



*One month after Nils joined
the group - he wrote this
ML*

September 25, 1961

Proposal for Research

SRI No. ESU 61-123


MATHEMATICAL TECHNIQUES OF SELF-ORGANIZING SYSTEMS

Prepared for:

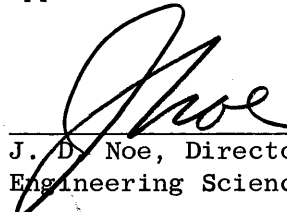
Rome Air Development Center
Griffiss Air Force Base
New York

Prepared by:


Nils J. Nilsson
Research Engineer


C. A. Rosen, Manager
Applied Physics Laboratory

Approved:


J. D. Noe, Director
Engineering Sciences Division

Copy No. 24

Proposal for Research

SRI No. ESU 61-123

MATHEMATICAL TECHNIQUES OF SELF-ORGANIZING SYSTEMS

I INTRODUCTION AND BACKGROUND

In response to Rome Air Development Center Purchase Request No. 152083, dated 10 July 1961, this proposal outlines a program of research aimed at the development of a mathematical structure sufficiently comprehensive to serve as a means for subsequently realizing a useful, economical, self-organizing machine.

A number of models for such machines have been proposed and are being actively explored. In particular, Taylor¹, Widrow², and Rosenblatt³ have versions which embody many important concepts. A good review has been presented by Hawkins⁴.

Basically, these models are quite similar, so that their study permits viewing the same major problems from different aspects. All these models make use, in effect, of a threshold logic module, with variable analog stores or weights to constitute the memory. It is proposed to make use of the most important common aspects of these models to serve initially for the formulation of a mathematical basis.

¹W. K. Taylor, "Pattern Recognition by Means of Automatic Analogue Apparatus," Proc. I.E.E. (London), Vol. 106, Pt. B (March 1959).

²B. Widrow and M. E. Hoff, "Adaptive Switching Circuits," Stanford Electronics Lab., Stanford, California, Report No. 2104-1 (July 1960).

³F. Rosenblatt, "Principles of Neurodynamics," Cornell Aeronautical Lab., Buffalo, New York, Report No. VG 1196-G-8 (March 1961).

⁴J. K. Hawkins, "Self-Organizing Systems--A Review and Commentary," Proc. IRE 49, pp. 31-47 (1961).

The term "self-organizing" can be interpreted to relate to a rather broad class of systems, e.g., biological systems, automata, physical systems (such as in the growth of crystals) and others. It is proposed to limit this study to a highly important subclass, namely, the learning machine system.*

At present the Institute is engaged in several sponsored programs involving the study, development and application of learning machines. These are:

- (1) Graphical Data Processing Research Study and Experimental Investigation, Contract DA 36-039 SC 78343, U.S. Army Signal Research and Development Laboratory, Fort Monmouth, New Jersey.
- (2) Research in Self-Organizing Machines, Contract Nonr 3438(00) Office of Naval Research, Washington, D.C.

The objective of the first program is to study and experimentally verify techniques for the recognition and classification of graphical patterns which arise in military applications, such as in reconnaissance photographs and military maps. This program has already resulted in the building of a small experimental learning machine which includes some

*The term "self-organizing machine" would, if taken literally, imply that a group of components and interconnections arranged in an undifferentiated mass could organize without human or machine intervention into a useful machine merely by being exposed to a set of external signals. Such organization would occur by virtue of the fact that the interconnections, logical properties of the components, or both, were adaptive or "variable, with a purpose." It may be conjectured that partial programming of a structure is necessary to permit development, by subsequent adaptation, of a useful organism (or machine) in a reasonable length of time. It is suggested that a partially preorganized system with a capability for internal alteration of its organization and memory stores be termed a "learning machine system," and that for this system a "teaching" process would be required, involving the purposeful application of signals from the external environment through human intervention or with the aid of non-biological transducers (or sensors).

novel logic and memory devices^{5,6} and has produced various contributions to learning theory. It is expected that this program will be continued with primary emphasis on the development of sampling techniques, new logic devices, and an improved machine organization; a much larger feasibility machine is now planned to implement these developments.

The second program is concerned primarily with a mathematical investigation and implementation of a preprocessing technique using concepts derived from studies of integral geometry. These techniques would permit the recognition of invariants of patterns under transformations of position, rotation, and size.

A third program, internally sponsored by the Institute, has been under way for the past two years; it is involved in the study of learning machines, their implementation and applications. Initial mathematical models, devices, and digital computer simulation studies have provided a great deal of useful data.

Both the past and future work in these above projects will be of considerable aid in providing the necessary technical background and tools for the proposed work. On the other hand, the proposed development of a mathematical framework will complement these projects.

II OBJECTIVE

The objective of this program is to develop a mathematical basis for the design of learning machines. The program will have the following specific goals:

- (1) Development of a geometrical framework
- (2) Isolation of the pertinent problems
- (3) Solution of the pertinent problems
- (4) Organization of the results to permit implementation and test of the mathematical theory.

⁵A. E. Brain, "The Simulation of Neural Elements by Electrical Networks Based on Multi-Aperture Magnetic Cores," Proc. IRE, 49, pp. 49-52 (1961).

⁶A. E. Brain, et al, "Graphical Data Processing Research Study and Experimental Investigation," Quarterly Reports, Contract DA 36-039 SC 78343, U.S. Army Signal Research and Development Laboratory, Fort Monmouth, New Jersey.

III METHOD OF APPROACH

A. Characteristics of the Mathematical Basis for Design

For the purposes of this study, we shall assume that a learning machine is a machine that is able to make decisions about a given input environment based on the machine's past experiences in that environment. We assume that the environment includes a teacher, either human or machine, and that the learning machine can be taught by simple methods.

A mathematical basis for the design of learning machines should resolve the following questions:

- (1) What are some economical learning machine configurations for a variety of different problems which may include the following as examples?
 - (a) Visual pattern recognition
 - (b) Electric signal recognition
 - (c) Function approximation
 - (d) Process control
- (2) How should these machines be trained so that they remember learned responses and generalize appropriately to others?
- (3) What are some suitable performance indices of these machines by which different designs may be compared?

