

The Perceptron: A Probabilistic Model For Information Storage and Organization in the Brain.

F. Rosenblatt (1958)

Presented by
Manu Reddy, Sep 9 2015

Outline

- Questions for understanding perceptual recognition, recall and thinking.
- Assumptions for the Perceptron
- Organization of the Perceptron
- The Perceptron's response to Stimuli – analysis.
- H.D Block's proof for perceptron – results.

Two important questions.

To understand Perceptual recognition, we must answer

- How is information about physical world sensed by biological system.
- In what form is it stored.
- How does information stored influence recognition and behaviour.

Rosenblatt discusses the 2nd and 3rd questions.

Question 2: In what form is information stored?

There are two alternative positions to this question.

1) Coded representation or images (Coded Memory Theorists)

- One-to-one mapping between the sensory stimulus and the stored pattern.

2) Images of stimuli are never stored (Connectionists)

- The information is contained in connections or associations.

How does the stored information influence recognition

Corresponding to the 2 positions on information retention, there are 2 hypothesis to this question.

- 1) Involves pattern matching of the stored contents with incoming sensory patterns.
- 2) Automatically activate the appropriate response based on the new connections or pathways created. No need for pattern matching.

Perceptron

- A hypothetical nervous system to illustrate some of the fundamental properties of intelligent systems.

Language

- The gross organization of the system is known but the precise structure isn't.
- Rosenblatt feels the language of symbolic logic and boolean algebra is not well suited to describe the theory of perceptron.
- Formulated the current model of perceptron using probability theory.

Assumptions for the perceptron model.

- The connections of the nervous systems differ from one organism to another. The construction process of these networks is largely random.
- The system is capable of changing. The probability that a stimulus applied to one set of cells causes a response in another set of cells is likely to change, due to some relatively long-lasting changes in the neurons.
- Over time, similar stimuli will tend to form pathways to the same sets of responding cells and dissimilar stimuli will tend to develop connections to different sets of responding cells.
- Positive and negative reinforcements affects the formations of connections.
- Similarity is represented by nervous system.

Organization of a perceptron

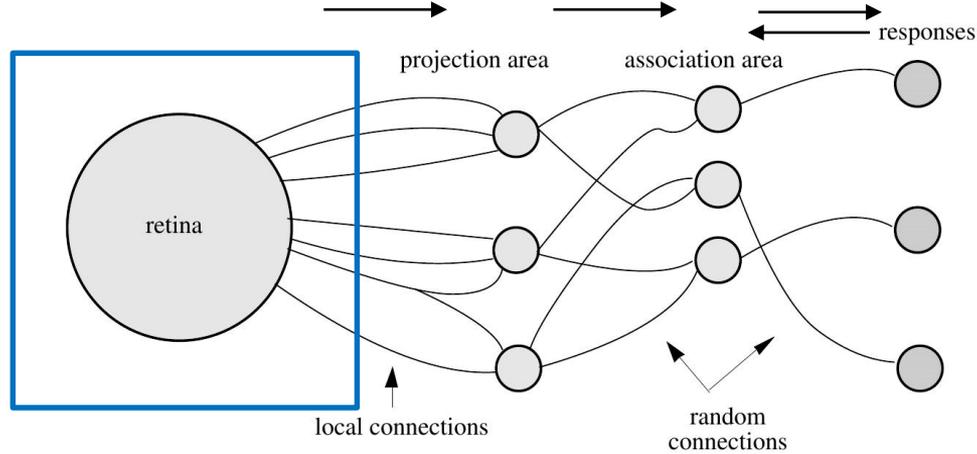


Fig. 3.1. The classical perceptron [after Rosenblatt 1958]

S Points

- Stimuli impinge on the retina of sensory units.
- They respond on an all-or-nothing basis.

Organization of a perceptron

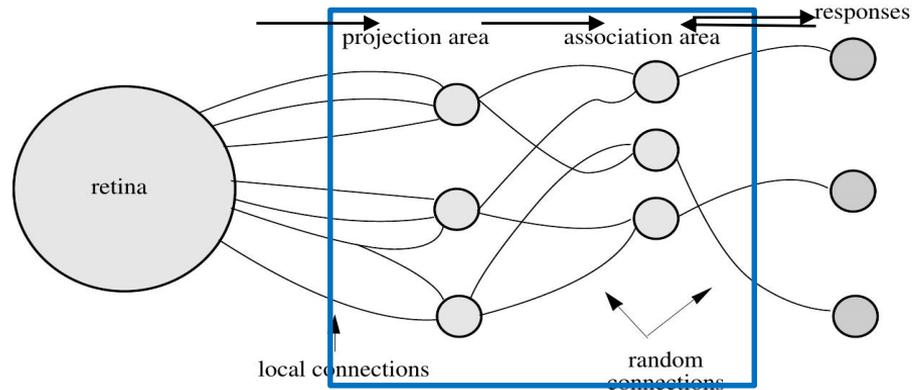


Fig. 3.1. The classical perceptron [after Rosenblatt 1958]

