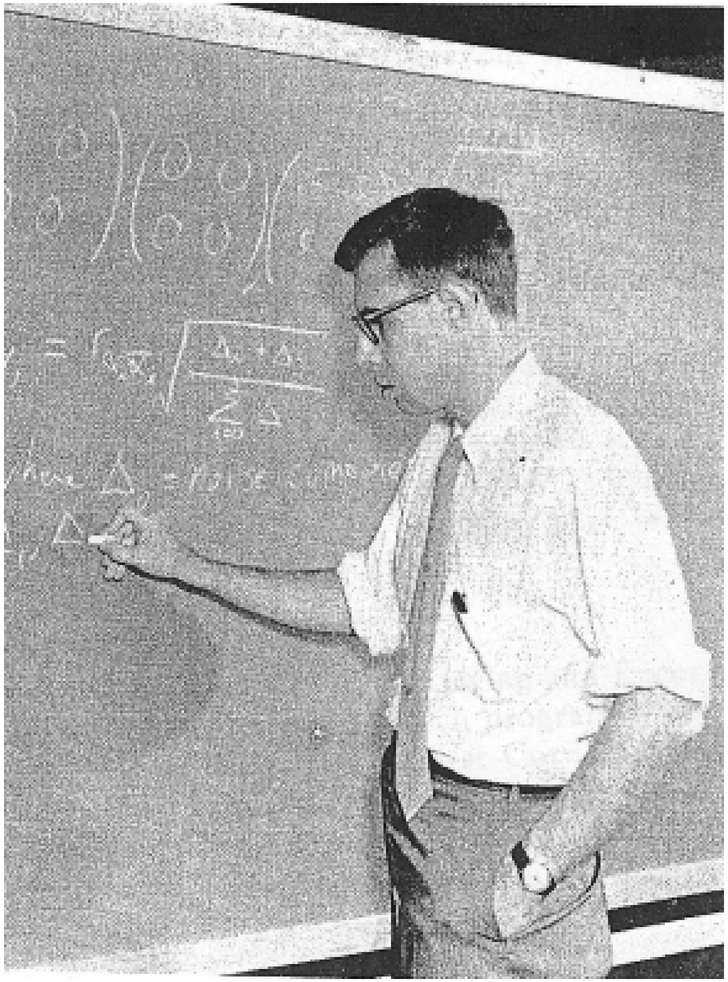
Frank Rosenblatt, my distinguished advisor

George Nagy
PhD 1962, Cornell



Frank Rosenblatt, my distinguished advisor

George Nagy, PhD Cornell 1962

Abstract - Frank Rosenblatt (1928-1971) pioneered the connectionist approach for explaining how biological systems sense, process, organize and use information. He based his models on the results of previous research on neurophysiology and on published observations of the performance of cognitive tasks. His models of diverse aspects of biological information processing, that he called perceptrons, consisted of networks of fixed connections, threshold units, and adaptive non-linear storage elements. He analyzed the models using mathematical tools (probability and linear algebra), large-scale computer simulations, and via hardware implementations. He demonstrated that various families of perceptrons with parameters assigned according to given probability distributions could learn to discriminate classes of patterns, exhibit selective attention, associate geometrically similar and temporally contiguous patterns, and recall entire sequences of sensory input. In the later stages of his career, he conducted experiments on rats that sought to demonstrate the biochemical bases of learning. He published his results in venues ranging from technical reports through conference proceedings to scientific and technical journals like Psychological Review, Science, Nature, Reviews of Modern Physics, and the Proceedings of the IRE. He summarized his results up to 1961 in the book Neurodynamics. Some of his publications are still widely cited. In addition to his influential contributions to biological information processing, Rosenblatt is credited with the invention of trainable artificial neural networks that have found many applications far from the realm of biology and neurophysiology. This aspect of his work is commemorated by the Frank Rosenblatt Award of the IEEE Neural Networks Society. As an amateur astronomer, he built his own observatory and in 1971 proposed a method of detecting planetary systems that is now pursued by the NASA. He was an accomplished musician, mountain climber, and sailor, He participated in national political campaigns and in faculty governance at Cornell University. This presentation was prepared for a small symposium at Pace University dedicated by his former students to Frank's Rosenblatt's achievements.

TOC

MCMXXVIII – MCMLXX1

Epochs

A.B. 1950
Social
Psychology

- Bronx High School of Science
- Cornell student (1946 – 1956)
- Cornell Aeronautical Laboratories
- Cognitive Systems Research Program
- Neurobiology

political campaigns in NY, NH, VT, CA

music (piano, composition)

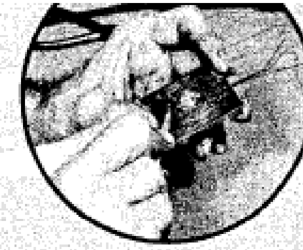
astronomy and cosmology

mountain climbing and sailing

Publications

- Parallax and Perspective during Aircraft Landings, *The American Journal of Psychology* (with **J.J. Gibson**, & P. Olum), 1955
- The K-coefficient Design and Trial Application of a New Technique for Multivariate Analysis – PhD Thesis, Cornell, 210 pages, 1956
- The Perceptron, A **P**erceiving **a**nd **R**ecognizing **A**utomaton (Project Para), CAL Rept. No. 85-460-1, January 1957 EPAC
- The perceptron: A probabilistic model for information storage and organization in the brain, *Psychological Review*, 1958
- **Principles of Neurodynamics,** 616 pages, Spartan, 1962
- A two-color photometric method for detection of extra-solar planetary systems, *Icarus*, 1971
- ~ 30 publications in *Rev. Mod. Phys., Science, Nature, Procs. IRE,* ...
- ~40 technical reports, memoranda, and patents

Vol. VI, No. 2, Summer 1958



research trends

CORNELL AERONAUTICAL LABORATORY, INC., BUFFALO 21, NEW YORK

The Design of an Intelligent AUTOMATON

by FRANK ROSENBLATT

EDITOR’S N OTE

Because of the unusual significance of Dr. Rosenblatt’ s article, Research Trends is proud to devote this entire issue to it.

Introducing the perceptron — A machine which senses, recognizes, remembers, and responds like the human mind.

Manifesto (1958)

perceptron models

1. are economical (relatively few units)
2. are robust (to failure of particular units)
3. don’t *memorize* inputs
4. don’t violate known information about the CNS
5. are capable of spontaneous organization and symbolization of their environment

The book: pp. 382-383

$$P_{jk} = \begin{cases} p & j \leq K, \quad k = K+j \\ p & j > K, \quad k = j-K \\ (1-p)/(2K-1) & j \leq K, \quad k \neq K+j \\ (1-p)/(2K-1) & j > K, \quad k \neq j-K \end{cases}$$

Let $w = p-(1-p)/(2K-1)$; then the P_{jk} can be expressed as follows. For $1 \leq j \leq K$, $1 \leq k \leq K$, we have

$$P_{jk} = P_{j+K,\, k+K} = r$$

$$P_{j,\, K+k} = P_{j+K,\, k} = r + w\delta_{jk}$$

where $r = (1-w)/2K = (1-p)/(2K-1)$. This means that the transition probability from a stimulus to its transform, or vice versa, is $r + w$, while for any two unrelated stimuli, the transition probability is r .

Then from (16.16) we have

$$\gamma^{(i)} = \frac{N_a \eta}{2K\delta}\left[\sum_{j=1}^{K}\sum_{k=1}^{K} + \sum_{j=1}^{K}\sum_{k=K+1}^{2K} + \sum_{j=K+1}^{2K}\sum_{k=1}^{K} + \sum_{j=K+1}^{2K}\sum_{k=K+1}^{2K}\right] Q_{ij}^{(1)} P_{jk}\, \phi(\alpha^{(k)})$$

$$= \frac{N_a \eta}{2K\delta}\left[\sum_{j=1}^{K}\sum_{k=1}^{K} Q_{ij}^{(1)} P_{jk}\, \phi(\alpha^{(k)}) + \sum_{j=1}^{K}\sum_{k=1}^{K} Q_{ij}^{(1)} P_{j,\,k+K}\, \phi(\alpha^{(k+K)})\right.$$

$$\left. + \sum_{j=1}^{K}\sum_{k=1}^{K} Q_{i,\,j+k}^{(1)} P_{j+k,\,k}\, \phi(\alpha^{(k)}) + \sum_{j=1}^{K}\sum_{k=1}^{K} Q_{i,\,j+K}^{(1)} P_{j+K,\,k+K}\, \phi(\alpha^{(k+K)})\right]$$

