Psychological Review
Vol. 65, No. 6, 1958

THE PERCEPTRON: A PROBABILISTIC MODEL FOR INFORMATION STORAGE AND ORGANIZATION IN THE BRAIN [1]

F. ROSENBLATT

Cornell Aeronautical Laboratory

Presenter: Daniel Maturana

Frank Rosenblatt, PhD (1928 – 1971)

- Background: Psychology
- Undergrad, PhD and teacher at Cornell

Introduction: 3 questions about biological systems

1. How is information about the physical world sensed by the biological system?

Introduction: 3 questions about biological systems

1. How is information about the physical world sensed by the biological system?
2. In what form is information stored?

Introduction: 3 questions about biological systems

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

Introduction: 3 questions about biological systems

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

Two answers for Question 2 (memory)

- Coded (symbolic?) representations with 1-1 mapping to world

Two answers for Question 2 (memory)

- Coded (symbolic?) representations with 1-1 mapping to world
- “Associationist”, “connectionist”: no simple mapping from memory to world

Two answers for Question 2 (memory)

- Coded (symbolic?) representations with 1-1 mapping to world
- “Associationist”, “connectionist”: no simple mapping from memory to world

Two answers for question 3 (behavior)

- “if-then” pattern matching

Two answers for question 3 (behavior)

- “if-then” pattern matching
- Same answer as question 2 - “automatic activation” of correct response.

Two answers for question 3 (behavior)

- “if-then” pattern matching
- Same answer as question 2 - “automatic activation” of correct response.

The *perceptron* framework

- **The Perceptron:** A hypothetical nervous system or machine designed to illustrate the fundamental properties of intelligent systems in general, without becoming too deeply enmeshed in the details of biological systems.

The *perceptron* framework

- **The Perceptron:** A hypothetical nervous system or machine designed to illustrate the fundamental properties of intelligent systems in general, without becoming too deeply enmeshed in the details of biological systems.
- The textbook linear perceptron is a special case!

The Perceptron is “probabilistic”

- Rosenblatt rejects symbolic logic and Boolean algebra as a suitable language for analysis of this system

The Perceptron is “probabilistic”

- Rosenblatt rejects symbolic logic and Boolean algebra as a suitable language for analysis of this system
- Advocates use a probabilistic analysis instead, arguing this is more realistic for biological intelligence

The Perceptron is “probabilistic”

- Rosenblatt rejects symbolic logic and Boolean algebra as a suitable language for analysis of this system
- Advocates use a probabilistic analysis instead, arguing this is more realistic for biological intelligence
- Part of a deeper split between “discrete” A.I. and more “statistical” approaches.

Biological inspiration

- The perceptron is biologically inspired (Hebb, Hull, Eccles, etc.)

Biological inspiration

- The perceptron is biologically inspired (Hebb, Hull, Eccles, etc.)
- However, argues other biological theories are not rigorous enough

- The perceptron is biologically inspired (Hebb, Hull, Eccles, etc.)
- However, argues other biological theories are not rigorous enough
- The perceptron is a concrete, testable model

1. Construction of physical connections in the nervous system is largely random.

Five biological assumptions

1. Construction of physical connections in the nervous system is largely random.
2. The system of connections is capable of plasticity through long-lasting changes in the neurons.

Five biological assumptions

1. Construction of physical connections in the nervous system is largely random.
2. The system of connections is capable of plasticity through long-lasting changes in the neurons.
3. After exposure to large sample of stimuli, “similar” stimuli will activate same sets of cells.

Five biological assumptions

1. Construction of physical connections in the nervous system is largely random.
2. The system of connections is capable of plasticity through long-lasting changes in the neurons.
3. After exposure to large sample of stimuli, “similar” stimuli will activate same sets of cells.
4. Positive and negative reinforcement influences connection formation.

Five biological assumptions

1. Construction of physical connections in the nervous system is largely random.
2. The system of connections is capable of plasticity through long-lasting changes in the neurons.
3. After exposure to large sample of stimuli, “similar” stimuli will activate same sets of cells.
4. Positive and negative reinforcement influences connection formation.
5. *Similarity* is not a necessary attribute of the stimuli, but depends on the perceiving system, which evolves though interaction with a given environment.