A-Units

- Impulses are transmitted to a set of association cells (A-units) in a “projection area”.
- **Origin points:** The set of S points which transmit impulses to a specific A-unit.
- The connections from these origin points could be either *excitatory* or *inhibitory*.
- The A units fire, if the algebraic sum of the excitatory and inhibitory impulse intensities are greater than the threshold of the A-Unit.

Organization of a perceptron

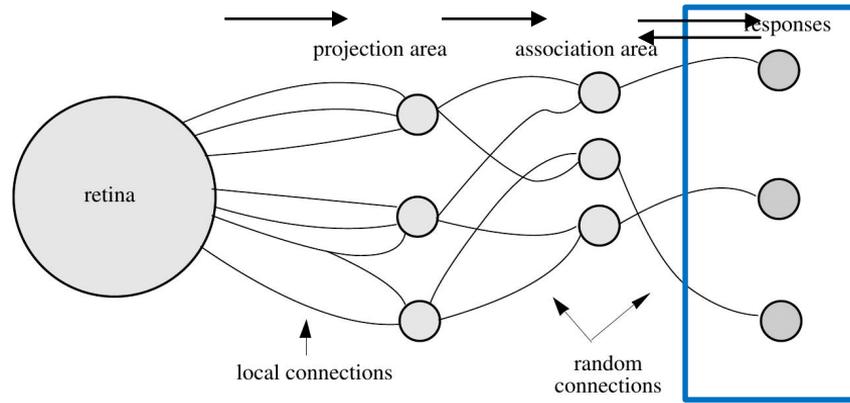


Fig. 3.1. The classical perceptron [after Rosenblatt 1958]

R-cells

The response cells respond in a similar fashion as the A-units.

Source set of a Response:

The set of A-units transmitting impulses to a particular response will be called the source-set for that response.

Note: Connections between the A-units and the R-units are bi-directional (feedback).

Rules governing Feedback connections

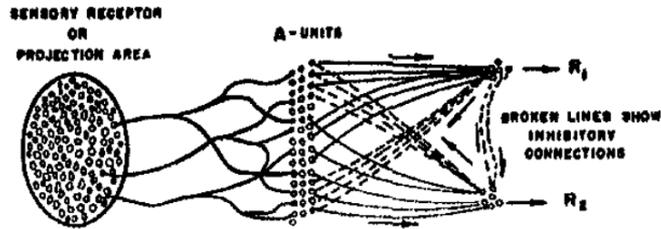


FIG. 2A. Schematic representation of connections in a simple perceptron.

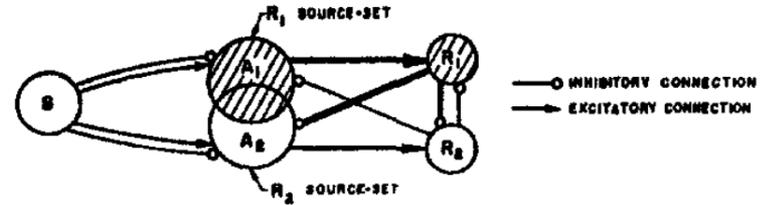


FIG. 2B. Venn diagram of the same perceptron (shading shows active sets for R_1 response).

- Each response has excitatory feedback connections to the cells in its own source set.
- Each response has inhibitory feedback connections to the complement of its own source-set (i.e., it tends to prohibit activity in any association cells which do not transmit to it).

Activation of A-units.

- For learning to be possible, the value of A-units should be modifiable.
- The values of A-units could be amplitude, frequency etc...
- There are 3 models proposed in the paper.
 1. α system: an active cell gains a unit value for every impulse and holds this indefinitely.
 2. β system: an active cell gains a mean value.
 3. γ system: an active cell gains in value, so that total value of source-set is constant.

2 Phases

Predominant Phase

Some proportion of A-units (denoted by solid dots) responds to the stimulus.