Assuming $S_i \in \{S_1, \ldots, S_K\}$ we have

$$\gamma^{(i)} = \frac{\eta}{2K\delta}\sum_{j=1}^{K}\sum_{k=1}^{K}\Big[(q+s\delta_{ij})\big(r\phi(\alpha^{(k)}) + (r+w\delta_{jk})\,\phi(\alpha^{(k+K)})\big)$$

$$+ q\big((r+w\delta_{jk})\,\phi(\alpha^{(k)}) + r\phi(\alpha^{(k+K)})\big)\Big]$$

$$\gamma^{(i)} = \frac{\eta}{2K\delta}\sum_{j=1}^{K}\sum_{k=1}^{K}\Big\{2qr\big[\phi(\alpha^{(k)}) + \phi(\alpha^{(k+K)})\big] + qw\delta_{jk}\big(\phi(\alpha^{(k+K)}) + \phi(\alpha^{(k)})\big)$$

$$+ s\delta_{ij}\big(r\phi(\alpha^{(k)}) + (r+w\delta_{jk})\,\phi(\alpha^{(k+K)})\big)\Big\}$$

$$= \frac{\eta}{2K\delta}(2Kqr + qw + sr)\sum_{k=1}^{K}\big[\phi(\alpha^{(k)}) + \phi(\alpha^{(k+K)})\big] + \frac{\eta s w}{2K\delta}\,\phi(\alpha^{(i+K)})$$

Thus if p (or w) is nearly 1 and s/q is large, S_i will generalize to its transform, and conversely $T(S_i)$ will generalize to S_i .

[illegible]

$$= \frac{\eta}{2K\delta}(2Kqr + qw + sr)\sum_{k=1}^{K}\big[\phi(\alpha^{(k)}) + \phi(\alpha^{(k+K)})\big] + \frac{\eta s w}{2K\delta}\,\phi(\alpha^{(i+K)})$$

[illegible] specific form of the conditions for such generalization to occur, [illegible] the term for $k = 1$ in $\sum_{k=1}^{K}$ and put it with the second term. This [illegible] required inequality,

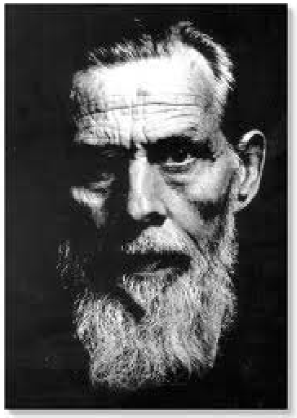
$$(\eta/2K\delta)(2Kqr + qw + sr + sw) \geq \theta$$

The quest

1. How is information about the physical world sensed by the biological system?
2. In what form is information stored and retrieved?
3. How does remembered information influence recognition and behavior?

Rosenblatt chose a genotypic approach to study these problems, through *perceptrons* (“a general class of networks, which include the brain as a special case”). He demonstrated that his biologically-plausible models could mimic various aspects of biological information processing.

Influences



- Bullock
- Cahal
- Clark and Farley
- Culbertson & Hayek
- Eccles
- Kohler
- Holland
- **Hubel and Wiesel (1959)**
- Lettvin and Maturana
- **McCulloch and Pitts (1943)**
- Minsky
- Penfield
- Rashevsky
- Rochester
- Uttley
- von Neumann
- **Hebb (1951)**

Biological information processing system

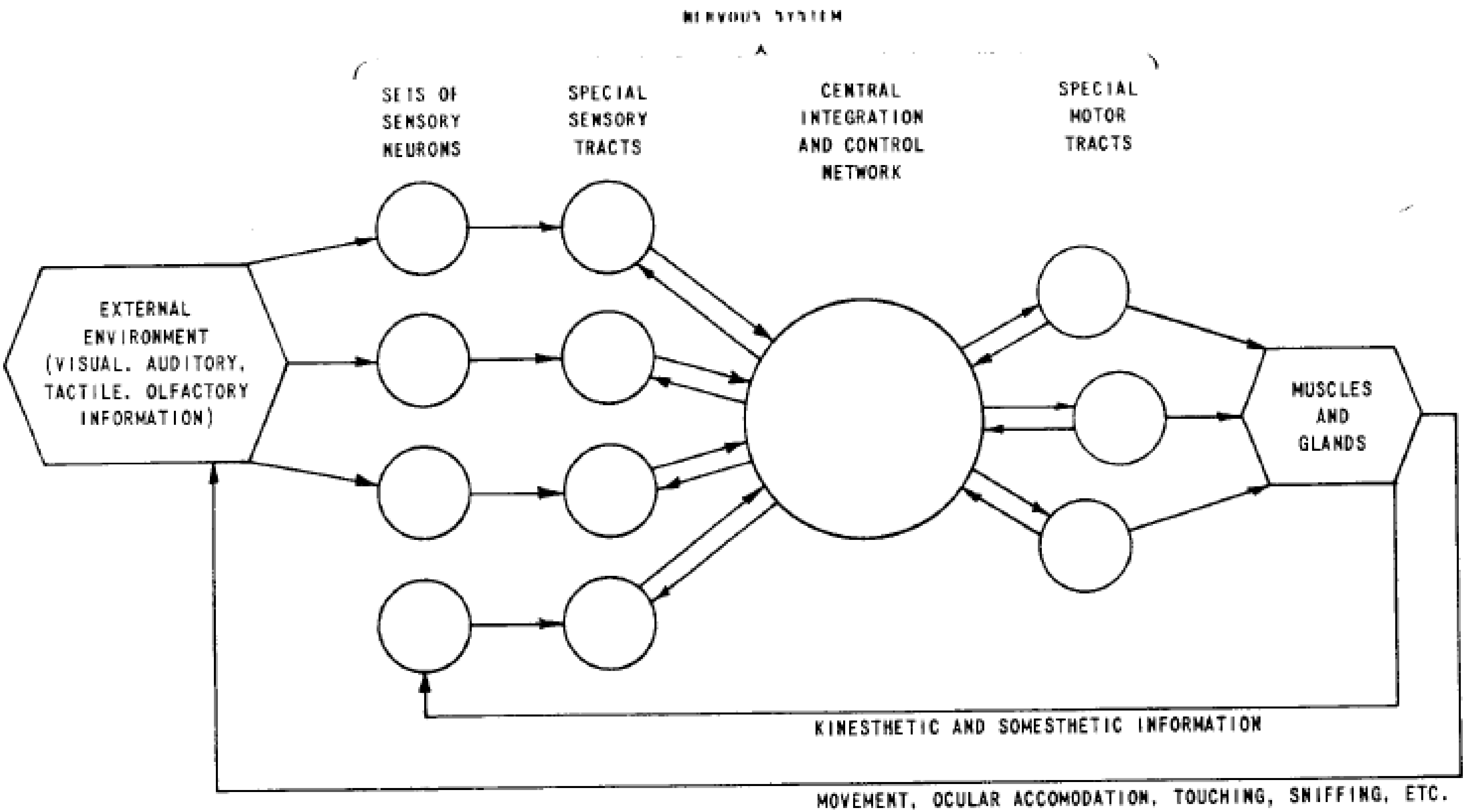


Figure 1 BASIC TOPOLOGICAL STRUCTURE OF THE NERVOUS SYSTEM AND ITS SOURCES OF INFORMATION

Biological information processing system (II)

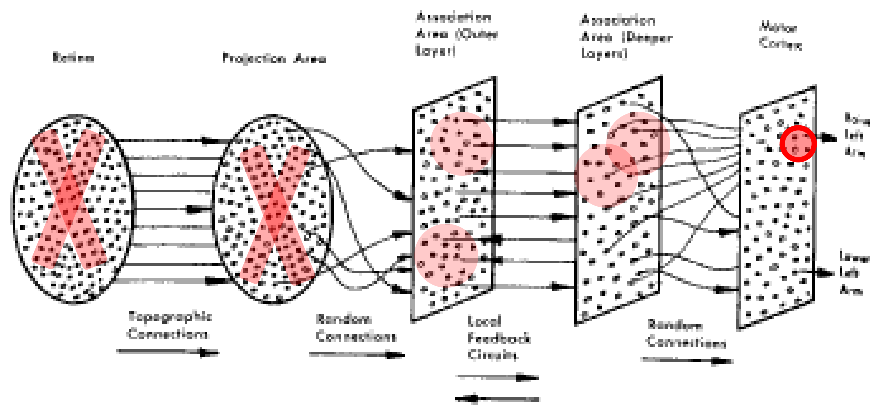


FIG. 1 — Organization of a biological brain. (Red areas indicate active cells, responding to the letter X.)