The organization of a perceptron

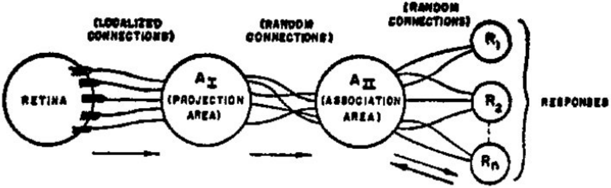


FIG. 1. Organization of a perceptron.

- S-points: sensory cells
- A-units: association cells
- R-units: response cells

The organization of a perceptron

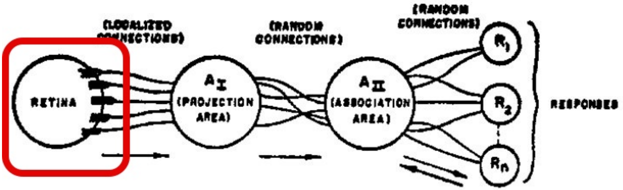


FIG. 1. Organization of a perceptron.

- S-points: sensory cells
- A-units: association cells
- R-units: response cells

The organization of a perceptron

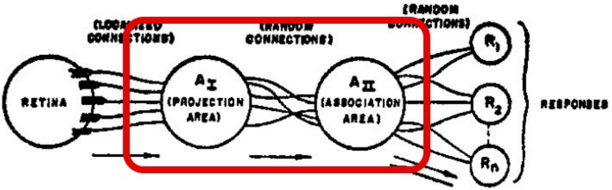


FIG. 1. Organization of a perceptron.

- S-points: sensory cells
- A-units: association cells
- R-units: response cells

The organization of a perceptron

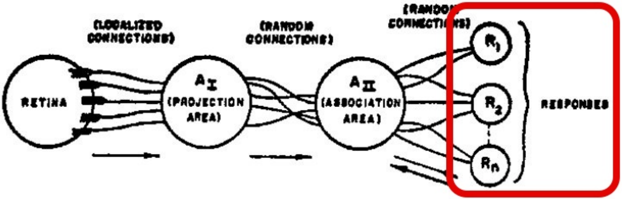


FIG. 1. Organization of a perceptron.

- S-points: sensory cells
- A-units: association cells
- R-units: response cells

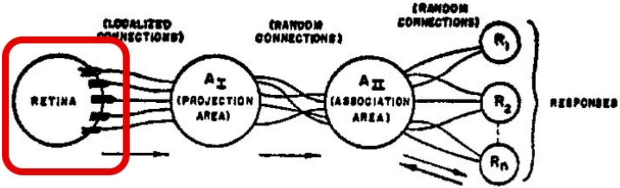


FIG. 1. Organization of a perceptron.

- “Sensory cells” activated by stimuli
- In practice modelled after the retina
- Activation is assumed to be discrete

A-units

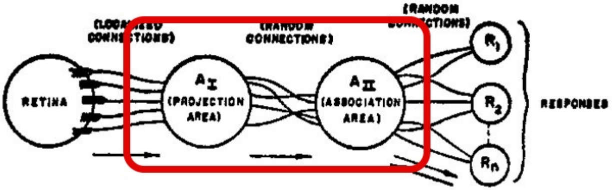


Fig. 1. Organization of a perceptron.

- “Association” units.
- Summation unit w/ threshold

$$\text{output} = \text{sgn}\left[\sum \text{excitatory} - \sum \text{inhibitory} - \theta\right]$$

A-units

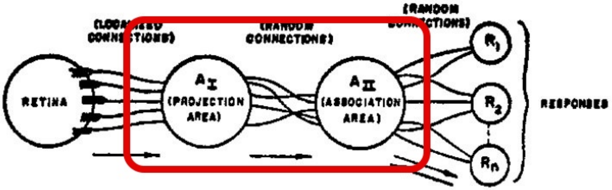


FIG. 1. Organization of a perceptron.