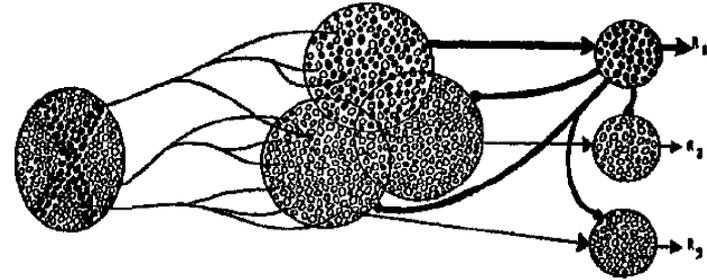


FIG. 3A. Predominant phase. Inhibitory connections are not shown. Solid black units are active.

Post Dominant Phase

One of the responses becomes active, inhibiting activity in the complement of its own source-set, and this prevents other responses from triggering.

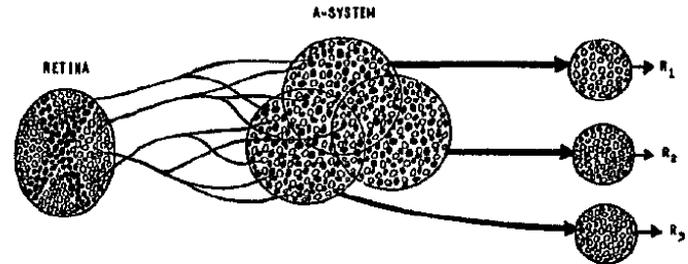


FIG. 3B. Postdominant phase. Dominant subset suppresses rival sets. Inhibitory connections shown only for R_1 .

Analysis of Predominant Phase

P_a : the expected proportion of A-units activated by a stimulus of a given size.

$$P_a = \sum_{e=\theta}^x \sum_{i=\theta}^{\min(y, x-e)} P(e, i) \quad \text{where}$$

$$P(e, i) = \binom{x}{e} R^e (1 - R)^{x-e} \times \binom{y}{i} R^i (1 - R)^{y-i}$$

and

e = excitatory components of the impulse received by A-unit.

i = inhibitory components of the impulse received by the A-units

R = proportion of S-points activated by the stimulus

x = number of excitatory connections to each A-unit

y = number of inhibitory connections to each A-unit

θ = threshold of A-units.

Analysis of Predominant Phase

P_c : the conditional probability that an A-unit which responds to a given stimulus, S_1 , will also respond to another stimulus, S_2 .

$$P_c = \frac{1}{P_a} \sum_{e=\theta}^x \sum_{i=e-\theta}^y \sum_{l_e=0}^e \sum_{l_i=0}^i$$

$$\sum_{g_e=0}^{x-e} \sum_{g_i=0}^{y-i} P(e, i, l_e, l_i, g_e, g_i)$$

$$(e - i - l_e + l_i + g_e - g_i \geq \theta)$$

where

$$\begin{aligned} P(e, i, l_e, l_i, g_e, g_i) &= \binom{x}{e} R^e (1 - R)^{x-e} \\ &\times \binom{y}{i} R^i (1 - R)^{y-i} \\ &\times \binom{e}{l_e} L^{l_e} (1 - L)^{e-l_e} \\ &\times \binom{i}{l_i} L^{l_i} (1 - L)^{i-l_i} \\ &\times \binom{x-e}{g_e} G^{g_e} (1 - G)^{x-e-g_e} \\ &\times \binom{y-i}{g_i} G^{g_i} (1 - G)^{y-i-g_i} \end{aligned}$$

Analysis of Predominant Phase

L = Proportion of the S-points illuminated by the first stimulus, S_1 , which are not illuminated by S_2 .

G = Proportion of the residual S-set which is included in the second stimulus S_2 .

l_e and l_i are number of excitatory and inhibitory origin points lost by the A-unit when S_1 is replaced by S_2 .

g_e and g_i are number of excitatory and inhibitory origin points gained by the A-unit when S_1 is replaced by S_2 .