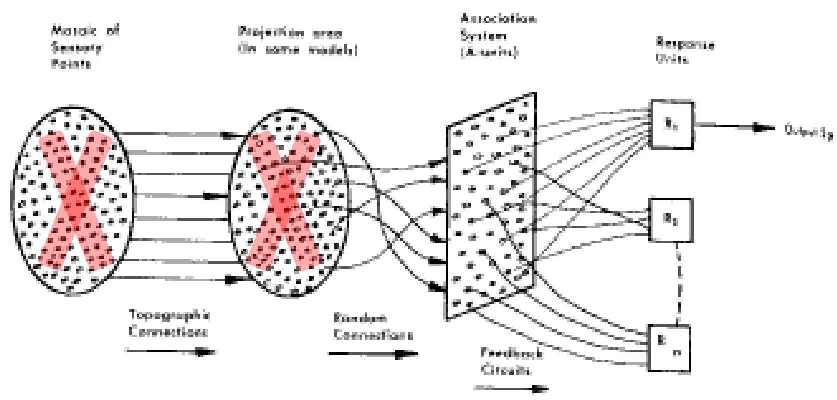


FIG. 2 — Organization of a perceptron.

Learning

"The Mark I perceptron* recognized letters of the alphabet."

This misses the point: commercial OCR was introduced several years earlier.

The point is that Mark I *learned* to recognize letters by being zapped when it made a mistake!

"The perceptron is first and foremost a brain model, not an invention for pattern recognition."

* now in the Smithsonian

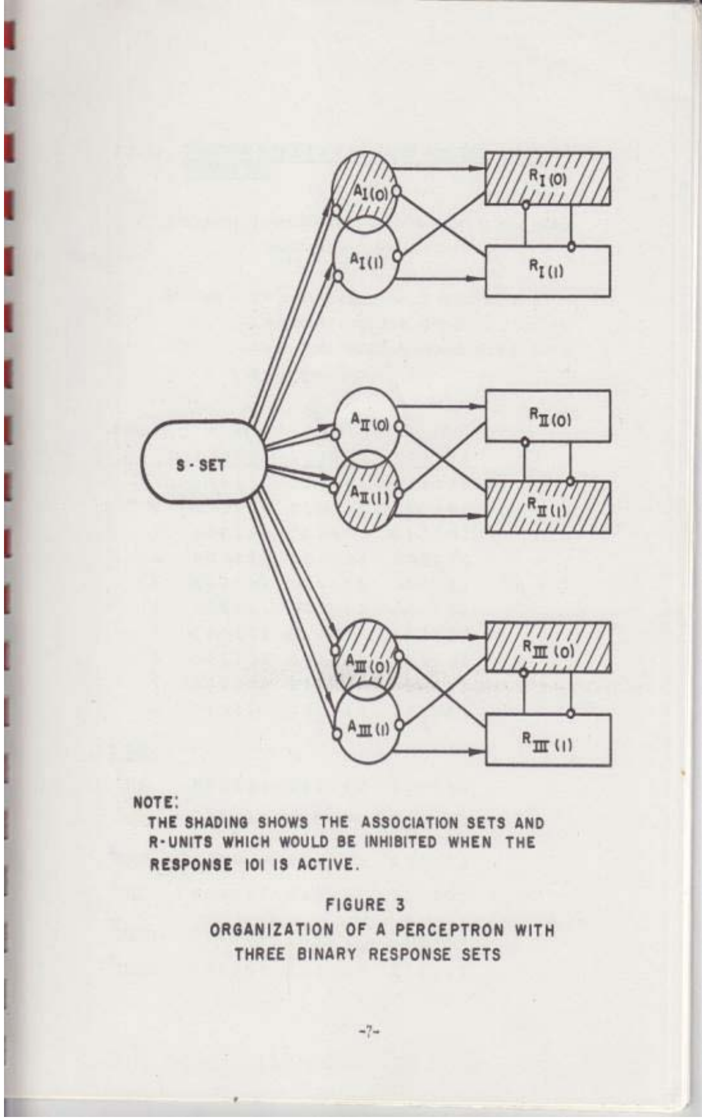
perceptron

Definition 17 (p. 83, Neurodynamics):

A perceptron is a network of S, A, and R units with a variable interaction matrix V which depends on the sequence of past activity states of the network.

Families of perceptrons

- Simple (*fixed S-A*, $u_i = f(\Sigma\alpha_{ij})$)
- Elementary (A, R, f =_⌐)
- Alpha, Gamma
- Universal
- Cross-coupled
- Back-coupled
- N-layer



January 1, 1957

Training

- S-controlled
- R-controlled
- Error-correcting
- Error-correcting, random sign
- Alpha
- Gamma

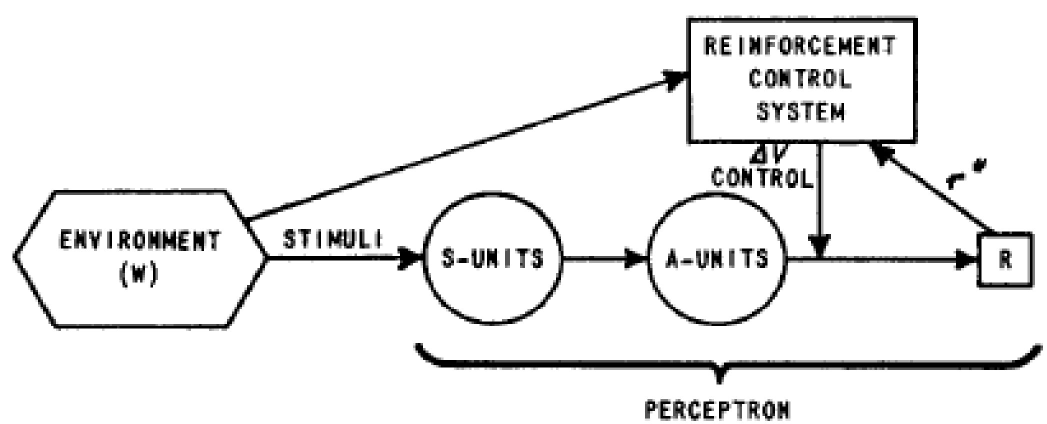


Figure 3 EXPERIMENTAL SYSTEM WITH A SIMPLE PERCEPTRON

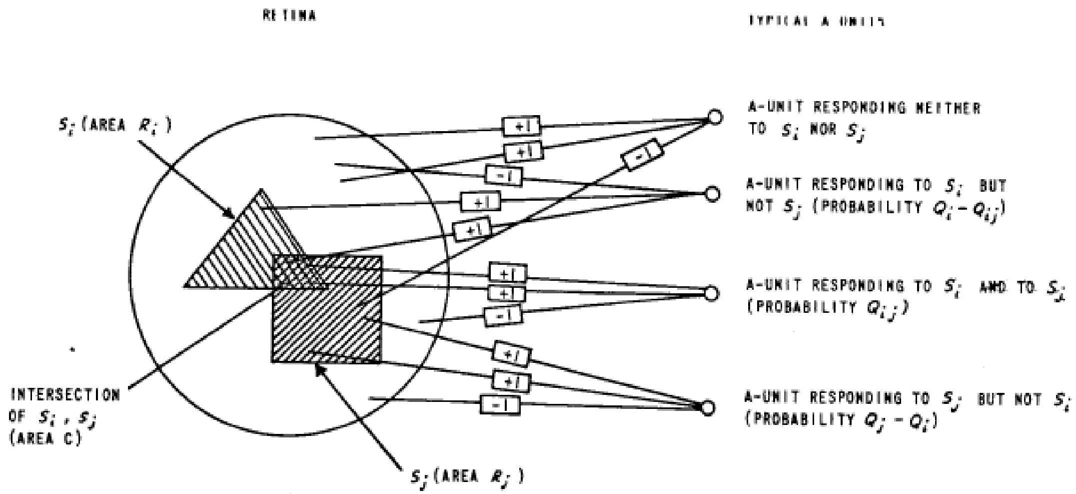
The existence of universal perceptrons

Theorem 1: Given a binary retina, the class of elementary perceptrons for which a solution exists for every classification C(W), of possible environments W, is non-empty.

Proof: By construction, with one A-unit for each pattern, connected by either +1 or -1
to the R-unit.

Corollary: *Three* layers is the minimum

Analysis: G-functions and Q-functions



G is a square matrix with elements g_{ij}

The *generalization coefficient* g_{ij} is the total change in value over all A units in the set responding to S_i if the set of units responding to S_j are each reinforced with η.

