED FIGURE 7

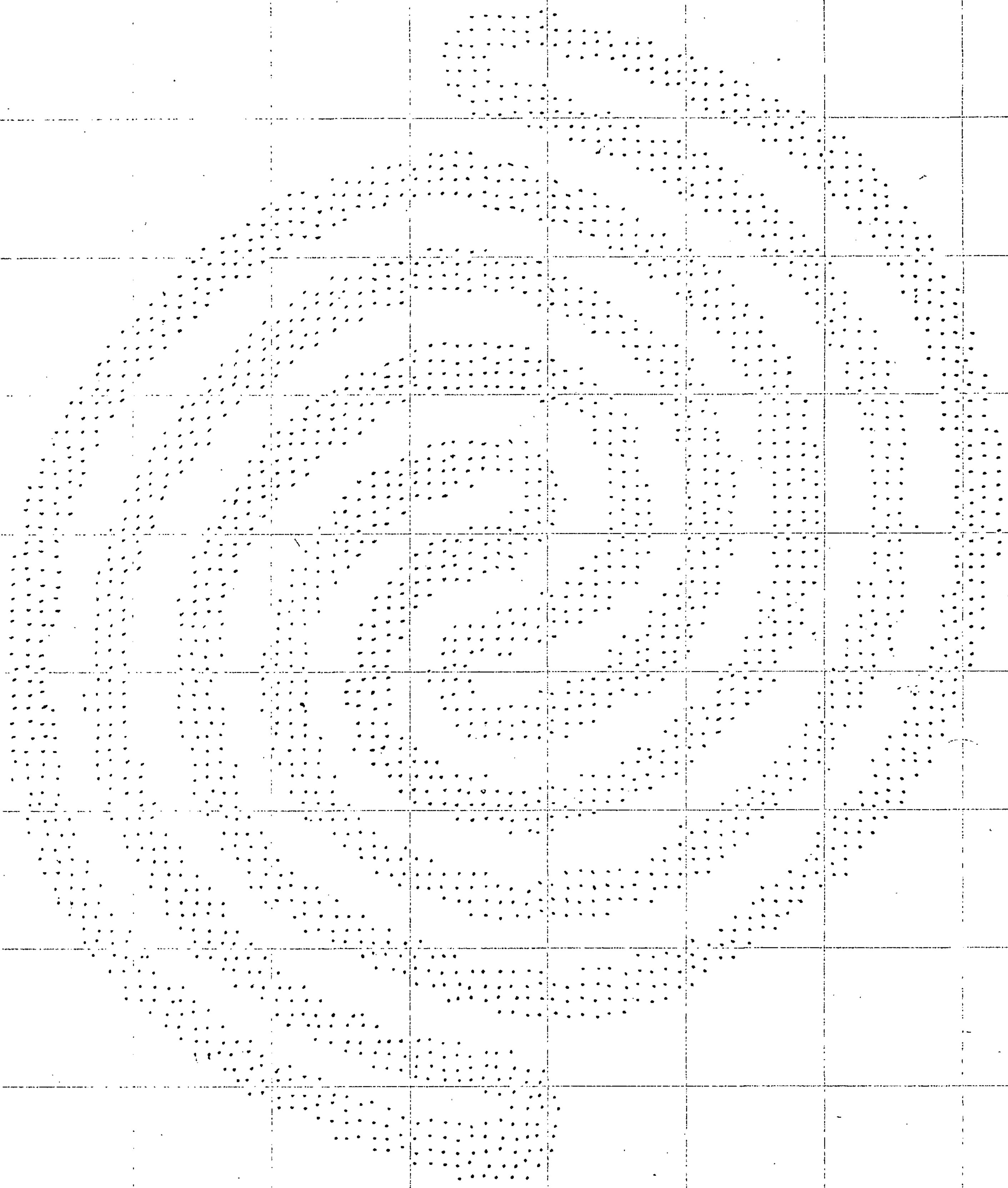