Analysis of P_a

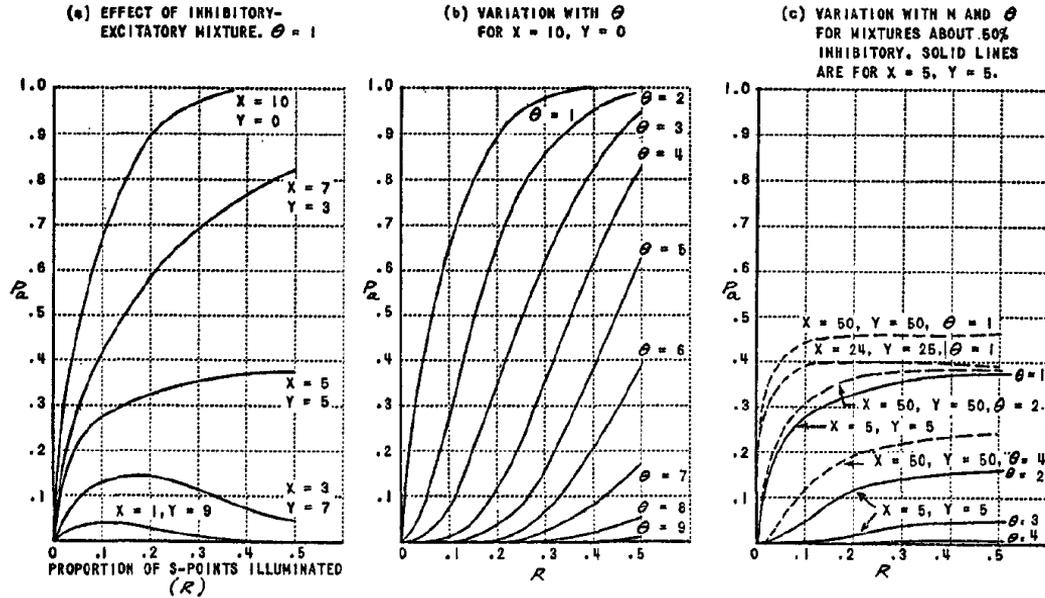


FIG. 4. P_a as function of retinal area illuminated.

- P_a can be reduced in magnitude by either increasing the threshold, θ or by increasing the number of inhibitory connections.
- If excitation is equal to inhibition, then the curves for P_a are flattened out.

Analysis of P_c

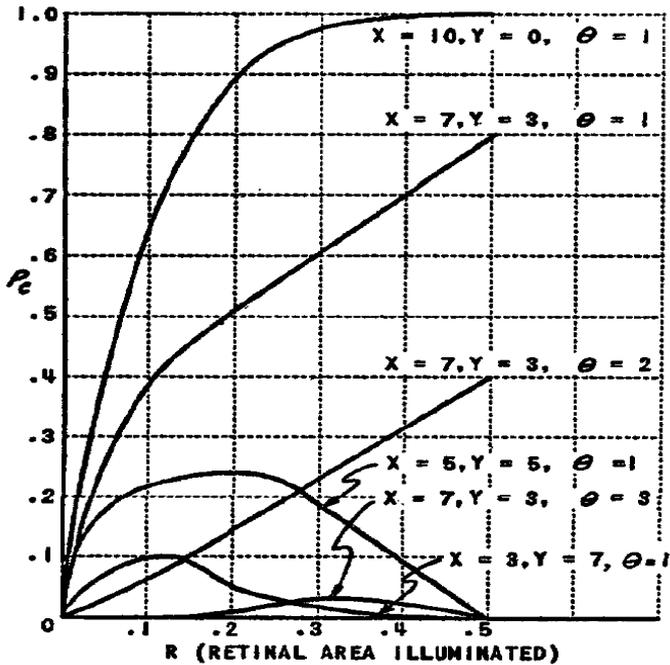


FIG. 5. P_c as a function of R , for nonoverlapping stimuli.

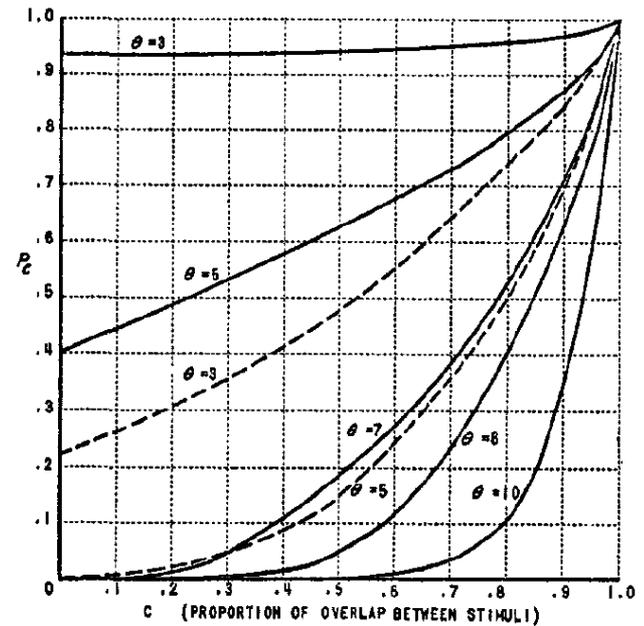


FIG. 6. P_c as a function of C . $X = 10$, $Y = 0$. Solid lines: $R = .5$; broken lines: $R = .2$.