$Q_{ij} = E[g_{ij}]$ is the probability that an A-unit in a given class of perceptrons responds to both S_i and S_j.

Convergence

Theorem 4: If a solution C(W) exists, and if each stimulus reoccurs in finite time, then an error-correction procedure, beginning in an arbitrary state, will always yield a solution to C(W) in finite time.

Proof: by contradiction via G-functions and Schwarz's inequality.

Bound

$$N \leq (k + M/\sqrt{n})^2 \, / \, \alpha M$$

where

n is the number of stimuli

M is the maximum diagonal element of H

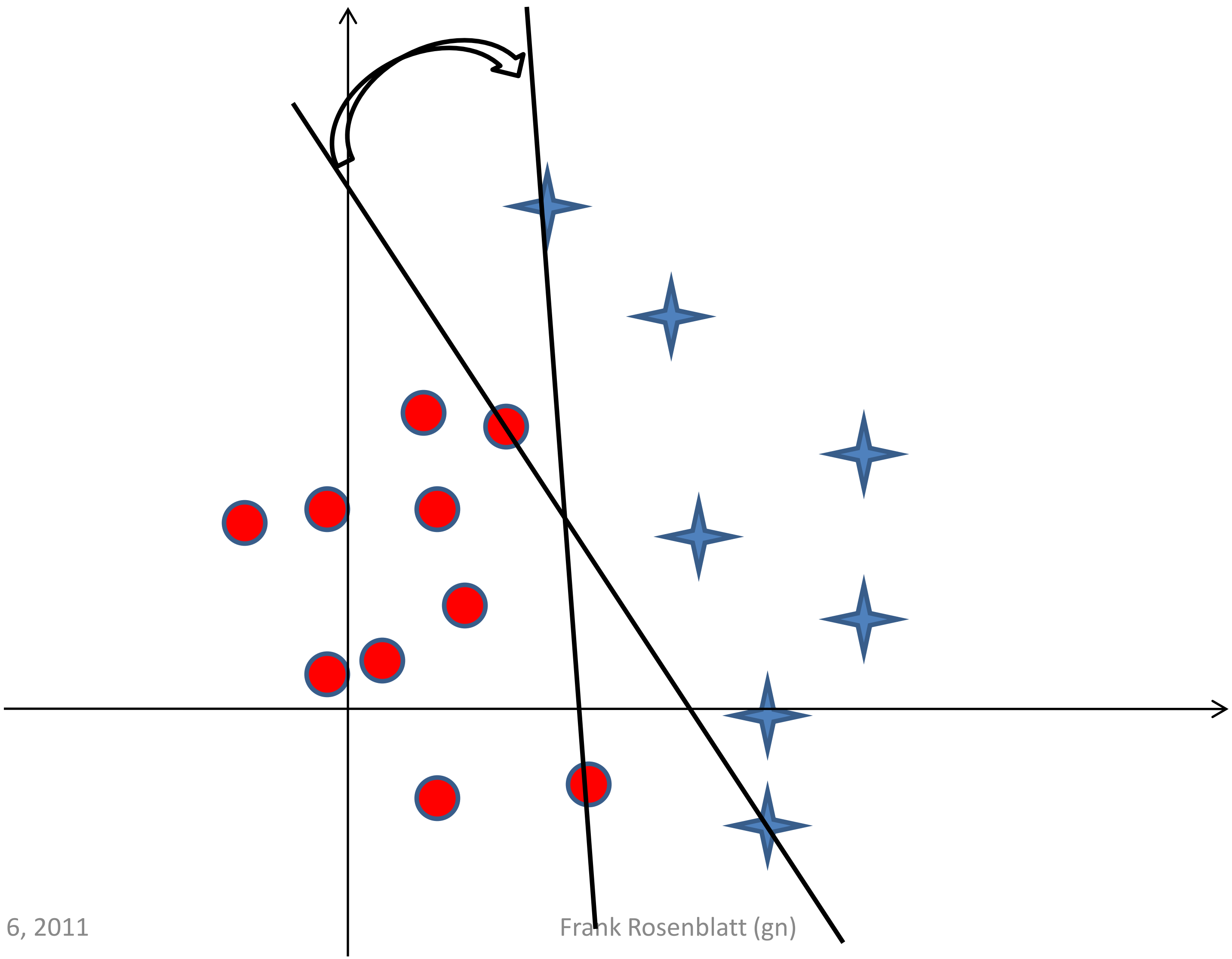
$H = D\ G\ D = BB'$ where $b_{ij} = \rho_i\, a_j{*}\ (s_i)$

$\alpha > 0$ is such that $x'Hx \geq \alpha|H|^2$

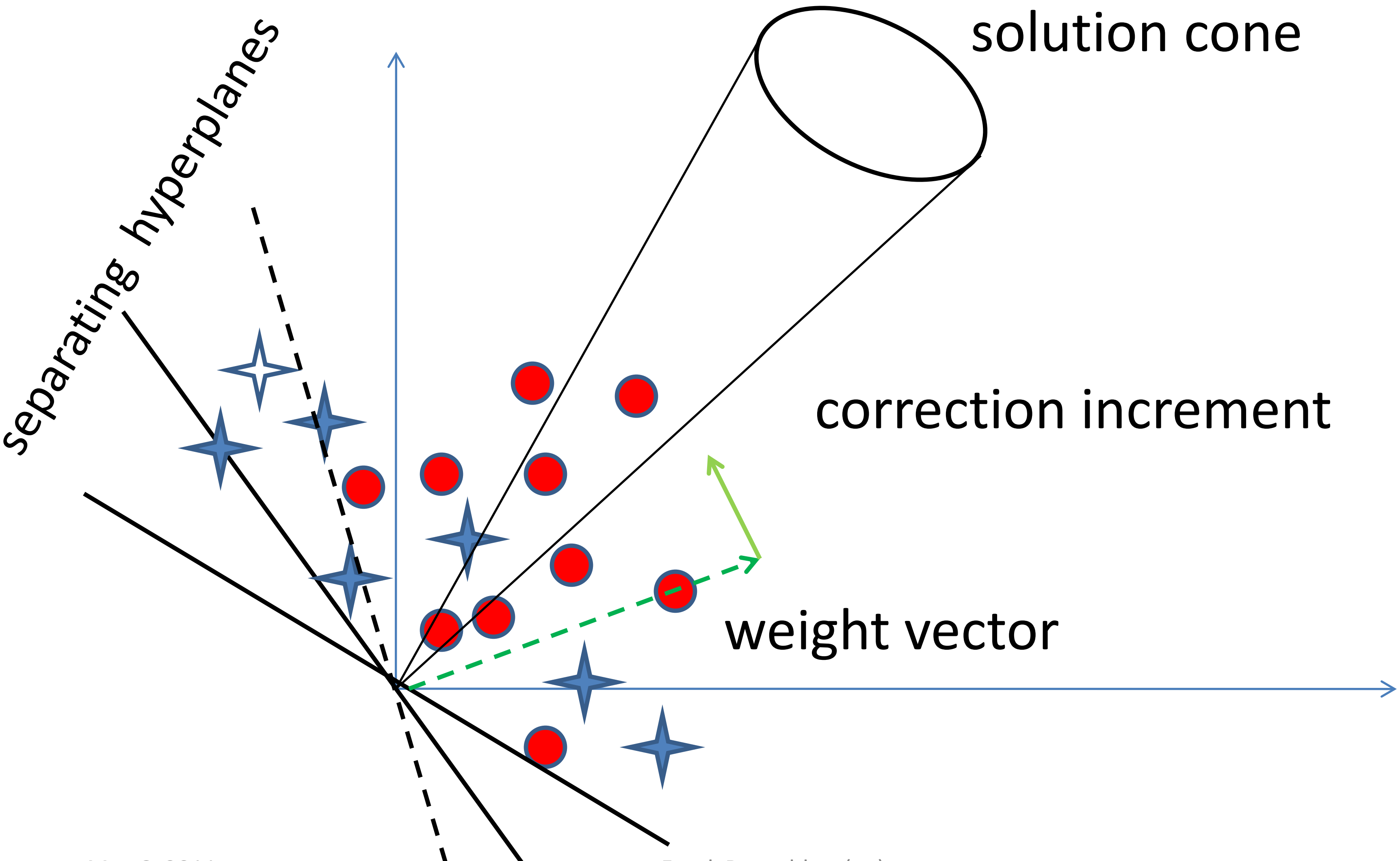
k such that $|Hx^\circ|\cdot|x| = k|x|$ when