B. Development of a Framework

Recent work at the Institute has established the beginnings of a geometric framework describing the properties of learning machines. This development is described in detail in the attached Appendix. One of the most important concepts discussed in the Appendix is the notion of separability of multi-dimensional spaces by hyper-surfaces. At present, this notion leads to the idea of constructing machines from a multiplicity of basic modules called threshold logic units. In its present stage of development, the framework serves to emphasize the essential similarities of some of the previously proposed models of learning machines^{1,2,3} and also points to some key questions about their design. These questions are:

- (1) Assuming a fixed total number of threshold logic units organized in a multi-layer machine what, if any, are the advantages of allocating these units to many layers of relatively few units each instead of to few layers of many units each?
- (2) What do various training procedures, forced or automatic, imply about the generalizing capabilities of learning machines? And conversely, how can the results of a decision-theoretic approach to the generalization problem be interpreted in terms of specific training procedures?
- (3) How can learning machines be used for the purpose of training other learning machines?
- (4) With regard to those inductive capabilities of a machine which depend on invariants and not on training, how can fixed (not adaptable) wiring be used in conjunction with the adaptive part of the machine?

C. Solution of the Problems

After the key problems are isolated, the following are illustrative of techniques available for their solution:

- (1) Linear-input logic
- (2) M-dimensional geometric analysis
- (3) Statistical decision theory
- (4) Integral geometry.

In addition to the above rather-formalized disciplines, a great store of practical experience in building and operating learning machines can be brought to bear on the problems. It is also intended to model promising solutions by digital computer simulation.

IV PERSONNEL

This work will be performed by staff members of the Applied Physics Laboratory and Mathematical Sciences Department of the Engineering Sciences Division. External consultation is available, and will be employed as needed. Biographies of key personnel follow:

Nilsson, Nils J. - Research Engineer, Applied Physics Laboratory

Dr. Nilsson received an M.S. degree in Electrical Engineering in 1956 and a Ph.D. degree in 1958, both from Stanford University. While a graduate student at Stanford he held a National Science Foundation Fellowship. His graduate field of study was the application of statistical techniques to radar and communications problems.

In July 1961 Dr. Nilsson completed a three-year term of active duty as a Lieutenant in the United States Air Force. He was stationed at the Rome Air Development Center, Griffiss Air Force Base, New York. His duties entailed research in advanced radar techniques, signal analysis, and the application of statistical techniques to radar problems. He has written several papers on various aspects of radar signal processing. While stationed at the Rome Air Development Center, Dr. Nilsson held an appointment as Lecturer in the Electrical Engineering Department of Syracuse University.

In August 1961 he joined the staff of Stanford Research Institute, where he is participating in the studies of pattern recognition and self-organizing machines.

Dr. Nilsson is a member of Sigma Xi, Tau Beta Pi, and the Institute of Radio Engineers.

Bliss, James C. - Research Engineer, Control Systems Laboratory

Dr. Bliss received a B.S. degree from Northwestern University in 1956, an M.S. degree from Stanford University in 1958, and a Ph.D. degree from the Massachusetts Institute of Technology in 1961, all in Electrical Engineering.

From 1953 to 1956 he was a Cooperative Student at the Argonne National Laboratory in Lemont, Illinois, where he worked on electrometer circuits, electrical conduction in solids, and automatic data read-out circuits.

In 1956 he joined the staff of the Control Systems Laboratory of Stanford Research Institute. He had full responsibility for a major portion of a project on alphanumeric reading. He was also responsible for the major part of the development of a frequency digitizer for an airborne system which could rapidly and accurately measure a high-frequency signal.

In 1958 he took a leave of absence from SRI to accept a National Science Foundation fellowship to do graduate work toward a Doctor of Philosophy degree at MIT and did his thesis work in the Sensory Aids Research Group on "Communication via the Kinesthetic and Tactile Senses."

His fields of specialty are communication theory, sensory processes, electronic systems, and human factors. He is the author of a paper on speech recognition in Automatic Control, co-author of a paper which has been accepted for publication in the Journal of the Optical Society of America; and he has submitted a paper for the IRE Professional Group on Information Theory special issue on Sensory Information Processing. He is also the author of a patent now pending on a technique for automatic character reading.

Dr. Bliss is a member of Phi Eta Sigma, Pi Mu Epsilon, Eta Kappa Nu, Tau Beta Pi, Sigma Xi, and Institute of Radio Engineers.

Fraser, Edward C. - Research Engineer, Electronics Group,
Control Systems Laboratory

Mr. Fraser attended the Worcester Polytechnic Institute at Worcester, Massachusetts, where he received his B.S. degree in Electrical Engineering in 1958. Following his graduation, Mr. Fraser did graduate work at the Massachusetts Institute of Technology, receiving his M.S. in September of 1960. He is presently working toward a Ph.D. at Stanford University.

Prior to joining the staff of Stanford Research Institute in October 1960, his experience included the analysis of aircraft electrical-power systems; a high-power servo-drive system for a radar antenna; and the development of a high-speed high-current drive scheme for computer memory cores. His most recent work at Lincoln Laboratory, M.I.T., was on an automatic missile-tracking system requiring design of an optimum predictor using a digital computer as a design tool for the later design of an optimum analog tracker.

At the Institute, Mr. Fraser has worked on projects including: the design of an adaptive controller for chemical processes; nonlinear application of semiconductor devices to obtain linear power amplification; analysis of the control requirements of a 50-BEV linear electron accelerator; and the application of analog-computation techniques to the solution of nonlinear, time-varying differential equations. His areas of specialization are nonlinear and adaptive systems.

Mr. Fraser is a member of Tau Beta Pi, Eta Kappa Nu, Sigma Xi, the Institute of Radio Engineers, and the American Institute of Electrical Engineers.

Forsen, George E. - Research Engineer, Applied Physics Laboratory

Mr. Forsen received both an S.B. and an S.M. degree in Electrical Engineering from the Massachusetts Institute of Technology in 1957, and the degree of Electrical Engineer from M.I.T. in 1959.