- P_c drops sharply than P_a as threshold is increased.
- P_c also decreases as the proportion of inhibitory connections increases.
- P_c approaches unity as stimuli overlap completely.
- $P_c > 0$, even when stimuli are completely different.

Analysis of Postdominant Phase

- The response of the perceptron in the predominant phase, where A-units respond to the stimulus, gives way to the postdominant phase.
- Two Systems have been proposed.
 - Mean discriminating phase: the response whose inputs have the greatest mean value respond first.
 - Sum discriminating phase: the response whose inputs have the greatest sum or net value respond first.

Methods of evaluation

Experiment 1(exact learning)

- Learn random stimuli

Experiment 2(learning classes)

- Learn classes

Experiment 1

- Expose perceptron to series of stimulus patterns.
- Force it give desired response by applying reinforcements.
- Test the system by presenting the same series of stimuli in an exact fashion as in training.
- Evaluated by P_r , the probability of correct choice of response between two alternatives.

Experiment 2

- Present same series of stimuli as in experiment 1
- Use a different evaluation criteria
- Test on new stimuli drawn from the same classes.
- Evaluated by P_g , the probability of correct generalization, i.e., the probability that the perceptron will give the correct response for the class of stimuli.

Results of Experiment 1

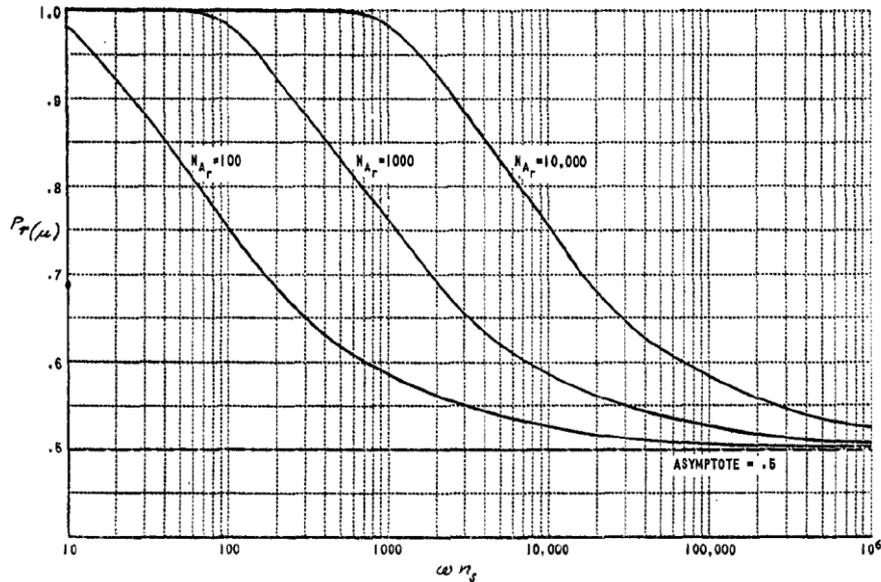


FIG. 8. $P_r(\mu)$ as function of ωn_s . (For $P_a = .07$, $\omega_e = 0$. Ideal environment assumed.)