x is the total reinforcement vector

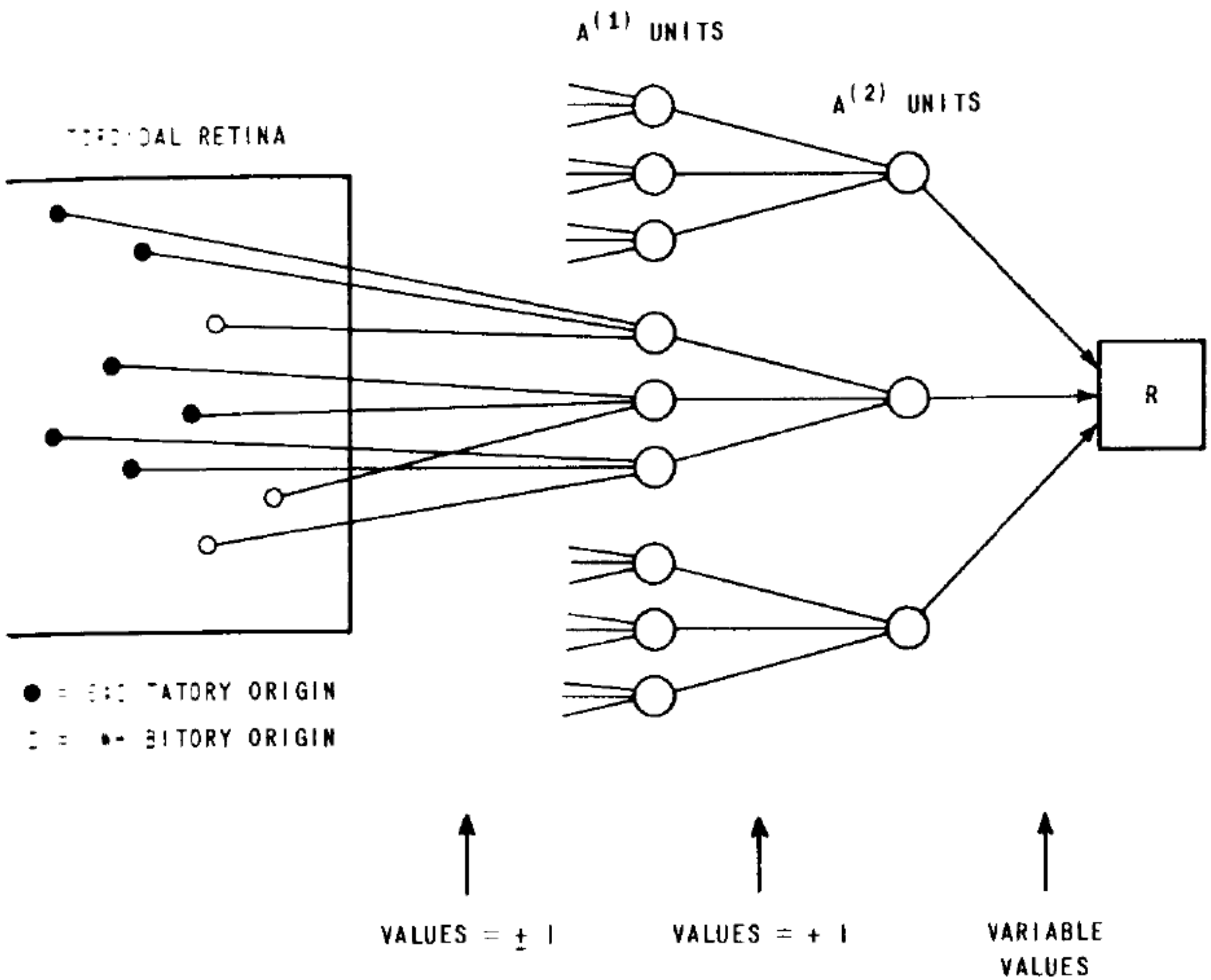
Linear separability in 2-D



A “modern” view



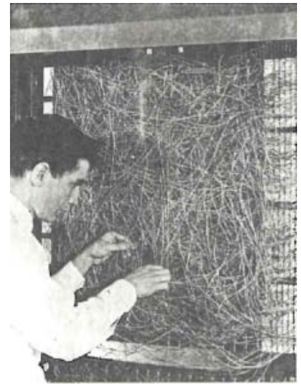
Four-layer similarity-constrained perceptron (cf. Hubel and Wiesel)

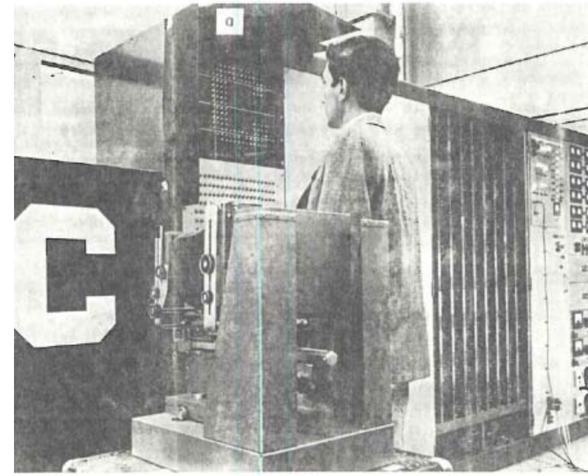


Machines

Mark I: 400 S-units,

512 A-units, 8 R-units





Tobermory:

45 band-pass filters

80 difference detectors

80x20 = 1600 A^1 units,

1000 A^2-units

12 R-units

12,000 weights between A^2 and R layers

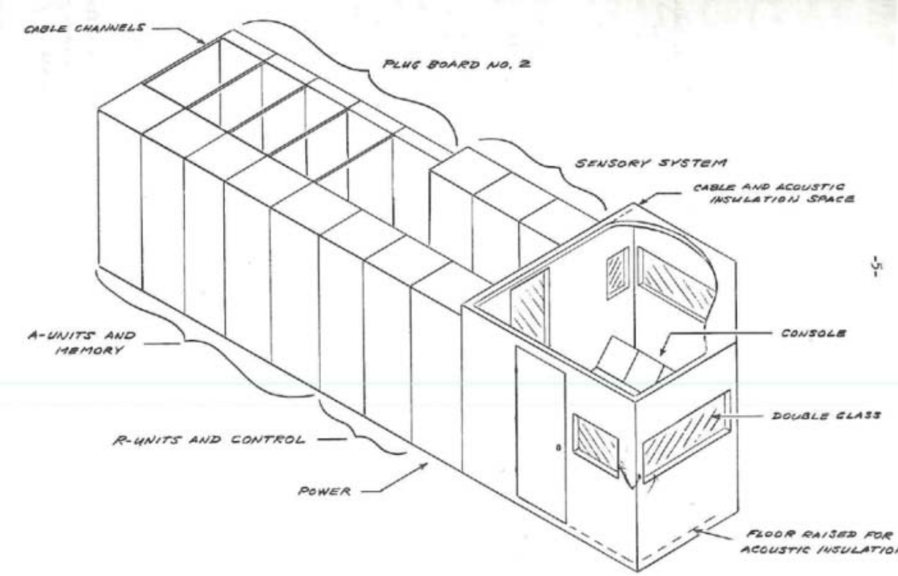
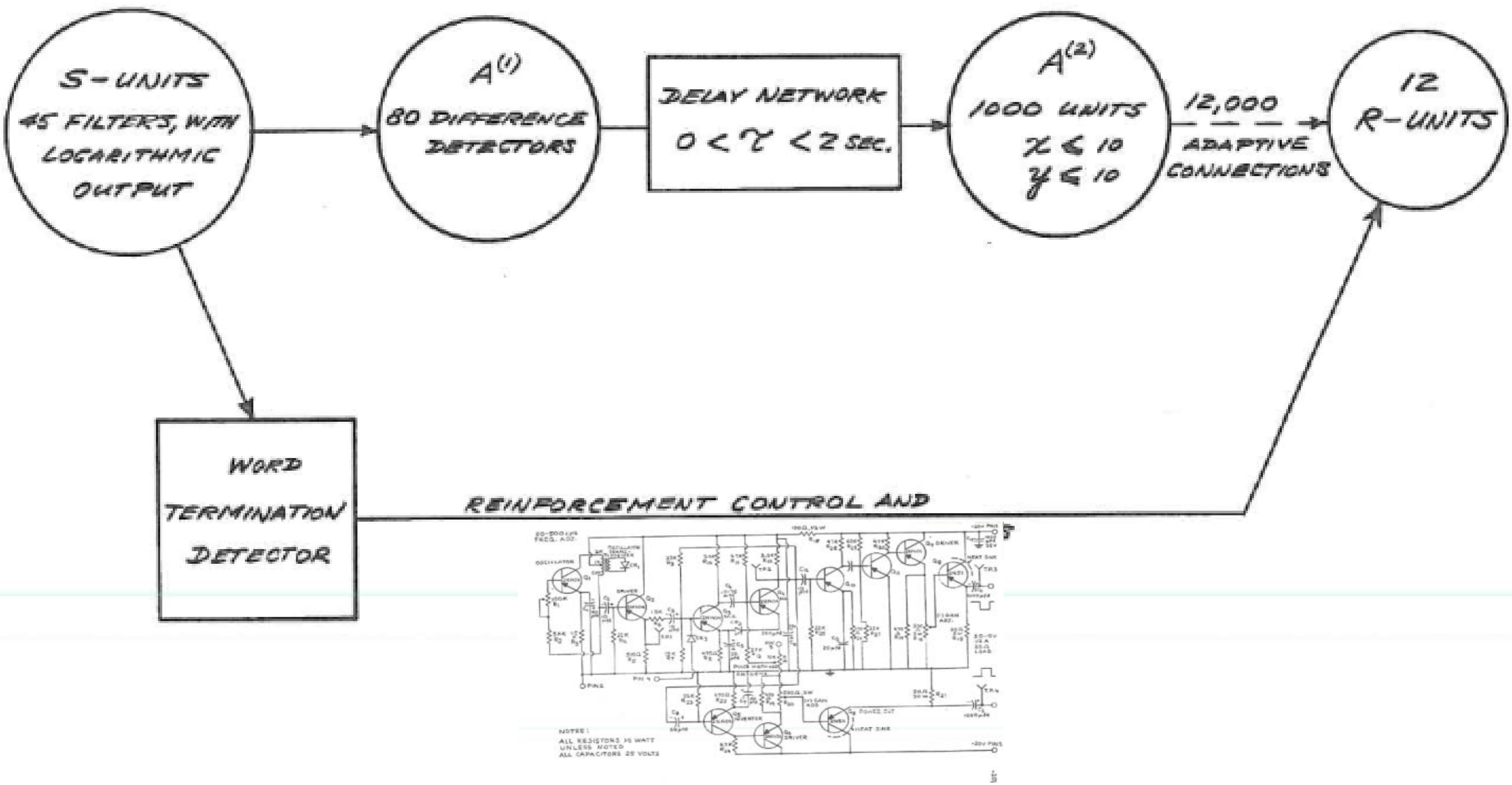
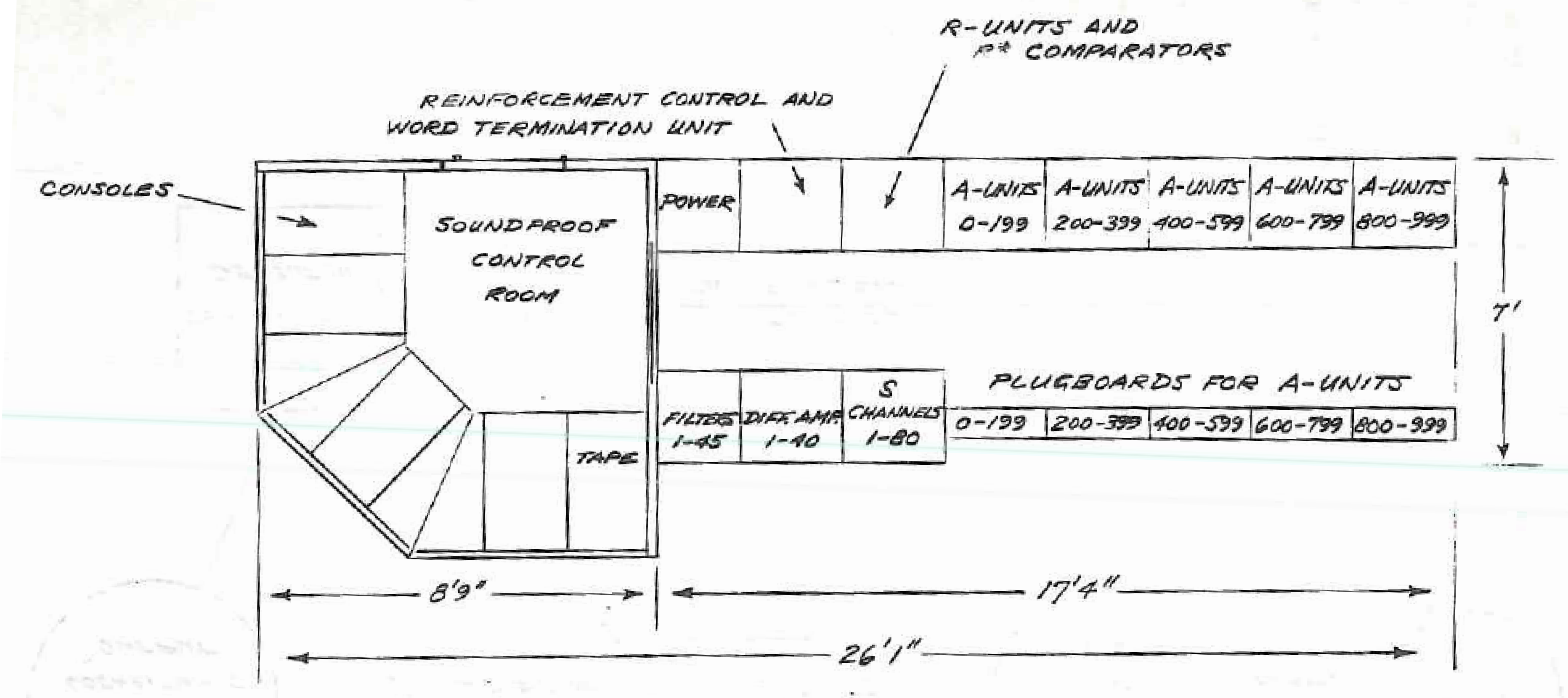


FIG. 0.0.3: ISOMETRIC VIEW OF TOBERMORY, PHASE I

Tobermory Schematic Diagram



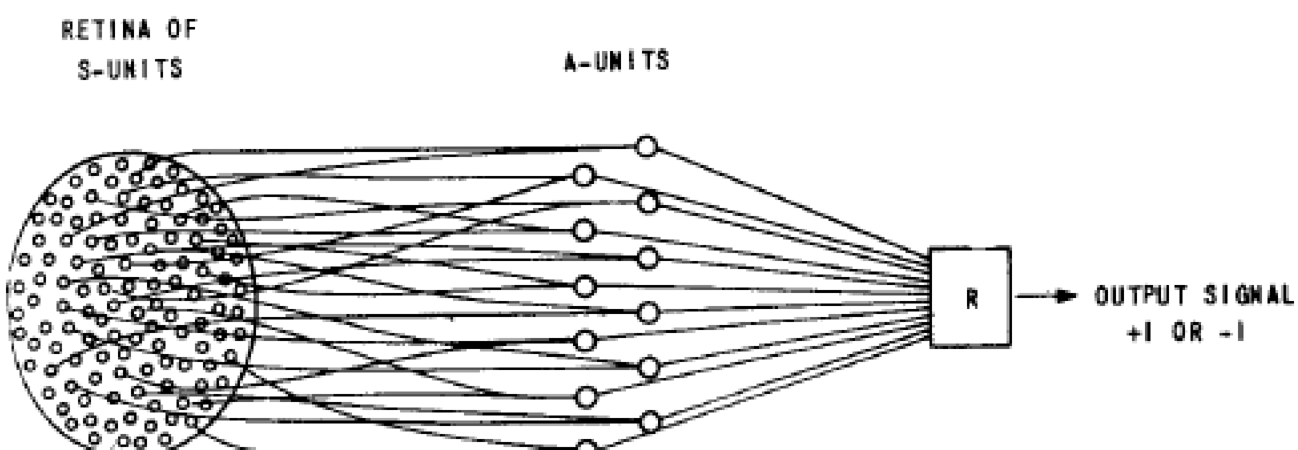
Tobermory Phase I Floor Plan



Backpropagation

There is no indication that Rosenblatt ever considered a gradient–descent algorithm with a quadratic cost function. If he had, he may well have found an argument for its biological plausibility.

(Backpropagation is credited to Bryson and Ho in 1969, but it was not until the mid-seventies that it was popularized by Werbos, Rumelhart †, and Hinton.)