On the Cooperative Plan with M.I.T. he was employed part time in 1954-1956 by the General Electric Company. While with G.E. he was a

member of the Small Aircraft Engine Department (Lynn, Massachusetts), the General Engineering Laboratory (Schenectady, New York), and the Electronics Laboratory (Syracuse, New York), working on standards, non-destructive testing methods, and measurement techniques for heat flow in power transistors, respectively.

In 1958-1959 he was a member of the Communications Biophysics Group, Research Laboratory of Electronics at M.I.T., as a Research Assistant and staff member. There he designed electronic instrumentation for the study of neuroelectric and psychophysical phenomena related to nervous systems. From 1957 to 1959 he was also employed by the Electrical Engineering Department of M.I.T. as a Teaching Assistant.

In October 1959 Mr. Forsen joined the staff of Stanford Research Institute. At the Institute he is currently engaged in the study of field emission and neuron-like devices.

Mr. Forsen is a member of the Institute of Radio Engineers and Sigma Xi.

Singleton, Richard C. - Research Mathematical Statistician,
Mathematical Sciences Department

Dr. Singleton received both B.S. and M.S. degrees in Electrical Engineering in 1950 from the Massachusetts Institute of Technology. In 1952 he received the M.B.A. degree from Stanford University Graduate School of Business. He holds also the degree of Ph.D. in Mathematical Statistics from Stanford University, conferred in 1960. His Ph.D. research was in the field of stochastic models of inventory processes, applying the general theory of Markov processes; this work was done under Professor Samuel Karlin.

Dr. Singleton has been a member of the staff of Stanford Research Institute since January 1952. During this period, he has engaged in operations research studies, in the application of electronic computers to business data processing, and in general consulting in the area of mathematical statistics.

His experience at the Institute includes: (1) a study of the market and possible applications for a new digital computer; (2) a study of the potential computer applications in a large bank; (3) a computer feasibility study and implementation project for an electric utility firm; (4) a study of the equipment requirements for the mechanization of the passenger reservation system for a major airline; (5) a computer feasibility study and implementation project for an insurance company; and (6) an operations research study of the supply system of one of the military services. He has written several articles for professional journals.

Before joining the Institute staff in 1952, Dr. Singleton's industrial experience included work in the product engineering and industrial engineering departments at Philco Corporation in Philadelphia, and employment as the chief engineer for a radio broadcasting station. He acted as an instructor while doing graduate work at M.I.T.

Dr. Singleton is a member of a number of professional societies, including the Institute of Radio Engineers, the Operations Research Society of America, the Research Society of America, and Eta Kappa Nu.

Myhill, John - Consultant

Dr. Myhill received a B.A. from Cambridge in 1944 and a Ph.D. from Harvard University in 1949 both in Philosophy. He taught at Vassar College from 1948 to 1949, Temple University from 1949 to 1951, Yale University from 1951 to 1954, the University of California at Berkeley from 1954 to 1960. In 1960 he became Professor of Philosophy and Foundations of Mathematics, Stanford University.

Dr. Myhill held a Guggenheim Fellowship at the University of Chicago in 1953-1954. From 1956 to 1957, he served as consultant in air weapons research at the University of Chicago. In 1957 he became Director of National Science Projects 3466 and 7277 at Princeton, New Jersey, where he served until 1959. He was a Member of the Institute for Advanced Study at Princeton, New Jersey, from 1957 to 1959.

He is co-author of several books: "Recursive Equivalence Types," J. Myhill and J.C.E. Dekker, University of California Publications in Mathematics, Vol. 3 (N.S.) No. 3, pp. 67-214 (1960) and "Recursion Theory," J. Myhill and J.C.E. Dekker (in preparation).

He has published over 30 papers pertaining to Mathematics and Logic.

Rosen, Charles A. - Manager, Applied Physics Laboratory

Dr. Rosen received a B.E.E. degree from the Cooper Union Institute of Technology in 1940. He received an M.Eng. in Communications from McGill University in 1950, and a Ph.D. degree in Electrical Engineering (minor, Solid-State Physics) from Syracuse University in 1956.

During 1940-1943 he served with the British Air Commission as a Senior Examiner dealing with inspection, and technical investigations of aircraft radio systems, components, and instrumentation. From 1943 to 1946 he was successively in charge of the Radio Department, Spot-Weld Engineering Group, and Aircraft Electrical and Radio Design at Fairchild Aircraft, Ltd., Longueuil, Quebec, Canada. During the period 1946-1950 he was a co-partner in Electrolabs Reg'd., Montreal, in charge of development of intercommunication and electronic control systems. During this period he also acted as a self-employed consulting engineer in these fields. In 1950 he was employed at the Electronics Laboratory, General

Electric Co., Syracuse, New York, where he was successively Assistant Head of the Transistor Circuit Group, Head of the Dielectric Devices Group, and Consulting Engineer, Dielectric and Magnetic Devices Subsection. In August 1957 Dr. Rosen joined the staff of Stanford Research Institute, where he has been working on applied physics projects.

His fields of specialty include dielectric and piezoelectric devices, electro-mechanical filters, and a detailed acquaintance with the solid-state device field. He has contributed substantially as co-author to two books, Principles of Transistor Circuits, R. F. Shea, editor (John Wiley and Sons, Inc., 1953) and Solid State Dielectric and Magnetic Devices, H. Katz, editor (John Wiley and Sons, Inc., 1959).

Dr. Rosen is a Senior Member of the Institute of Radio Engineers, a member of the American Physical Society, American Institute of Electrical Engineers, and the Research Society of America. He has helped to organize and has been the co-chairman of the Dielectric Devices Subcommittee (28.5 IRE).

V REPORTS

It is proposed that Monthly Progress Letters and a Final Technical Report be submitted in accordance with the requirements of Exhibit RADC 3002. As the study proceeds, interim Technical Reports will be issued when a reasonably self-contained phase or topic has been completed.