- A_1-units: receive connections from S-points. ("feature detectors")
- A_2-units: receive random input from A_1 units.

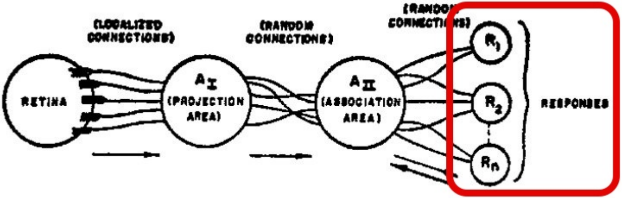


FIG. 1. Organization of a perceptron.

- "Response-cells", or output

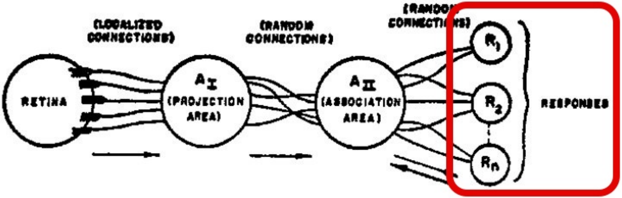


FIG. 1. Organization of a perceptron.

- "Response-cells", or output
- Reciprocal connections to A_2 cells.

R-units

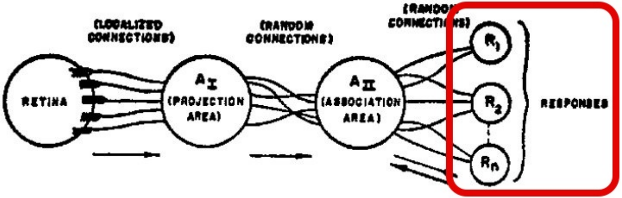


FIG. 1. Organization of a perceptron.

- "Response-cells", or output
- Reciprocal connections to A_2 cells.
- The feedback connections create a "winner-take-all" situtation

Learning with feedback

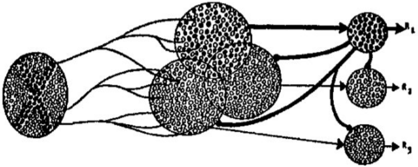


FIG. 3A. Predominant phase. Inhibitory connections are not shown. Solid black units are active.

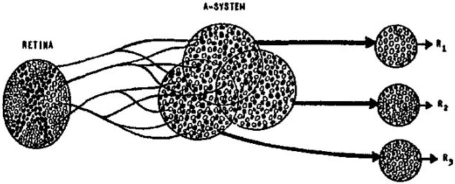


FIG. 3B. Postdominant phase. Dominant subset suppresses rival sets. Inhibitory connections shown only for R_1.

FIG. 3. Phases of response to a stimulus.

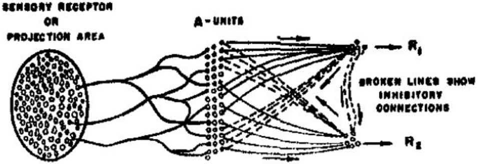


FIG. 2A. Schematic representation of connections in a simple perceptron.

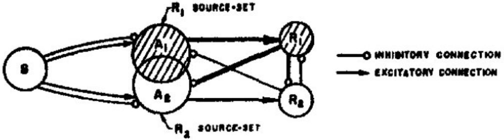


FIG. 2B. Venn diagram of the same perceptron (shading shows active sets for R_1 response).

The perceptron learning rule?

- The textbook equation is not in this paper!

The perceptron learning rule?

- The textbook equation is not in this paper!
- The paper considers and analyzes variations of the perceptron, none exactly the textbook perceptron

- "Learning" occurs via changes in the "activation" (denoted by V for Value) of A-cells.