FIGURE 8

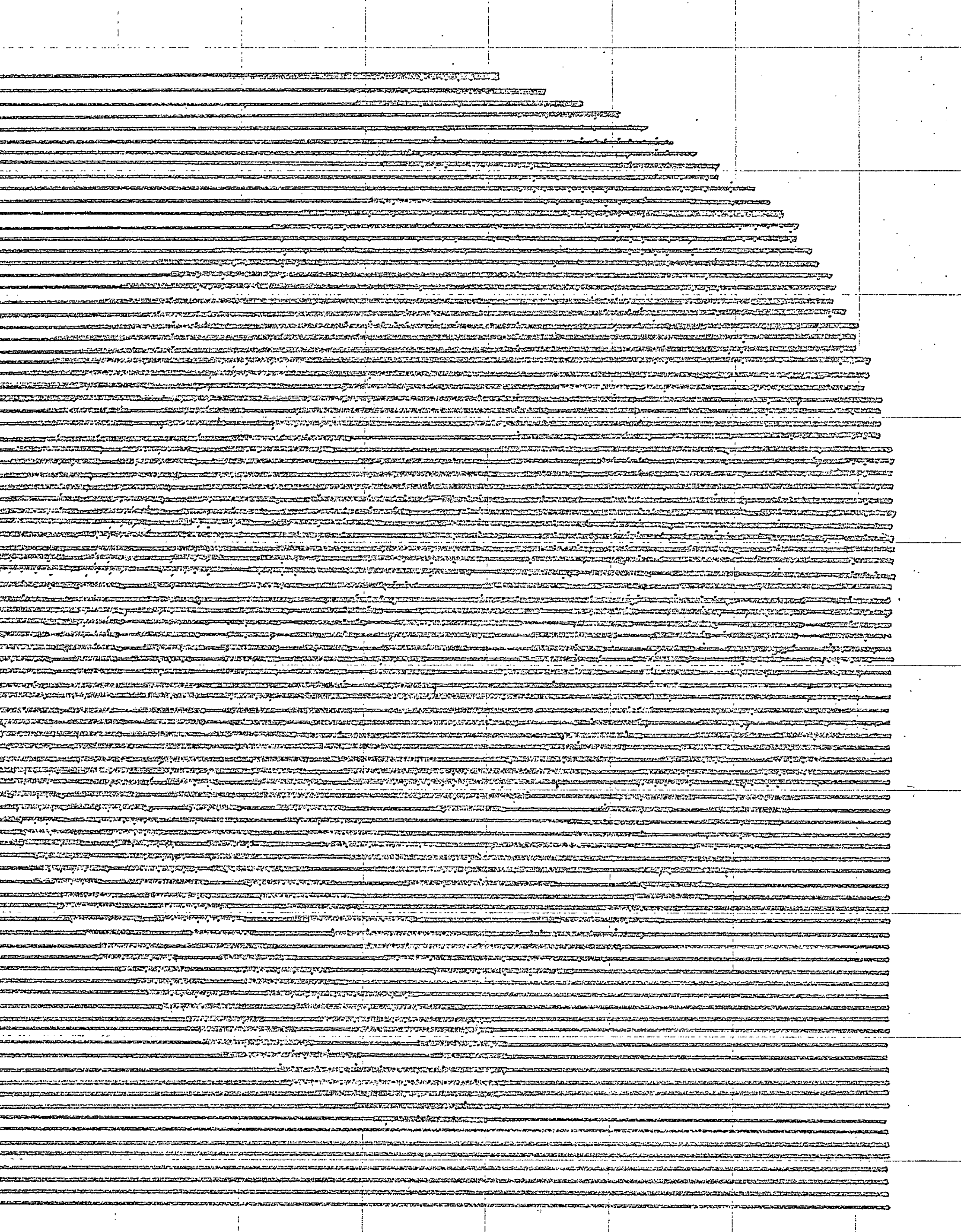