VI ESTIMATED TIME AND CHARGES

The estimated time required to complete this project and report its results is 13 months. The Institute could begin work within one week following the acceptance of the contract. The estimated costs are detailed in the attached Cost Sheet. It is requested that any contract resulting from this proposal be written on a cost-plus-fixed-fee basis under the Basic Agreement No. AF 33(600)-7435, between the United States Air Force and Stanford Research Institute.

VII ACCEPTANCE PERIOD

This proposal will remain in effect until 30 November 1961. If consideration of the proposal requires a longer period, the Institute will be glad to consider a request for an extension in time.

COST BREAKDOWN

Personnel Costs

Supervisory, 1 man-month at \$		\$
Research Mathematician, 6 man-months at \$		
Research Engineer, 6 man-months at \$		
Research Engineer, 2 man-months at \$		
Research Engineer, 7 man-months at \$		
Research Engineer, 4 man-months at \$		
Editorial, 1/2 man-month at \$		
Secretarial and Clerical, 1-1/2 man-month at \$		

*Total Direct Labor

**Overhead at 100% of Direct Labor

Total

Direct Costs

Travel and Subsistence		
2 Transcontinental Trips at \$		
Telephone and Telegraph		
Computer Time--25 hours at \$/hr.		
Consultant's Fee--estimated 20 days at \$/day		
Report Production Costs		

Total Direct Costs

Total Estimated Costs

Fixed Fee at 7% Total Estimated Cost

TOTAL ESTIMATED PRICE

* Included in direct labor are all salary base costs such as vacation, holiday, and sick leave pay, social security taxes, and contributions to employee benefit plans.

** The overhead rate quoted represents current cost experience. It is requested that the contract provide for reimbursement at this rate on a provisional basis, subject to retroactive adjustment to fixed rates negotiated on the basis of historical cost data (in accordance with ASPR 3-704). The contract should also specifically provide for the inclusion of general research costs as an allowable indirect expense to the extent determined reasonable.

APPENDIX

AN APPROACH TOWARD A MATHEMATICAL THEORY OF LEARNING MACHINES

by

Nils J. Nilsson

I FORMULATION OF THE PROBLEM

A. Introduction

Many of the tasks which humans and some machines can perform are pattern recognition tasks. By pattern is meant some input to the senses of a human or to the transducers of a machine. By recognition is meant some appropriate response which is evoked by the input. Examples of pattern recognition are the following: (1) a human upon examining a photograph (input pattern) suddenly exclaims (the response) that he sees an airplane; (2) a machine examining signals on magnetic tape (input pattern) decides (response) that the signal is representative of the type presumed to emanate from a new enemy radar; and (3) a control system continuously monitoring an aircraft's altitude (input) adjusts the aircraft's control surfaces (output). In all of the above examples, the response is some (possibly complicated) function of the input. A learning machine or an adaptive machine would share the human ability to change the functional relation between present input and output, in accordance with the accumulated information stored from past experience.

B. Sensory Space and Response Space

Any input to a human or machine can conveniently be represented as a point in a multi-dimensional space. For example, if the input is an electric voltage waveform $S(t)$, it can be represented by perhaps N samples $S(t_1), S(t_2), \dots, S(t_N)$; these samples, in turn, can be thought of as the coordinates of a point in an N -dimensional space. If the input consists of many waveforms, the collection of all the sample values can similarly be represented as one point in a higher-order multi-dimensional space. A photograph or two-dimensional visual pattern can also be represented by a finite number of samples which can be thought of as the coordinates of a point in a multi-dimensional space. Whatever the form of the input, we think of it as a point in a space called an input or signal space. Let us call this space the S -space.

If there are K different responses which the input patterns are supposed to evoke, then the output of the machine or human can be thought of as one of K points in a response, or R -space. For example, if the output is a positioning of a potentiometer at any integer value between 0 and 100 ohms, the output space contains 99 points. A function of the

machine or human is then to transform a point in S-space to a point in R-space. If the machine used to accomplish this task is a learning machine, then some of the rules by which this transformation is made have to be learned by the machine.

C. Sensory Matrix

We now introduce a matrix which will be useful later. Suppose S-space is M-dimensional and contains the n points defined by the vectors $(s_{1j}, s_{2j}, \dots, s_{Mj})$ for $j = 1, 2, \dots, n$. These n vectors can be thought of as column vectors comprising an $M \times n$ matrix \mathcal{S} . That is,

$$\mathcal{S} = \begin{pmatrix} s_{11} & \dots & s_{1n} \\ \dots & s_{ij} & \dots \\ s_{M1} & \dots & s_{Mn} \end{pmatrix} \quad (1)$$

where the element s_{ij} is the i th sample of the j th input pattern. The \mathcal{S} -matrix contains the sample values of the input signals to the machine.

D. Statement of the Problem

The machine, then, is a device which can transform a point in S-space into one in R-space. The machine is described by a set of boundaries in S-space which divide the S-space into regions.

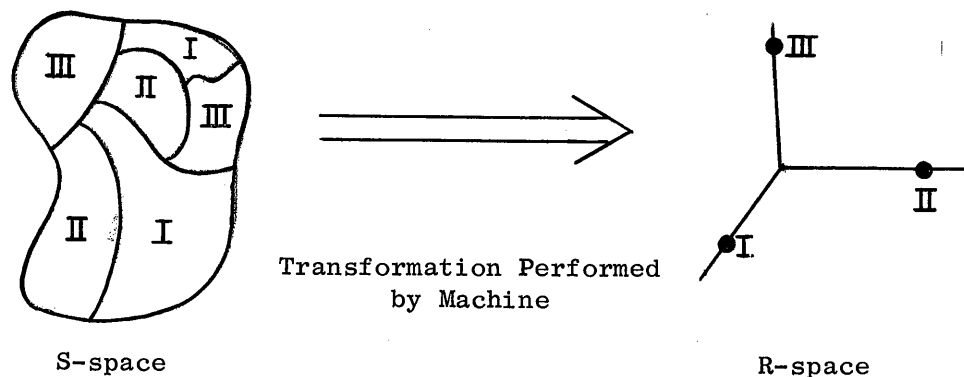


Fig. 1

Transforming S-space into R-space

The machine implementing the transformation in Fig. 1 will transform an arbitrary point in S-space into one of the three points in R-space, depending on the region of the point in S-space. As a result of this type

of treatment, we see that: (1) A categorization machine is one which is capable of drawing boundaries in S-space, (2) The specification of the machine is equivalent to the specification of the boundaries which make the appropriate transformation, (3) A learning machine is one which can change its boundaries to satisfy the dictates of a teacher.