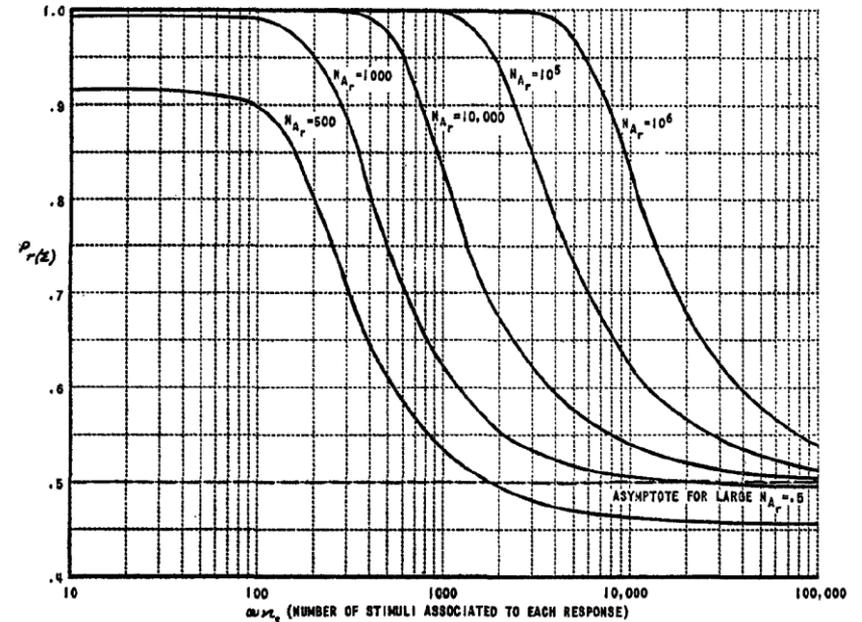


FIG. 7. $P_r(\Sigma)$ as function of ωn_s , for discrete subsets. ($\omega_e = 0$, $P_a = .005$. Ideal environment assumed.)

As the number of stimuli increases, the probability falls to 0.5 (random selection).

N_{A_R} = total number of A-units in source set of a response.

Comparison of α , β , γ systems.

- γ model performs better than α , β models.
- The μ (mean discriminating phase) performs better than the Σ (sum discriminating phase) system.

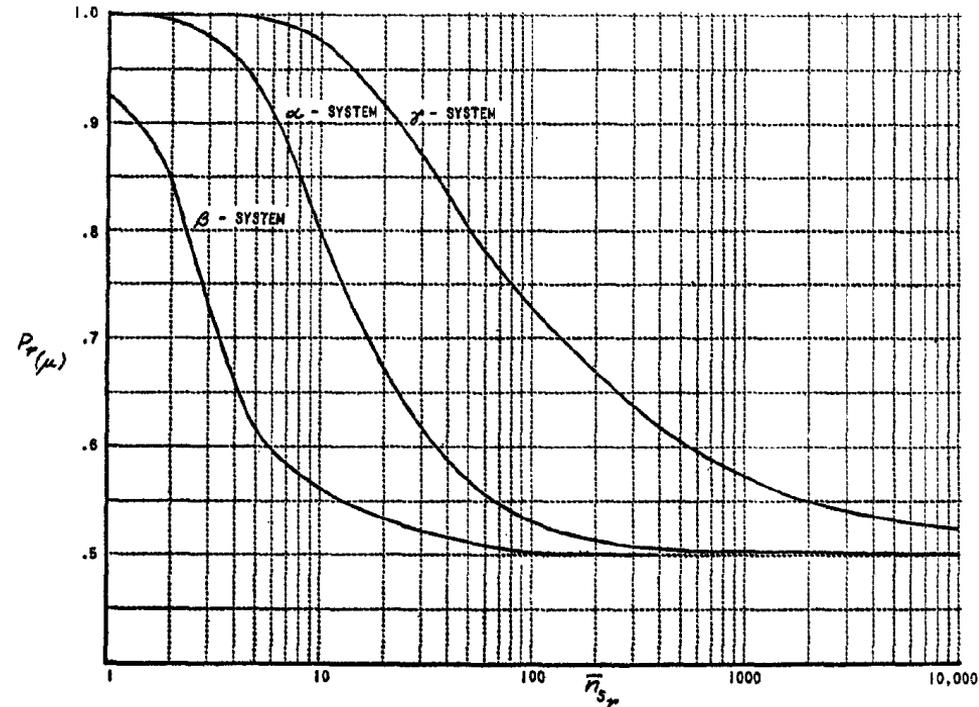


FIG. 10. Comparison of α , β , and γ systems, for variable n_s .
($N_R = 100$, $\sigma_{n_s} = .5\bar{n}_s$, $N_A = 10,000$, $P_b = .07$, $\omega = .2$).

Experiment 2

- After infinite learning, it makes no difference whether the perceptron has seen a particular test stimulus before or not; if the stimuli are drawn from a differentiated environment, their performance will be equally good in either cases.
- $P_{c\alpha\beta}$ = expected value of P_c between pairs of stimuli drawn at random from classes α and β .
- If $P_{c11} > P_a > P_{c12}$, performance of the perceptron will be better than chance.

Results of Experiment 2

- P_r and P_g approach the same asymptote.
- As number of association cells increase, the learning limit approaches unity rapidly.
- N_e = number of effective A-units in the source set of a response.
- n_{Sr} = number of stimuli associated to each response.