Hardware vs. software

- Most of the ONR funding went to machine construction.
- Most of the significant results were established by simulation experiments on large general purpose computers at Cornell Aerolabs, Rome AF Research Center, NYU, & Cornell Medical Center.
- This has been typical in the history of Artificial Intelligence, Pattern Recognition, Image Processing, and Computer Vision.

perceptron milestones

Error-correcting training procedure	
Proof of convergence	(upper bound)
Universal perceptrons	(can learn any dichotomy)
Geometrically constrained networks	
	- similarity generalization and cat's cortex
Unsupervised training	(cross-coupled perceptrons)
	- generalization from temporal proximity
Selective attention	back-coupled perceptrons
Fully cross-coupled perceptrons	(Hopfield nets?)
Multilayer learning	(phonemes/words)
Elastic perturbation reinforcement	(simulated annealing?)
Sequence recapitulation	(probabilistic C-system)

Legacy

MRI, DOT, and IEAs have improved our knowledge of brain topology and function, and inform current models of the brain as a network of *adaptive* non-linear elements. “Connectionist” research remains active and productive. Today, Frank Rosenblatt’s pioneering efforts are widely acknowledged (over 2800 citations of his 1958 article). Nevertheless, many of the problems of bio-info processing that he studied remain unsolved.

ANNs are also exploited in Machine Learning and have found many applications far from biology.

Magnetic Resonance Imaging, Diffuse Optical Tomography, Implanted Electrode Arrays