A mathematical treatment of learning machines must then address itself to the following questions. Considering the responses to be learned for a set of points in S-space, how should a machine be built which draws boundaries in S-space in such a way that

- (1) Learned responses are remembered by the machine with sufficient reliability?
- (2) "Appropriate" responses are made for new (not yet learned) points in S-space?

The first question has to do with memory, the second with generalization or induction. The inductions can, in general, be made according to statistical decision theory¹ and are based on invariants built into the machine and generalizations learned by the machine. Both questions are phrased in the context of "how should a machine be built which draws boundaries?" and much of the mathematical theory to be developed is concerned with boundary-drawing machines generally.

II MACHINES THAT DRAW BOUNDARIES

A. Simplifying the Boundary Drawing Problem

To respond adequately to learned inputs and generalize appropriately to others, it is possible that quite complicated boundaries may have to be drawn in S-space. Synthesis of a machine with such boundaries is not an immediately straightforward process. We can synthesize a machine that draws complicated boundaries by a trick of proceeding from S- to R-space through intermediate spaces where the transformation from one intermediate space to the next depends only on simple boundaries. Examples of simple boundaries are hyperplanes, hyperspheres, or other surfaces which are simply instrumented. We shall proceed first with a discussion of instrumenting some simple boundaries.

B. Hyperplane Boundaries

If the components of the input signal vector $\vec{S} = (s_1, s_2, \dots, s_M)$ are each in turn weighted by the components of a weight vector

¹For an example of a decision-theoretic approach to the generalization or induction problem, see P. J. Braverman, "Machine Learning and Automatic Pattern Recognition," Tech. Report #2003-1, Stanford Electronics Labs., Stanford University, Stanford, California (17 February 1961).

$\vec{T} = (t_1, t_2, \dots, t_M)$ we have a generalized dot product:

$$\vec{T} \cdot \vec{S} = t_1 s_1 + t_2 s_2 + \dots + t_M s_M \quad (2)$$

When this weighted sum is equal to a constant, d , we have the equation of a hyperplane [such a hyperplane is called an $(M-1)$ -flat] in an M -dimensional S -space. If the weighted sum $\vec{T} \cdot \vec{S}$ is greater than d , then the end of the vector \vec{S} is on one side of the plane; if it is less than d , the end of the vector \vec{S} is on the other side of the plane. If $\vec{T} \cdot \vec{S} \geq d$, let us say that the point \vec{S} satisfies the positive condition with respect to the hyperplane determined by \vec{T} and d . The device shown in Fig. 2 is capable of deciding on which side of a hyperplane an arbitrary point in S -space lies. The orientation and position of the hyperplane can be changed or adapted by varying the weights (t_1, \dots, t_M) and/or the threshold d . Such a device is called a threshold logic unit.² It responds with an output = + 1 if the input satisfies the positive condition with respect to the plane, otherwise, the output is 0.

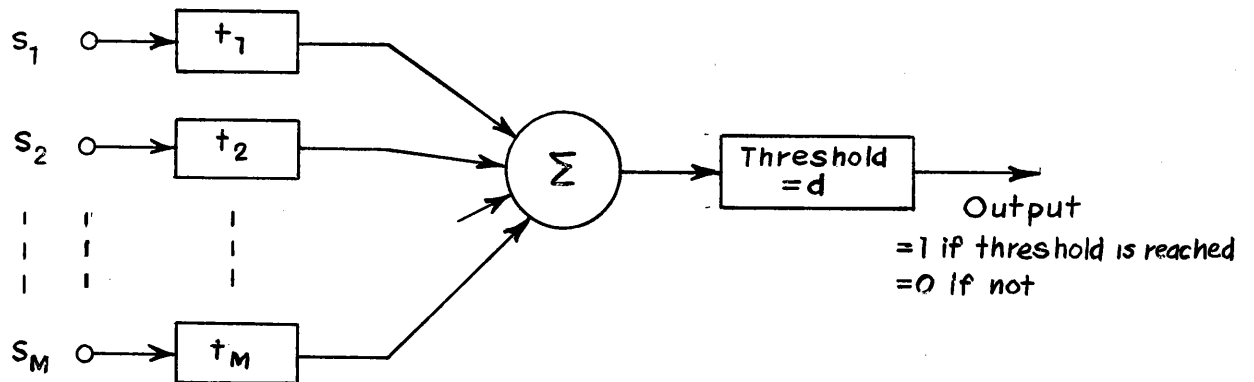
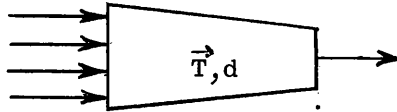


Fig. 2
Threshold Logic Unit

² Also called a linear input logic device. See, for example, R. C. Minnick, "Linear Input Logic," IRE Trans. on Elect. Computers, Vol. EC-10, Number 1 (March 1961).

The threshold logic unit will be represented schematically by the symbol



where \vec{T} is the weight vector and d the threshold.

The threshold logic unit, then, is a device which draws a plane in S -space. The S -space is thus divided into two regions. Let us represent the output of the threshold logic unit as one of the two points (0 or 1) in a one dimensional space called an A_1 -space. All of the points in the positive region of S -space (on the positive side of the plane) transform into the point "1" in A_1 -space. The points in the other region of S -space transform into the point "0" in A_1 -space.

