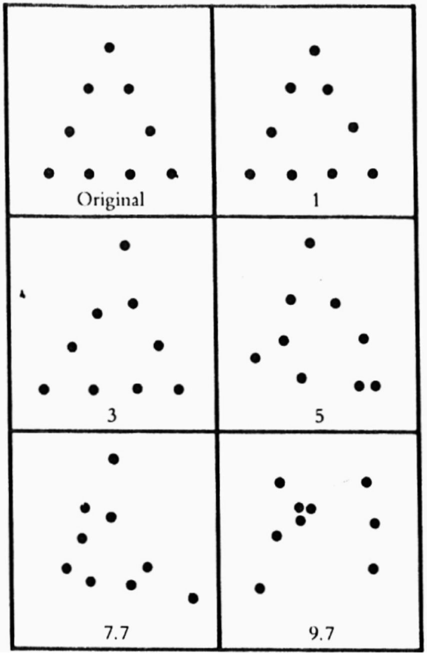
Learning from noisy exemplars

Posner & Keele (1968, J. Exp. Psy.)

- Exemplar dot patterns created by distorting "prototype" pattern to varying degrees
- Participants learn to categorize exemplars vs. random dot patterns
  - Given correct/incorrect feedback but never shown prototype
- Correct categorization of actual prototype as novel exemplar is faster than actual trained exemplars
  - Suggests participants induced and stored the prototype pattern
  - However, not true if specific trained exemplar was presented very often during training



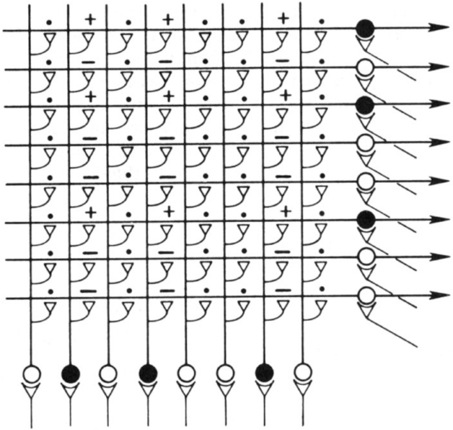
Representing both general and specific knowledge

Alternative classes of theories

- Store prototypes (e.g., Rosch)
  - Problem: Many demonstrations of effects of specific examples (e.g., congruity of test stimuli with particular trained stimuli)
- Store exemplars (e.g., Jacoby, Hintzman)
  - Problem: "Enumeration of specific experiences" require unlimited amount of storage and unrealistically powerful search mechanism
- Store both (e.g., Anderson)
  - Specific instances are stored (as productions) and then generalized

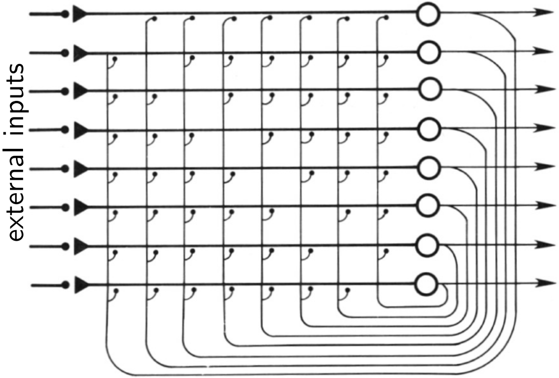
Auto-associator ("associates" each pattern with itself)

Standard pattern associator



- Feedforward processing from inputs to outputs

Auto-associator



- Recurrent network (settles over time)
- Same units serve as input and output
- Delta rule uses external input as targets

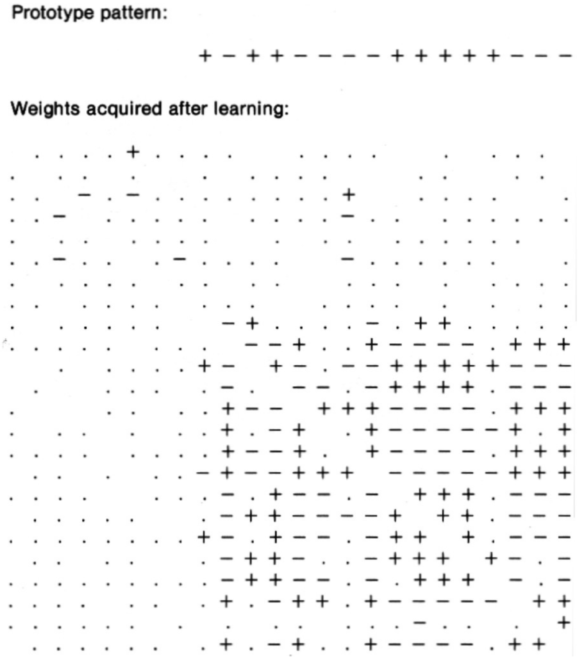
Distributed memory model (McClelland & Rumelhart, 1985)

- Delta-rule learning in an auto-associator
- Trained on distorted versions of prototype patterns
- Decay in weight increments (exponentially decreasing with time; for modeling priming)
- Properties
  1. Can extract prototype (central tendency) of set of patterns that are random distortions of prototype (e.g., semantic category)
  2. Can extract several different prototypes without labels for which prototype (category) each pattern "belongs to"
  3. Representations of specific, repeated exemplars can co-exist in the same set of connections as general knowledge of the prototype
  4. Accounts for various empirical priming phenomena