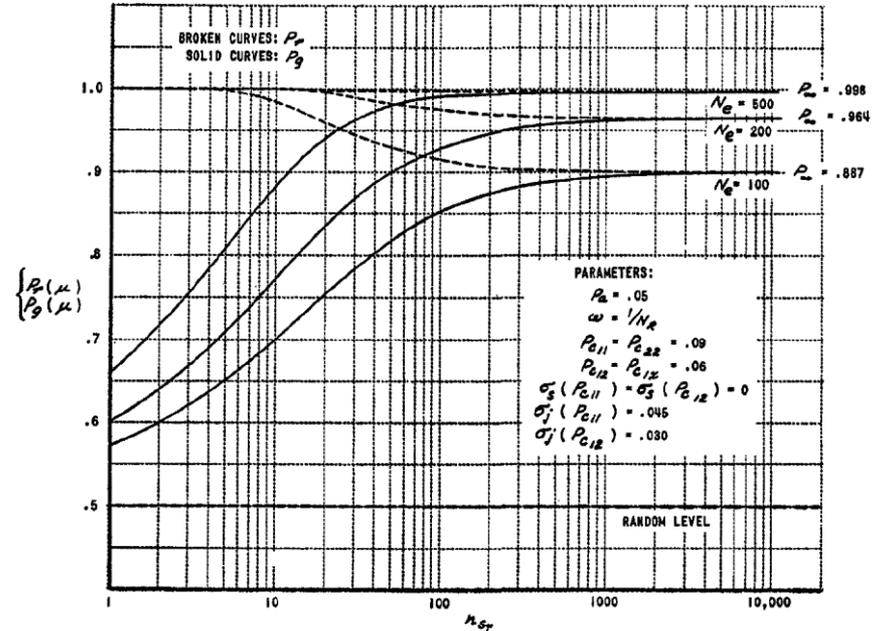


FIG. 11. P_r and P_g as function of n_{Sr} . Parameters based on square-circle discrimination.

Bivalent Systems

- Till now active A-units always increased in value.
- In a bivalent system, positive and negative reinforcements can be applied causing the active A-units values to either increase or decrease in value.
- This causes a perceptron to learn by trial and error.

Conclusion

For a given mode of organization ($\alpha, \beta, \gamma, \Sigma$ or μ) the fundamental phenomenon of learning, perceptual discrimination and generalization can be predicted entirely from six basic **physical parameters**, namely:

x : the number of excitatory connections per A-unit,

y : the number of inhibitory connections per A-unit,

θ : the expected threshold of an A-unit,

ω : the proportion of R-units to which an A-unit is connected,

N_A : The number of A-units in the system,

N_R : the number of R-units in the system

Capabilities of the perceptron.

- It Can do
 - Pattern recognition, associative learning, selective attention and selective recall.
- It is Capable of
 - Temporal pattern recognition, spatial recognition, multi modal input, trial and error learning.
- It Cannot do
 - Relative judgement, abstraction of relationships.

The proposed model is

- Parsimonious
 - All variable used are already present in structure of physical and biological sciences.
 - Introduced only 1 new variable, i.e. the value of A-unit which is assumed to have potentially measurable physical correlate.
- Verifiable
 - More confident of the validity compared to the previous models.
- Explanatory power and generality.
 - Not specific to any one organism or learning situation.
 - A theory of learning on these foundations would permit the synthesis of behaving systems.

The Perceptron: A Model for Brain Functioning

H.D Block, 1962

Analysis of a simple perceptron

- N_s = number of sensory units
- N_a = number of associator units.
- s_σ = sensory unit, a_μ = associator unit
- S_i = stimulus
- $C_{\sigma\mu} \in \{0,1\}$ = connection between s_σ and a_μ .

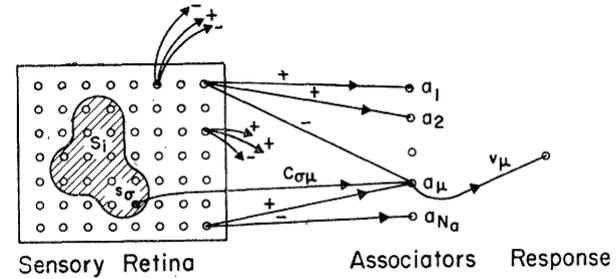


FIG. 4. Simple perceptron.

Signal transmitted to associator a_μ when S_i is applied is $\alpha_\mu^i = \sum_{\sigma, s_\sigma \in S_i} C_{\sigma\mu}$

Total signal arriving at response unit is $u = \sum_{\mu} v_\mu$, where summation is over active a-units.