We can easily divide S -space into more regions by passing more planes through it. Each plane is drawn by another threshold logic unit. An arbitrary point in S -space might then satisfy the positive conditions with respect to some of the planes and therefore the corresponding threshold logic-units will have +1 outputs. If we group together H_1 threshold logic units (H_1 -planes), we can represent the outputs of all the units as a point in an H_1 -dimensional A_1 -space. All points in A_1 -space lie on the vertices of an H_1 -dimensional hypercube. This set of threshold logic units is a machine which transforms a point in S -space into a point in A_1 -space. The planes in S -space form regions in S -space, and all of the points in the same region transform into one point in A_1 -space. Let us call each of the threshold logic units an A_1 -unit. A_1 -space will have as many dimensions as there are A_1 -units, each A_1 -unit corresponding to a plane in S -space. There are as many points in A_1 -space as there are regions formed by the planes in S -space. A machine for transforming S -space into A_1 -space is shown in Fig. 3. Let us call such a machine a two-layer threshold logic unit.

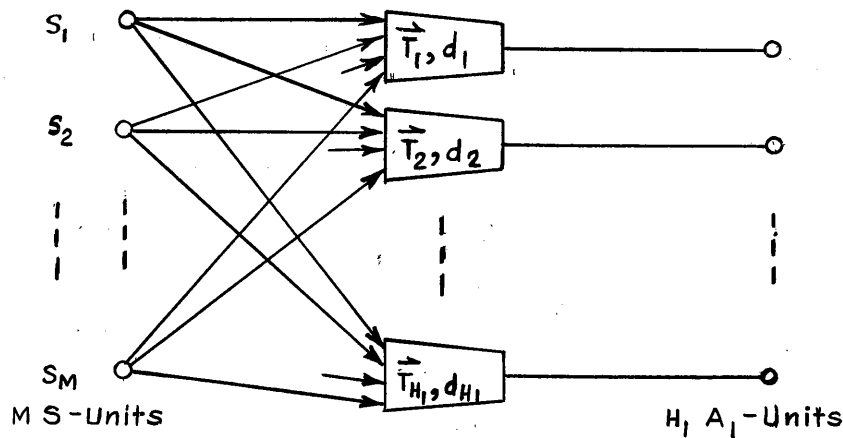
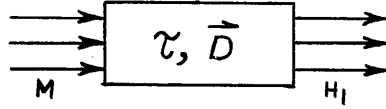


Fig. 3

Two-Layer Threshold Logic Unit

We can compound such a spatial transformation and proceed to an A_2 -space, then an A_3 -space, and so on. Let us represent the two-layer threshold unit illustrated above by the symbol



where τ is an $H_1 \times M$ matrix whose rows are the H_1 \vec{T} -vectors and \vec{D} is a vector composed of the H_1 thresholds. A general multi-layer device can then be represented schematically as in Fig. 4.

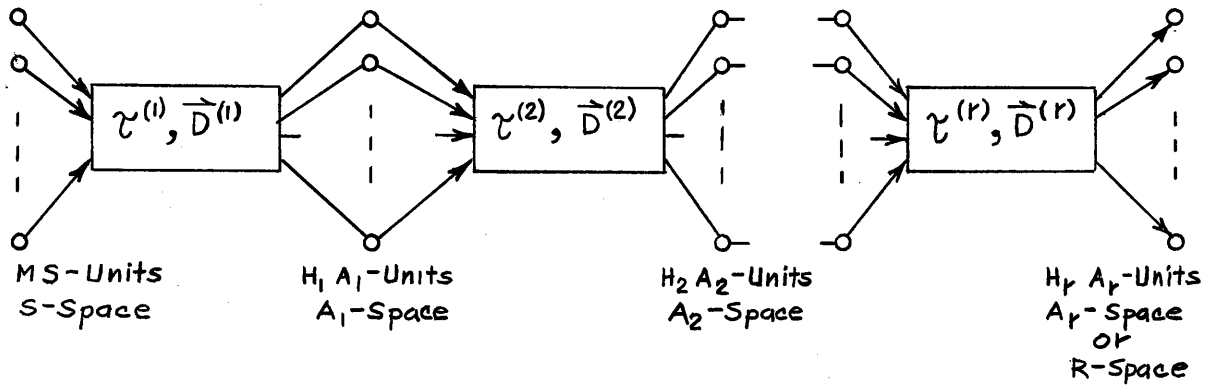


Fig. 4

Multi-Layer Threshold Logic Unit

In the multi-layer logic unit shown in Fig. 4, each of the components of the matrices and \vec{D} -vectors can, in general, be adjusted (adapted) to force the machine to categorize correctly learned responses. A systematic rule for changing these planes to force a desired response corresponds to the training procedure. The total effect in transforming from S-space to R-space will be as if quite complicated boundaries were used to separate S-space into regions.

C. Matrix Formulation

Let \mathcal{A} be the input signal matrix as defined by Eq. (1). $\tau^{(1)}$ is the linear operator (matrix of weights) which transforms points of S-space to points of B_1 -space. $B^{(1)}$ is a matrix, written as $[b_{ij}^{(1)}]$, consisting of all these points. $D^{(1)}$ is the non-linear operator which transforms points of B_1 -space to A_1 -space. $\tau^{(2)}$ is the linear operator on A_1 -space which transforms points to B_2 -space. $D^{(2)}$ is the non-linear operator which transforms points of B_2 -space into A_2 -space, and so on, until some A_r -space is the response space or R-space.

The linear operation performed by $\tau^{(1)}$ on \mathcal{L} , for example, is expressed in matrix form as follows:

$$\begin{bmatrix} t_{ik}^{(1)} \end{bmatrix} \begin{bmatrix} s_{kj} \end{bmatrix} = \begin{bmatrix} b_{ij}^{(1)} \end{bmatrix} \quad (3)$$

where

$t_{ik}^{(1)}$ = weight given by the i th A_1 -unit to the k th input component
 $b_{ij}^{(1)}$ = input to the i th A_1 -unit threshold when Pattern j is the input.

Equation (3) states that

$$b_{ij}^{(1)} = \sum_{k=1}^M t_{ik} s_{kj} \quad (4)$$

where M is the dimension of S -space.

$\begin{bmatrix} t_{ik}^{(1)} \end{bmatrix}$ premultiplies $\begin{bmatrix} s_{kj} \end{bmatrix}$ so to preserve the identity of the input signals as column vectors of $\begin{bmatrix} b_{ij}^{(1)} \end{bmatrix}$.