Learning in the perceptron

- “Learning” occurs via changes in the “activation” (denoted by V for Value) of A-cells.
- Similar to what we might call weights

Learning in the perceptron

- “Learning” occurs via changes in the “activation” (denoted by V for Value) of A-cells.
- Similar to what we might call weights
- Biological meaning of “activation” is vague: it could be amplitude, frequency, latency, probability of transmission

Learning in the perceptron

- In a broad characterization, learning is done by reinforcement of active cells (cf. Hebb)

Learning in the perceptron

- In a broad characterization, learning is done by reinforcement of active cells (cf. Hebb)
- Three systems with different value dynamics

Learning in the perceptron

- In a broad characterization, learning is done by reinforcement of active cells (cf. Hebb)
- Three systems with different value dynamics
  - α-system: each active cell gains unit activation

Learning in the perceptron

- In a broad characterization, learning is done by reinforcement of active cells (cf. Hebb)
- Three systems with different value dynamics
  - α-system: each active cell gains unit activation
  - β-system: active cells gain a “mean” activation

Learning in the perceptron

- In a broad characterization, learning is done by reinforcement of active cells (cf. Hebb)
- Three systems with different value dynamics
    - α-system: each active cell gains unit activation
    - β-system: active cells gain a “mean” activation
    - γ-system: total activation is constant; active cells gain value, inactive cells lose value

- α-system is closest to the textbook perceptron, in “bivalent” variant (described below).

Some notes

- α-system is closest to the textbook perceptron, in "bivalent" variant (described below).
- Nothing has been said yet about supervised training

Some notes

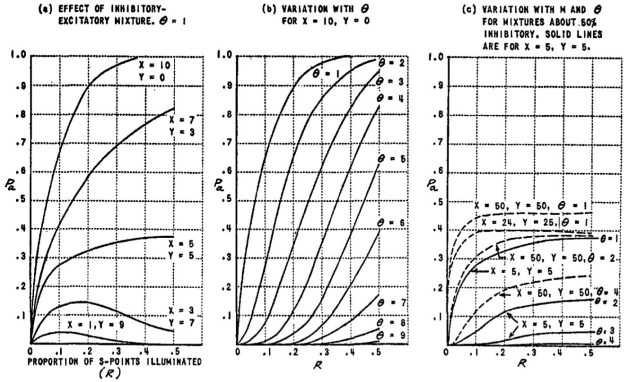
- α-system is closest to the textbook perceptron, in “bivalent” variant (described below).
- Nothing has been said yet about supervised training
- Dynamic, “self-organizing” aspects not explored in this paper

Analysis of “predominant” phase

- Before learning takes place
- Questions:
  - How does parameters of perceptron affect activation?
  - How does characteristics of stimuli affect activation?

Analysis of “predominant” phase

- Focuses on two quantities:
  - P_a: expected proportion of A-units activated.
  - P_c: conditional probability that A-unit that responds to stimulus S_1 responds to another stimulus S_2.



Depends on threshold θ, size of stimuli and type of connections

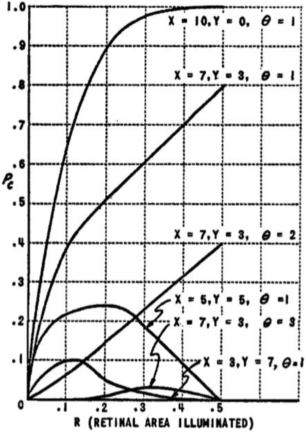


FIG. 5. P_c as a function of R, for nonoverlapping stimuli.

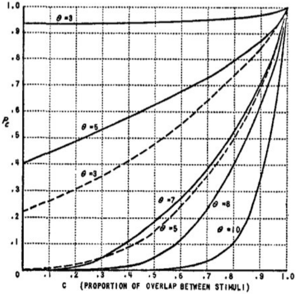


FIG. 6. P_c as a function of C. $X = 10$, $Y = 0$. Solid lines: R = .5; broken lines: R = .2.