In this simple model, $C_{\sigma\mu}$ do not change, therefore $A(S_i)$, set of a-units don't change.

$$e_{\mu i} = \begin{cases} 1 & \text{if } a_\mu \in A(S_i) \\ 0 & \text{if } a_\mu \notin A(S_i) \end{cases}, \text{ now } u \text{ becomes } u_i = \sum_{\mu} v_\mu e_{\mu i}$$

Discrimination: Learning by Error Correction

- 2 classes, +1 and -1.
- Each $S_i \in \rho_i$ where $\rho_i \in \{\pm 1\}$.
- Perceptron predicts correctly if $\rho_i u_i = \sum_{\mu} v_{\mu} e_{\mu i} \rho_i > \theta, \theta \in R_+$
- Error correction Procedure: A stimulus is shown, if the response is correct then no reinforcement is made, else, the v_{μ} for active associator a_{μ} is incremented by $\eta \rho_i$.
- The author assumes that classes are separable.

Discrimination: Learning by Error Correction

- After a certain finite number of steps the machine will give correct response to all stimuli thereafter, a bound on the number of mistakes.

$$\sum_i x_i \leq (n/m\eta) \max_i [2(\Theta - \rho_i u_i^0) + \eta N_i]. \quad (9)$$

where x_i is the number of times machine has incorrectly identified stimulus S_i .

Forced learning

- The value of each A-unit is incremented by $\eta\rho_i$ each time S_i is shown, regardless of machine's response.
- **Result:** If a perceptron can discriminate under forced learning, it will learn under error correction. If it can learn under error correction it can learn it under forced some choices of frequency vector p (frequency of occurrences of various stimuli), but not for others. Hence corrective learning is more effective.

Results

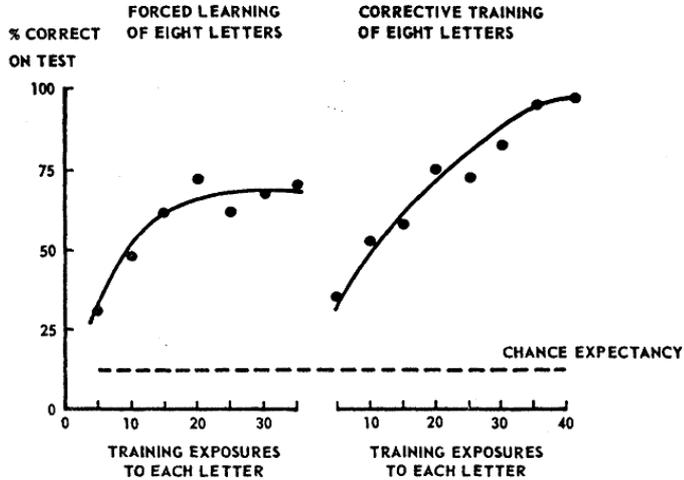


FIG. 9. Learning curves for eight letter identification task (each letter upright, but in a variety of locations).

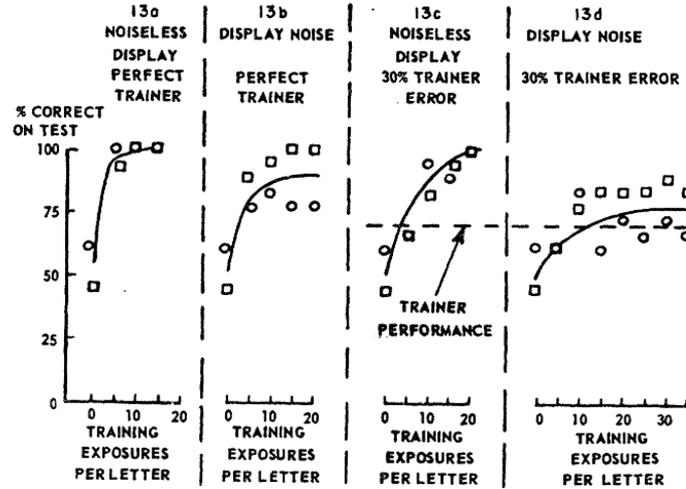


FIG. 12. Effects of noisy display and imperfect trainer on learning of "E"- "X" discrimination.

- Corrective training does better as compared to forced learning.
- Despite trainer error, accuracy numbers approach best performance.

Generalization.

- Simple perceptron generalizes well.
- Reason: more overlap with stimuli of same class and very little overlap between stimuli of opposite class.
- This causes perceptron to give correct response on basis of having seen the like stimuli.

References.

- Rosenblatt, Frank. "The perceptron: a probabilistic model for information storage and organization in the brain." *Psychological review* 65.6 (1958): 386.
- Block, H. D. "The perceptron: A model for brain functioning. I." *Reviews of Modern Physics* 34.1 (1962): 123.