MCMXXVIII + C = MMXXVIII

111 1000 1000 + 110 0100 = 111 1110 1100

Frank Rosenblatt

Centennial Symposium

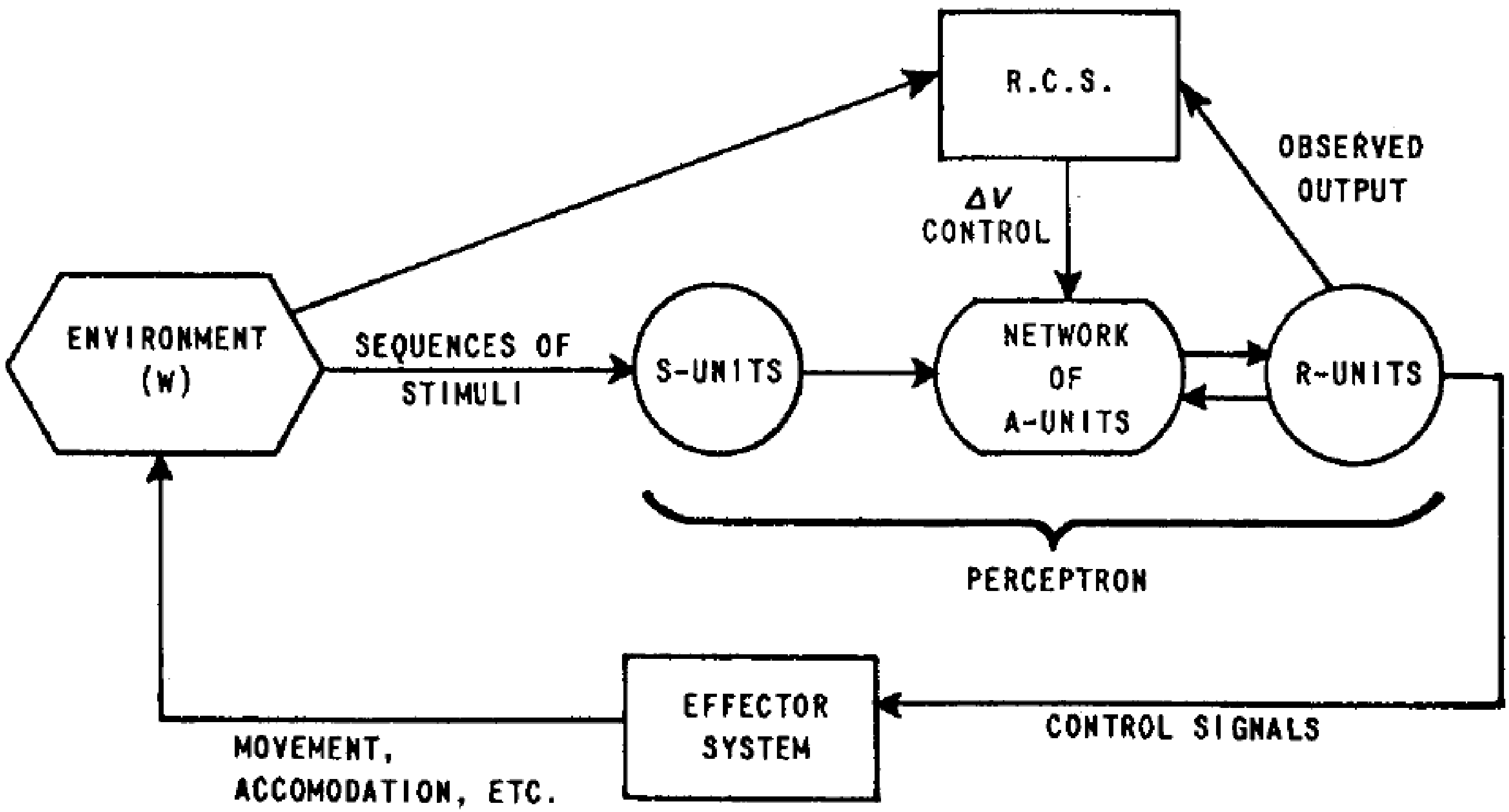
May 18-19, 2028

Cornell University, Ithaca, NY

CFP forthcoming: please register early

Thank you

General experimental system



Tobermory's Sensory Analyzer

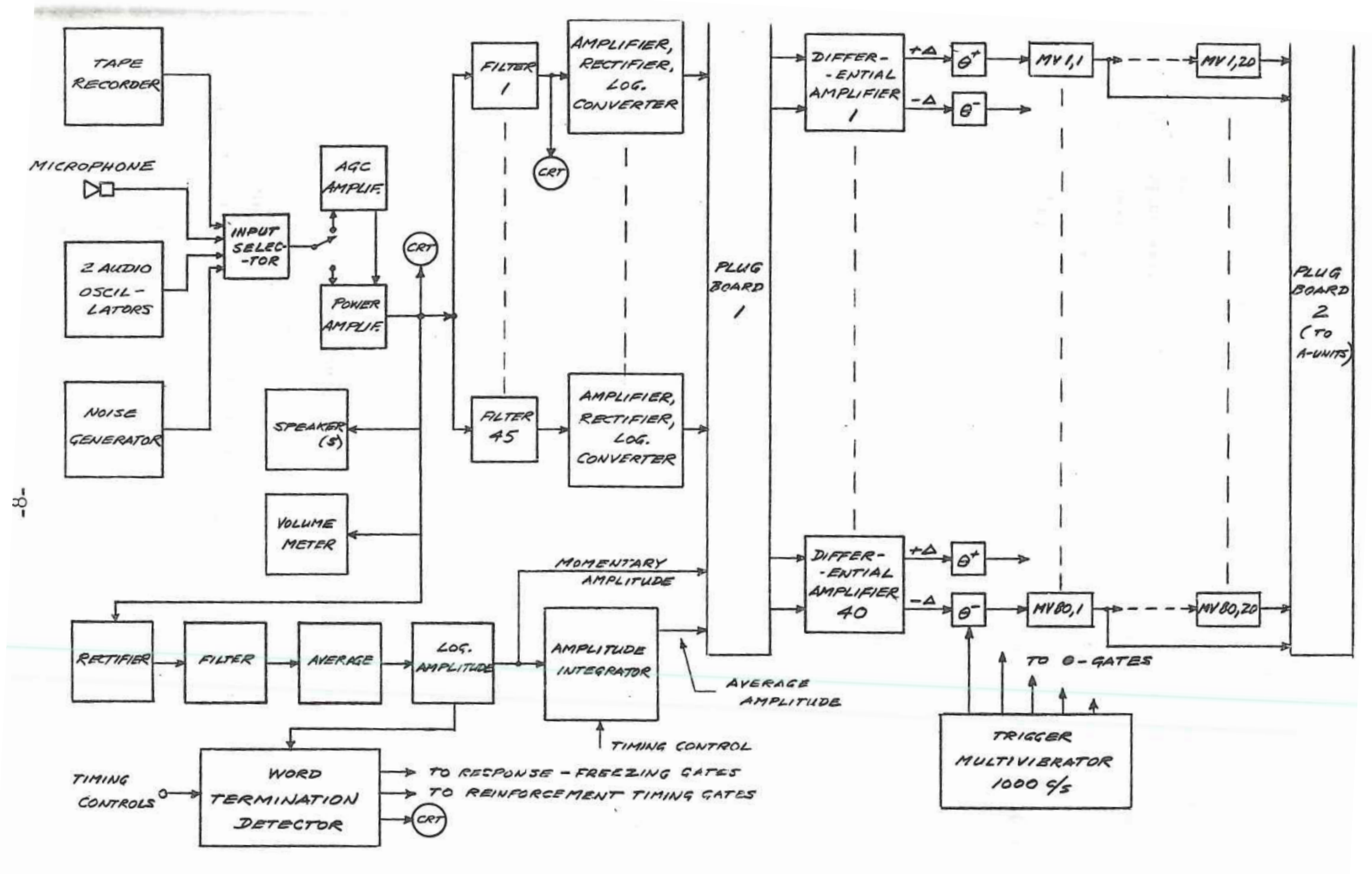
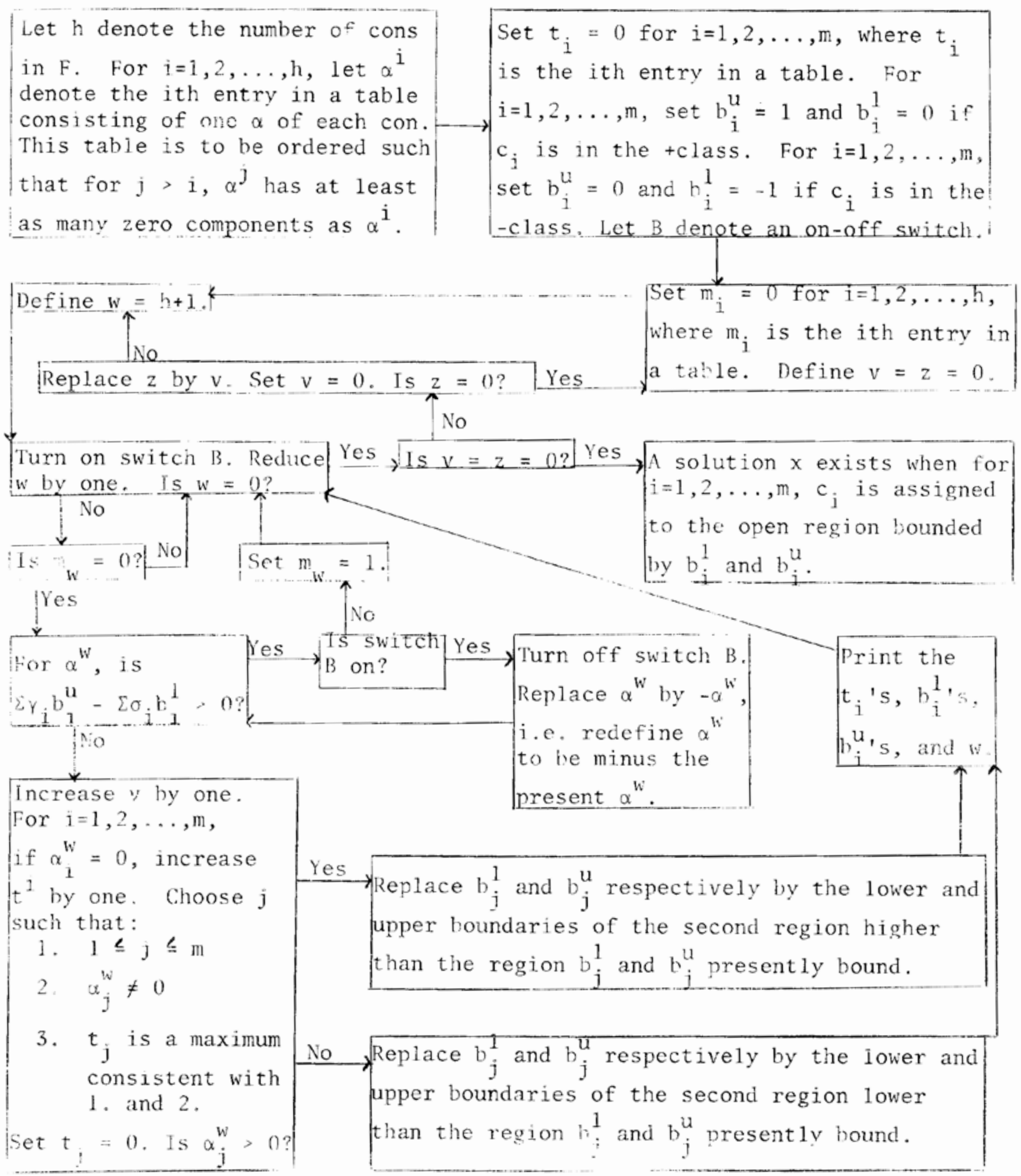


FIG. 1.0.1: THE SENSORY ANALYSER

Figure 1. Computer flow diagram used in Experiment 1



CarlKessler

J.J. Gibson (1904-1975)

One of the most important twentieth century psychologists in the field of visual perception.

Princeton dissertation on memory and learning.

Fulbright scholar at Oxford, fellow of the Institute of Advanced Study at Princeton University,

Member of the National Academy of Science.

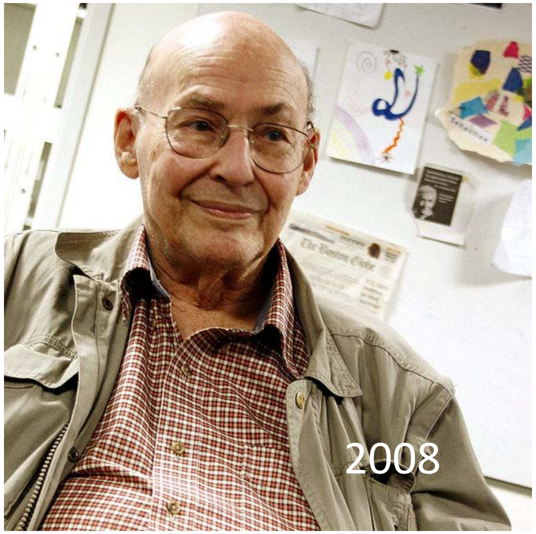
In WW II headed AF Research Unit in Aviation Psychology.

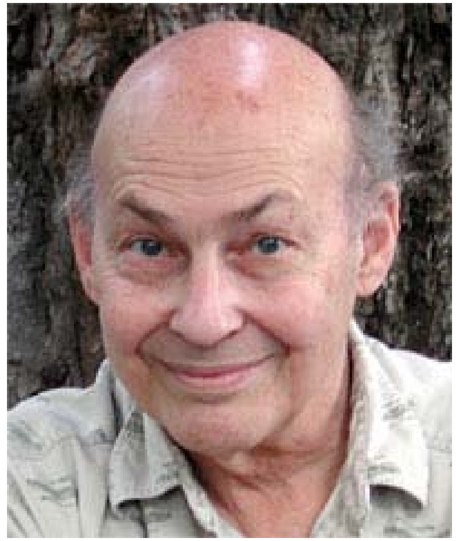
Books, incl. The Perception of the Visual World (1950).

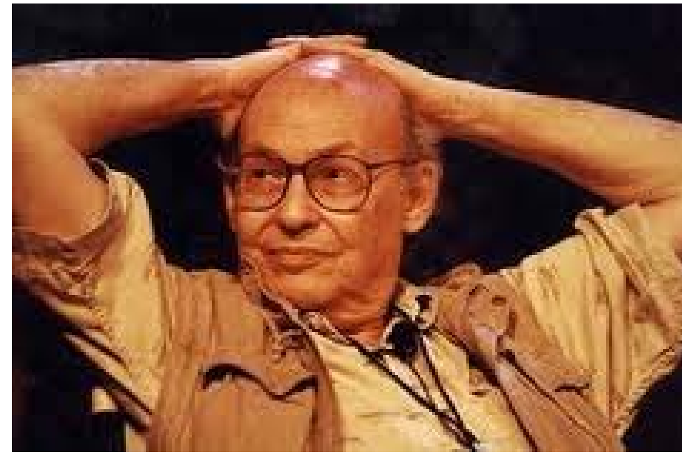
Coined the term *affordance*.

Marvin Minsky (1927 -)

In 1951 he built the SNARC, the first neural network simulator.







Symbolic vs. Connectionist 1990:

> Why is there so much excitement about Neural Networks today, and how is this related to research on Artificial Intelligence? Much has been said, in the popular press, as though these were conflicting activities. This seems exceedingly strange to me, because both are parts of the very same enterprise. What caused this misconception?