Depends mostly on overlap between stimuli

Analysis of “postdominant” phase

- After training, each response unit has its own set of (possibly overlapping) input/weights

Analysis of “postdominant” phase

- After training, each response unit has its own set of (possibly overlapping) input/weights
- (Even for two responses)

Analysis of “postdominant” phase

- After training, each response unit has its own set of (possibly overlapping) input/weights
- (Even for two responses)
- Two possible decision criteria for response:

Analysis of “postdominant” phase

- After training, each response unit has its own set of (possibly overlapping) input/weights
- (Even for two responses)
- Two possible decision criteria for response:
    - μ-system: greatest mean value response is first

Analysis of “postdominant” phase

- After training, each response unit has its own set of (possibly overlapping) input/weights
- (Even for two responses)
- Two possible decision criteria for response:
    - μ-system: greatest mean value response is first
    - Σ-system: greatest sum response is first

Analysis of “postdominant” phase

- Two experiments are considered, “ideal” (memorize random stimuli) and “differentiated” (learn “classes”).

Analysis of “postdominant” phase

- Two experiments are considered, “ideal” (memorize random stimuli) and “differentiated” (learn “classes”).
- Analysis of expected probability of correct recall (for ideal case) and correct generalization (for differentiated case).

Analysis of “postdominant” phase

- Two experiments are considered, “ideal” (memorize random stimuli) and “differentiated” (learn “classes”).
- Analysis of expected probability of correct recall (for ideal case) and correct generalization (for differentiated case).
- Basic equation is

$$P = \left(1 - (1 - P_a)^{N_s}\right) \phi \left(\frac{c_1 n_{s_r} + c_2}{\sqrt{c_3 n_{s_r}^3 + c_4 n_{s_r}}} \right)$$

with constants depending on details of the experiment.

An “ideal” experiment

1. Generate random stimuli

An “ideal” experiment

1. Generate random stimuli
2. Give an arbitrary label (desired response) to stimuli

An “ideal” experiment

1. Generate random stimuli
2. Give an arbitrary label (desired response) to stimuli
3. “Force” desired responses, and perform “learning”

An “ideal” experiment

1. Generate random stimuli
2. Give an arbitrary label (desired response) to stimuli
3. “Force” desired responses, and perform “learning”
4. After learning, test if system can recall labels

Results of “ideal” experiment

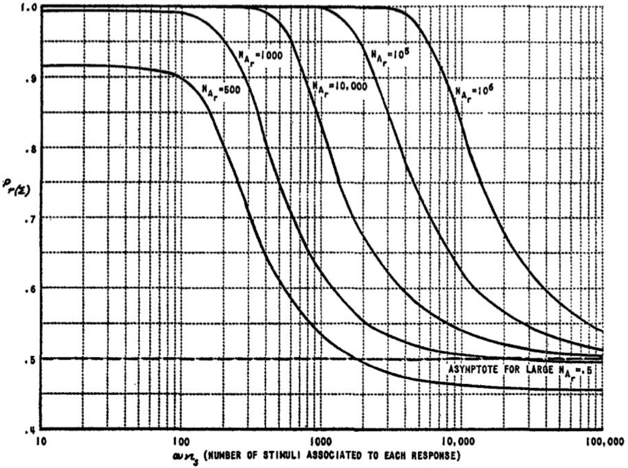


FIG. 7. $P_r(\Sigma)$ as function of ωn_s, for discrete subsets. ($\omega_c = 0$, $P_a = .005$. Ideal environment assumed.)

Performance gets worse as samples increase.

Results of “ideal” experiment

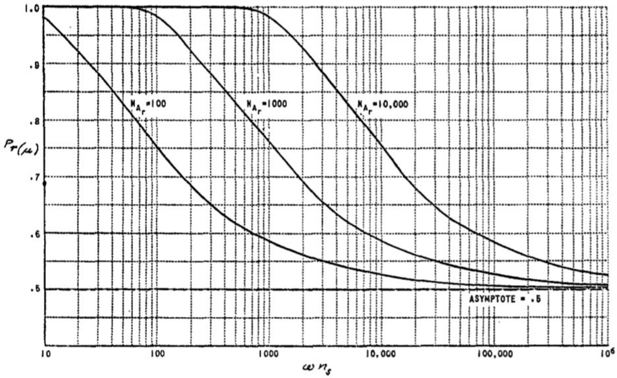


FIG. 8. $P_{r(\mu)}$ as function of ωn_s. (For $P_a = .07$, $\omega_c = 0$. Ideal environment assumed.)