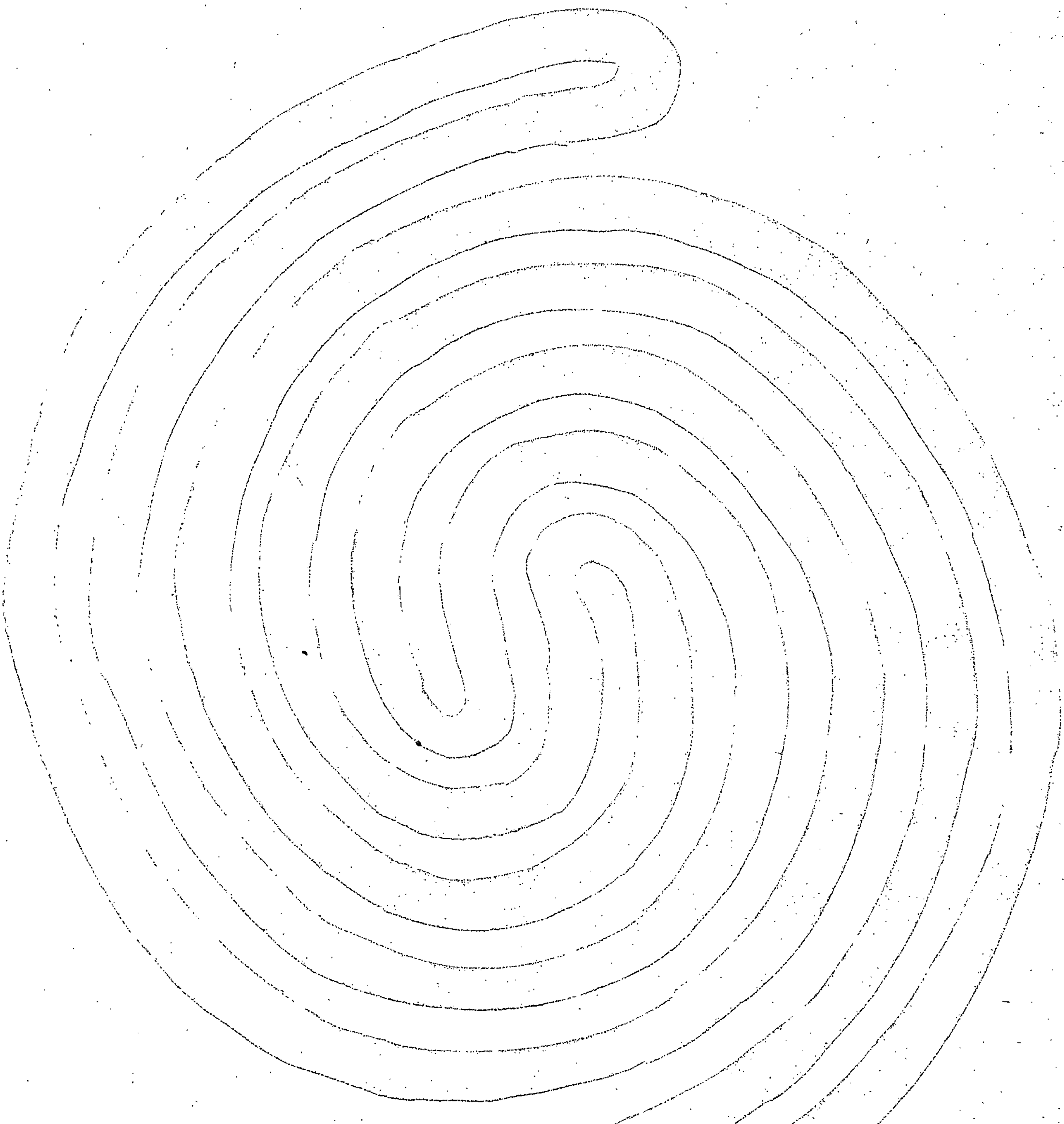