The non-linear operation performed by $D^{(1)}$ on $\begin{bmatrix} b_{ij}^{(1)} \end{bmatrix}$, for example, may be written as if the elements of $D^{(1)}$ formed a vector $\{d_1^{(1)}, d_2^{(1)}, \dots, d_{H_1}^{(1)}\}$, where H_1 is the dimension of B_1 -space. $D^{(1)}$ operates on the column vectors of $\begin{bmatrix} b_{ij}^{(1)} \end{bmatrix}$ so that

$$\left\{ D^{(1)} \right\} \begin{bmatrix} b_{ij}^{(1)} \end{bmatrix} = \begin{bmatrix} a_{ij}^{(1)} \end{bmatrix} \quad (5)$$

where $a_{ij}^{(1)} = \begin{cases} 1 & \text{if } b_{ij}^{(1)} \geq d_i^{(1)}, i = 1, 2, \dots, H_1 \\ 0 & \text{otherwise.} \end{cases}$

The D operator corresponds to a threshold operation on the weighted sums of the input. Equations (3) and (5) can now be applied using the linear operator $\tau^{(2)}$ and the non-linear operator $D^{(2)}$ to take us from A_1 -space through B_2 -space to A_2 -space, etc.

D. Separability of Spaces

Suppose S -space contains n points, each of which belongs to Category I or Category II. That is, one type of response is appropriate

for some of the input patterns and another is appropriate for the rest. The number of categories, K , is equal to 2. If a hyperplane can divide the points of one category from those of the other, then the S -space is said to be linearly separable. If S -space is linearly separable, the one-dimensional A_1 -space (with two points 0 and 1) is an R -space, and our problem is ended. If S -space is not linearly separable, then we must proceed, perhaps to an A_{r-1} -space which is linearly separable, making the A_r -space (one dimension, two points) the R -space. In general, if there are K categories, S -space may be K -linearly separable. A space is K -linearly separable if and only if it can be divided into K regions by planes with each region containing points of only one category. If S -space is K -linearly separable, then A_1 -space will have K -points, each corresponding to one of the categories, and, therefore, A_1 -space is an R -space. If S -space is not K -linearly separable, then we must proceed, perhaps to an A_{r-1} -space which is K -linearly separable, making the A_r -space the R -space. The above concepts will be illustrated in the next section.

III 3-LAYER THRESHOLD LOGIC DEVICES

A. A Fundamental Theorem

Henceforth, let us consider binary S -spaces, i.e., the points in S -space are constrained to lie on the vertices of the unit hypercube. With this restriction we can state and prove the following theorem.*

Theorem: Given a binary S -space with K -categories. An A_1 -space can always be obtained, by using separating planes in S -space, which is K -linearly separable. Thus, no more than 3-layers (S -space, A_1 -space, and R -space) are needed to give correct responses for any pattern.

Proof: S -space can be separated into $H_1 + 1$ regions by H_1 planes which do not intersect within the unit cube. These planes also have the property that they cut off one or more vertices of the S -space unit cube, and any point in S -space satisfies the positive condition with respect to one and only one of the planes. Each of the regions contains points of only one category, and, in general, $H_1 + 1$ is greater than K . Thus, A_1 -space has H_1 dimensions (one for each plane) and $H_1 + 1$ points (one for each region). H_1 of the points are each on one of the coordinate axes and the other is at the origin. It can easily be shown that such an A_1 -space is K -linearly separable.**

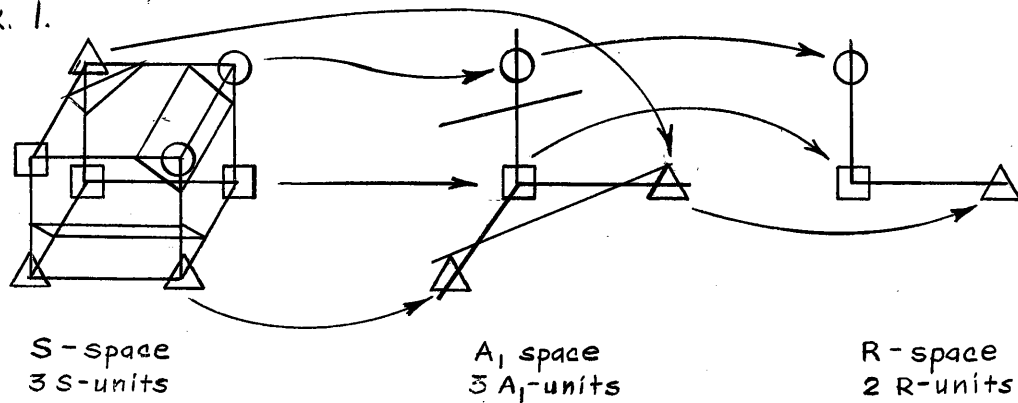
* This theorem is a generalization of Rosenblatt's theorem which states that a 3-layer α perception can always be built to dichotomize any input space. Note that the 3-layer threshold logic device is a 3-layer, α -perceptron.

** R. Singleton has in fact shown that this A_1 -space is always K -linearly separable by parallel planes.

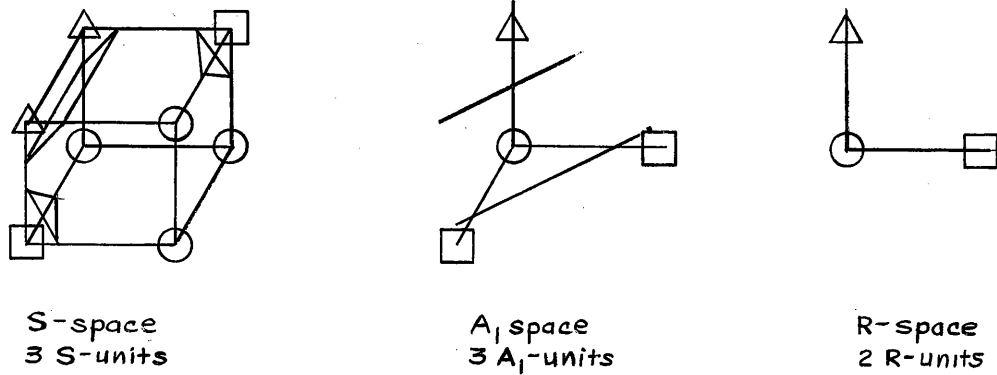
B. Examples

The following examples serve to illustrate the method of the above theorem. In the examples, the categories of the unit cube are marked by the symbols Δ , \square , or \circ . Note that the "positive" side of each plane (the side on which a point must be to turn on that plane's A_1 -unit) is always "away from" all of the other planes. Thus, only one A_1 -unit is turned on for each input pattern.