Results of “ideal” experiment

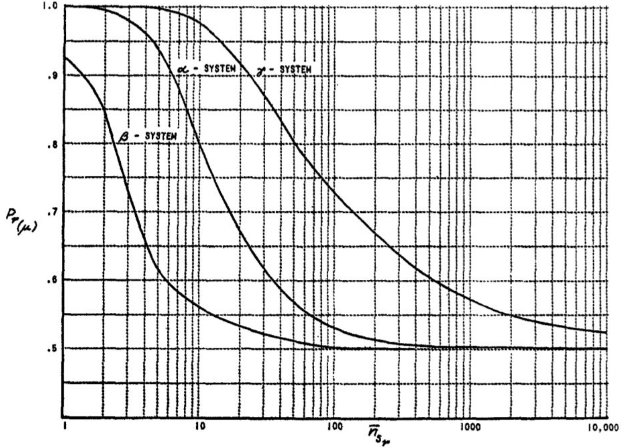


FIG. 10. Comparison of α, β, and γ systems, for variable n_{s_r} ($N_R = 100$, $\sigma_{n_{rs}} = .5\bar{n}_{s_r}$, $N_A = 10{,}000$, $P_a = .07$, $\omega = .2$).

Results of “ideal” experiment

- Empirically γ-system seems better
- Empirically μ rule is better, except for γ-systems

A “differentiated” experiment

1. Generate stimuli from a class (e.g. squares vs. circles)

A “differentiated” experiment

1. Generate stimuli from a class (e.g. squares vs. circles)
2. Give an arbitrary label (desired response) to stimuli

A “differentiated” experiment

1. Generate stimuli from a class (e.g. squares vs. circles)
2. Give an arbitrary label (desired response) to stimuli
3. “Force” desired responses, and perform “learning”

A “differentiated” experiment

1. Generate stimuli from a class (e.g. squares vs. circles)
2. Give an arbitrary label (desired response) to stimuli
3. “Force” desired responses, and perform “learning”
4. After learning, test if system can generalize labels to unseen stimuli

From equation of expected recall/generalization, as $n_s \to \infty$,

$$P = \left(1 - (1 - P_a)^{N_s}\right) \phi \left(\frac{c_1}{\sqrt{c_3}} \right)$$

> *"In the limit 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, the performance will be equally good in either case."*

- Let $P_{c\alpha\beta}$ be the expected value of P_c for pairs of stimuli drawn from classes α and β

- Let $P_{c\alpha\beta}$ be the expected value of P_c for pairs of stimuli drawn from classes α and β
- If $P_{c11} > P_a > P_{c12}$, performance of perceptron as samples tend to infinity will be better than chance.

Results of “differentiated” experiment

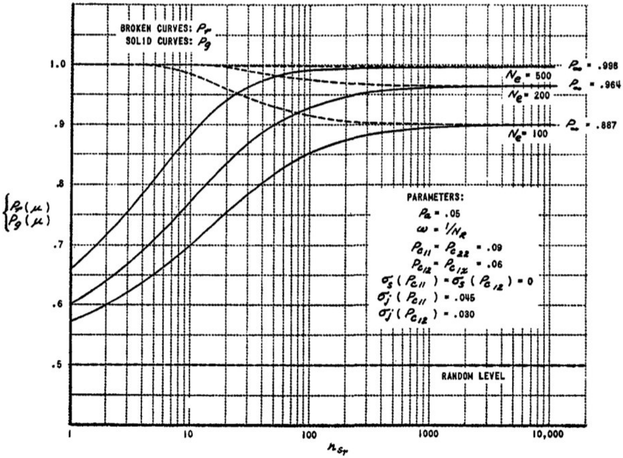


FIG. 11. P_r and P_g as function of n_{s_r}. Parameters based on square-circle discrimination.