FIGURE 9



DISCONNECTED

FIGURE 10

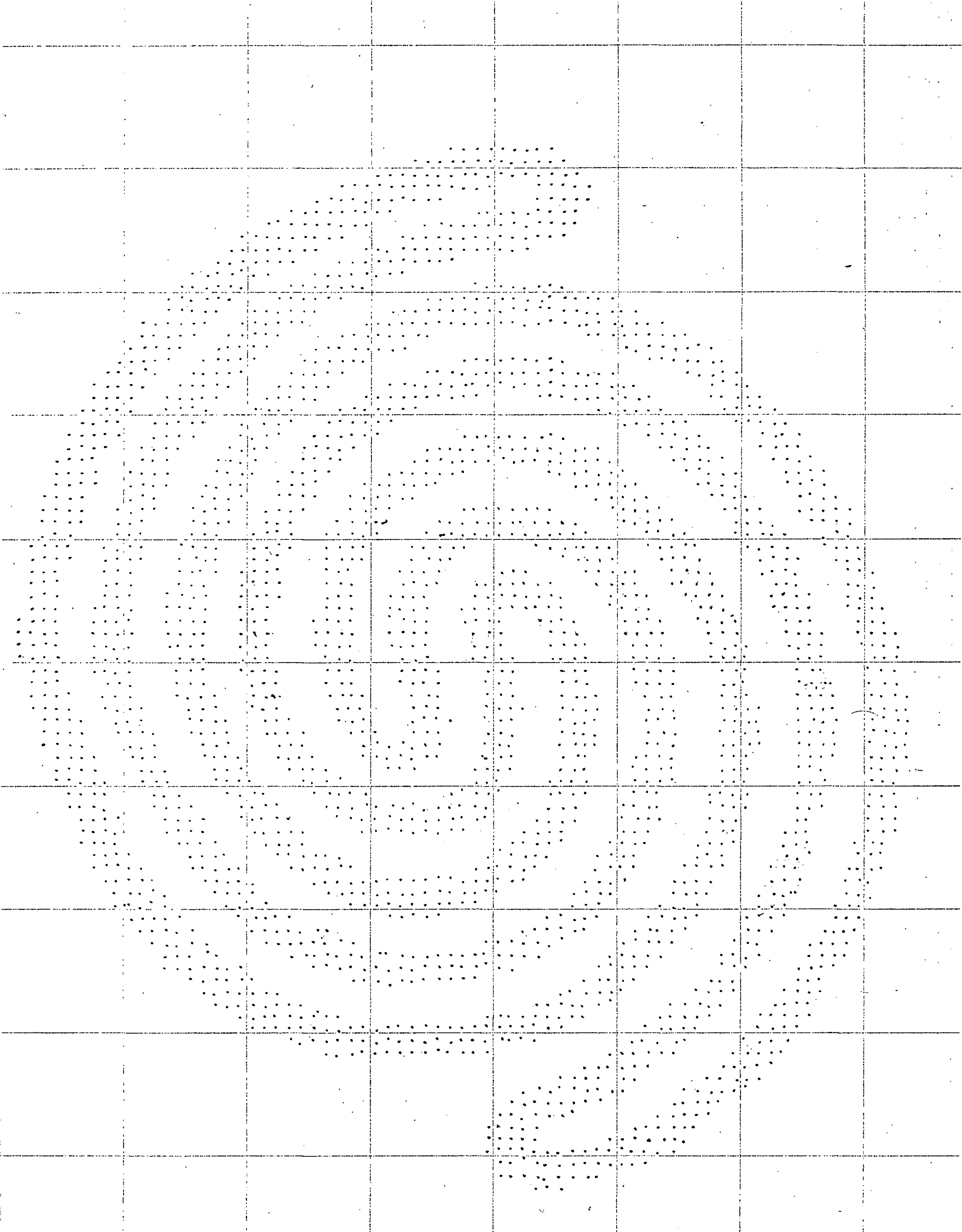