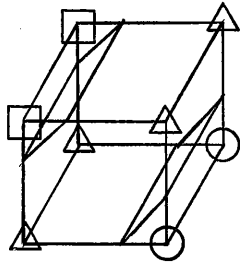
Ex. 1.



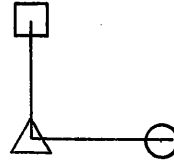
Ex. 2



Ex. 3



S-space
3 S-units



A_1 space or
R-space
2 A_1 -units

S-space is
3-linearly separable

The fundamental theorem and its method of proof are summarized by the entries in Table I.

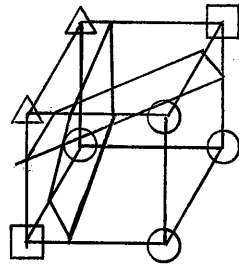
Table I
Relationships for a 3-Layer Threshold Logic Unit

	<u>S-space</u>	<u>A_1-space</u>	<u>R-space</u>
Number of dimensions	M	H_1	$K-1$
Number of points	n	$H_1 + 1$	K
Number of planes	H_1	$K-1$	
Number of regions	$H_1 + 1$	K	
Number of categories	K	K	K

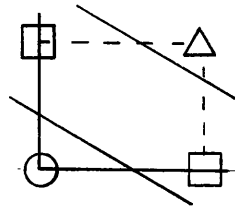
C. Use of Planes Which Intersect Within the Unit Cube

It is obvious that the $H_1 + 1$ regions in S-space could have been formed with fewer than H_1 planes if we allowed the planes to intersect within the unit cube. Thus, we might have been able to build a machine with fewer A_1 -units. As an example of how the use of intersecting planes can reduce the number of A_1 -units, we shall now repeat Example 2 of the last section using intersecting planes.

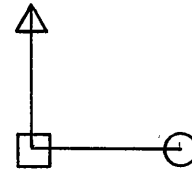
Ex. 4



S-space
3 S-units



A_1 -space
2 A_1 -units



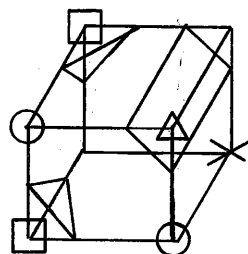
R-space
2 R-units

In the above example, only two A_1 -units were needed as opposed to the three needed in Ex. 2.

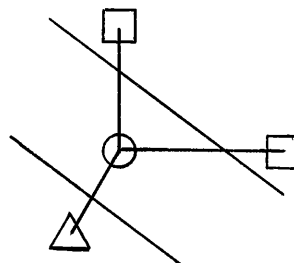
One must be careful with the use of intersecting planes, however. Their indiscriminate use might lead to an A_1 -space which is not K-linearly separable. But, even if A_1 -space is not K-linearly separable, we can proceed to another space that is. A trade-off is immediately apparent. We can reduce the number of A_1 -units needed while increasing the number of A_2 , A_3 , etc., units needed. In all probability we will find that multi-layer logic units are more economical of A-units required than is the simple 3-layer device. It is proposed that this question be explored fully.

IV THE INDUCTION AND TRAINING PROBLEM

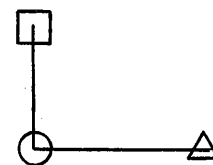
Consider the example shown in Fig. 5. Suppose that the categorized vertices in S-space represent points whose categories a machine must learn by changing its weights and thresholds. Suppose, using a certain training procedure and assuming a certain input sequence the state of the machine after training is represented by the planes shown in Fig. 5. The point X is as yet unlearned by the machine.



S-space



A_1 -space



R-space

Fig. 5

An Example Illustrating the Induction Problem

Suppose the fully trained machine is tested on point X. It will immediately say that X belongs to category "0". However, a different training procedure may have resulted in differently placed planes causing X to be classified, perhaps, as Δ . It is proposed that the problem of induction capabilities viewed as a function of the adaption rules of the machine be more fully investigated.

V CONCLUSIONS

The approach outlined above is a convenient medium in which to ask some of the fundamental questions about learning machines. It is proposed that the effort of constructing a Mathematical Theory pertinent to the class of Learning Machines previously described be directed towards answering the following basic questions:

- (1) Are there methods other than the one using non-intersecting planes which will guarantee K-linearly separability of A_1 -space?
- (2) If intersecting planes are used such that A_1 -space is not K-linearly separable, what are the possible trade-offs between the number of A_1 -units saved and the number of A_2 , A_3 , etc., units thus required?
- (3) What do various training procedures (for changing the positions of the planes) imply about the generalizing capabilities of learning machines? And, conversely, how can the results of a decision-theoretic approach to the generalization problem be interpreted in terms of specific training procedures?
- (4) How can learning machines be used to train other learning machines?
- (5) What are the relative advantages and disadvantages of using simple separating surfaces other than hyperplanes; hyperspheres, for example?
- (6) With regard to those inductive capabilities of a machine which depend on invariants and not on training, how can fixed (not adaptable) wiring be used in conjunction with the adaptive part of the machine? The fixed wiring may be all in the first layers in which case it is called "pre-processing." As an example of an "invariant" the machine may be told that all patterns be sensed as the same for all rigid motions in the plane. For this invariant, then, the machine is not willing to change its mind as a result of experience and thus keeps part of its wiring fixed.

It is felt that progress toward answers to the above questions will form a mathematical basis for the design of learning machines.