Some notes

- No analysis of what kind of things can be learned
- No analysis of convergence
- Advocates use of binary coding (as opposed to one hot) for better performance with large number of responses

- Learning by “trial-and-error”

- Learning by "trial-and-error"
- Positive *and* negative reinforcement

Bivalent systems

- Learning by “trial-and-error”
- Positive *and* negative reinforcement
- i.e. Increase activation if right, decrease if wrong.

Bivalent systems

- Learning by “trial-and-error”
- Positive *and* negative reinforcement
- i.e. Increase activation if right, decrease if wrong.
- with α-system, most similar to the textbook perceptron!

Other improvements and extensions

- Consideration of time, to analyze velocities, sequences, etc.

Other improvements and extensions

- Consideration of time, to analyze velocities, sequences, etc.
- Non-random origins of A-units on retina ("feature detectors")

Other improvements and extensions

- Consideration of time, to analyze velocities, sequences, etc.
- Non-random origins of A-units on retina (“feature detectors”)
- Self-supervised concept formation

Other improvements and extensions

- Consideration of time, to analyze velocities, sequences, etc.
- Non-random origins of A-units on retina (“feature detectors”)
- Self-supervised concept formation
- Intermodal (audio/photo) association

Limitations?

- Can do pattern recognition, associative learning, selective attention/recall, trial-and-error learning

Limitations?

- Can do pattern recognition, associative learning, selective attention/recall, trial-and-error learning
- But bad at “relative judgment” and “abstraction of relationships”.

- Can do pattern recognition, associative learning, selective attention/recall, trial-and-error learning
- But bad at “relative judgment” and “abstraction of relationships”.
- Similar to “brain-damaged patients”.

Conclusions (from paper)

- Main contribution: a model that predicts phenomena of *learning, perceptual discrimination* and *generalization* from six basic parameters:

Conclusions (from paper)

- Main contribution: a model that predicts phenomena of *learning, perceptual discrimination* and *generalization* from six basic parameters:
    - Number of excitatory inputs
    - Number of inhibitory inputs
    - Threshold of A-unit
    - Proportion of R-units to which each A-unit connects
    - Number of A-units
    - Number of R-units

Conclusions (from paper)

- Main contribution: a model that predicts phenomena of *learning*, *perceptual discrimination* and *generalization* from six basic parameters:
    - Number of excitatory inputs
    - Number of inhibitory inputs
    - Threshold of A-unit
    - Proportion of R-units to which each A-unit connects
    - Number of A-units
    - Number of R-units
- The model is parsimonious, verifiable and of high explanatory power.

- Highly influential paper

Conclusions, in retrospect

- Highly influential paper
    - Bio-inspired and useful

Conclusions, in retrospect

- Highly influential paper
  - Bio-inspired and useful
  - Statistical (as opposed to symbolic) approach

Conclusions, in retrospect

- Highly influential paper
  - Bio-inspired and useful
  - Statistical (as opposed to symbolic) approach
  - Use of probabilistic analysis

Conclusions, in retrospect

- Highly influential paper
  - Bio-inspired and useful
  - Statistical (as opposed to symbolic) approach
  - Use of probabilistic analysis
  - Use of computer simulation (besides hardware)

Conclusions, in retrospect

- Highly influential paper
    - Bio-inspired and useful
    - Statistical (as opposed to symbolic) approach
    - Use of probabilistic analysis
    - Use of computer simulation (besides hardware)
- Drawbacks

Conclusions, in retrospect

- Highly influential paper
  - Bio-inspired and useful
  - Statistical (as opposed to symbolic) approach
  - Use of probabilistic analysis
  - Use of computer simulation (besides hardware)
- Drawbacks
  - Didn’t “deliver” on promises

Conclusions, in retrospect

- Highly influential paper
  - Bio-inspired and useful
  - Statistical (as opposed to symbolic) approach
  - Use of probabilistic analysis
  - Use of computer simulation (besides hardware)
- Drawbacks
  - Didn’t “deliver” on promises
  - Practical and theoretical performance limitations