FIGURE 11

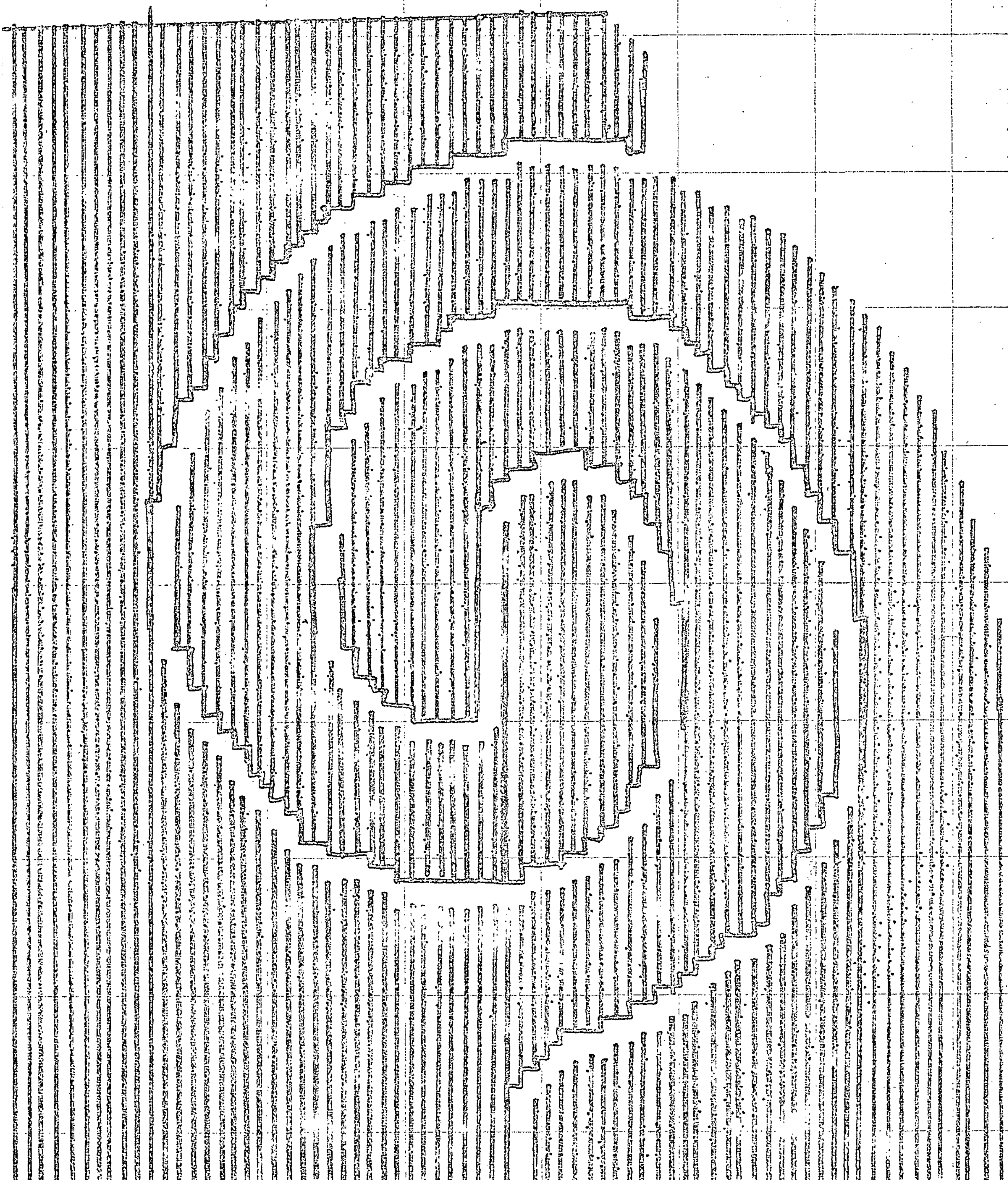


FIGURE 12