Prototype extraction

- "dog" as prototype
- 24 units: 16 general + 8 specific (name)
- 50 training patterns with p=0.2 distortion of *general* information
- weight changes decay to 5%
- weights capture correlational structure of prototype

- $+$ is positive weight
  $-$ is negative weight
  $\cdot$ is small weight
- First 8 rows/columns are "name" (random, no correlational structure)
- Last 16 rows/columns are general information (instances distored from prototype)
  - Weights reflect correlations among features



Multiple prototypes

- “dog” vs. “cat” vs. “bagel”
- correlation(dog,cat) $= 0.5$
- bagel *orthogonal* to dog and cat
- 16 general units $+$ 8 specific (name) units
- 50 training patterns for each prototype, all units distorted with $p=0.1$

Multiple prototypes

RESULTS OF TESTS AFTER LEARNING THE DOG, CAT, AND BAGEL PATTERNS

	Name Pattern	Visual Pattern
Pattern for dog prototype	+ − + − + − + −	+ − + + − − − − + + + + + − − −
Response to dog name		+3 −4 +4 +4 −4 −4 −4 −4 +4 +4 +4 +3 +4 −4 −4 −3
Response to dog visual pattern	+5 −4 +4 −5 +5 −4 +4 −4	
Pattern for cat prototype	+ + − − + + − −	+ − + + − − − − + − + − + + − +
Response to cat name		+4 −3 +4 +4 −4 −3 −3 −4 +4 −4 +4 −4 +4 +4 −4 +4
Response to cat visual pattern	+5 +4 −4 −5 +4 +4 −4 −4	
Pattern for bagel prototype	+ − − + + − − +	+ + − + − + + − + − − + + + + −
Response to bagel name		+3 +4 −4 +4 −4 +4 +4 −4 +4 −4 −4 +4 +3 +4 +4 −4

Note: $+3 = 0.3$; $-4 = -0.4$

Multiple prototypes (no labels)

RESULTS OF TESTS AFTER LEARNING
THE DOG, CAT, AND BAGEL PATTERNS WITHOUT NAMES

Dog visual pattern:	+	−	+	+	−	−	−	−	+	+	+	+	+	−	−	−
Probe:									+	+	+	+				
Response:	+3	−3	+3	+3	−3	−4	−3	−3	+6	+5	+6	+5	+3	−2	−3	−2
Cat visual pattern:	+	−	+	+	−	−	−	−	+	−	+	−	+	+	−	+
Probe:									+	−	+	−				
Response:	+3	−3	+3	+3	−3	−3	−3	−3	+6	−5	+6	−5	+3	+2	−3	+2
Bagel visual pattern:	+	+	−	+	−	+	+	−	+	−	−	+	+	+	+	−
Probe:									+	−	−	+				
Response:	+2	+3	−4	+3	−3	+3	+3	−3	+6	−6	−6	+6	+3	+3	+3	−3

Co-existence of prototype and exemplars

- "dog" prototype with Fido and Rover as specific dogs
- 3 names: "dog", "Fido", "Rover"
- other dogs: distortion p=0.2 of dog prototype
- 50 training trials of each
  - 50 for Fido and Rover each
  - 50 for distortions of dog

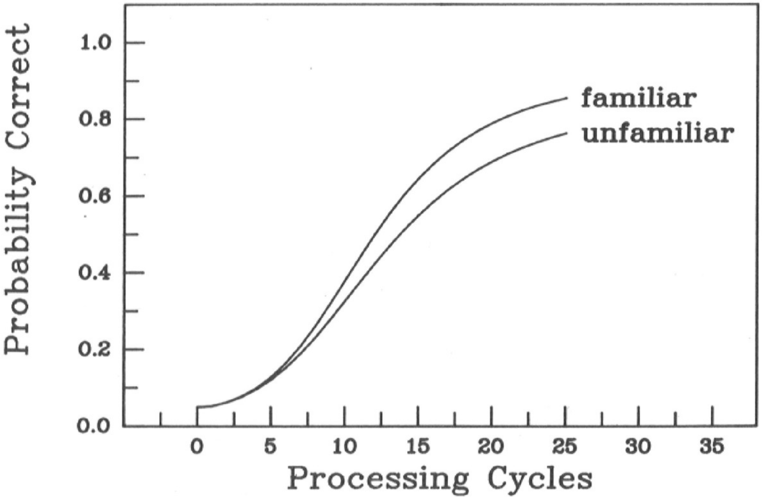
Co-existence of prototype and exemplars

RESULTS OF TESTS WITH PROTOTYPE AND SPECIFIC EXEMPLAR PATTERNS

	Name Pattern	Visual Pattern
Pattern for dog prototype	+ − + − + − + −	+ − + + − − − − + + + + + − − −
Response to prototype name		+4 −5 +3 +3 −4 −3 −3 −3 +3 +3 +4 +3 +4 −3 −4 −4
Response to prototype visual pattern	+5 −4 +4 −4 +5 −4 +4 −4	
Pattern for Fido exemplar	+ − − − + − − −	+ − (−) + − − − − + + + + + (+) − −
Response to Fido name		+4 −4 −4 +4 −4 −4 −4 −4 +4 +4 +4 +4 +4 +4 −4 −4
Response to Fido visual pattern	+5 −5 −3 −5 +4 −5 −3 −5	
Pattern for Rover exemplar	+ − − + + + − +	+ (+) + + − − − − + + + + + − − −
Response to Rover name		+4 +5 +4 +4 −4 −4 −4 −4 +4 +4 +4 +4 +4 −4 −4 −4

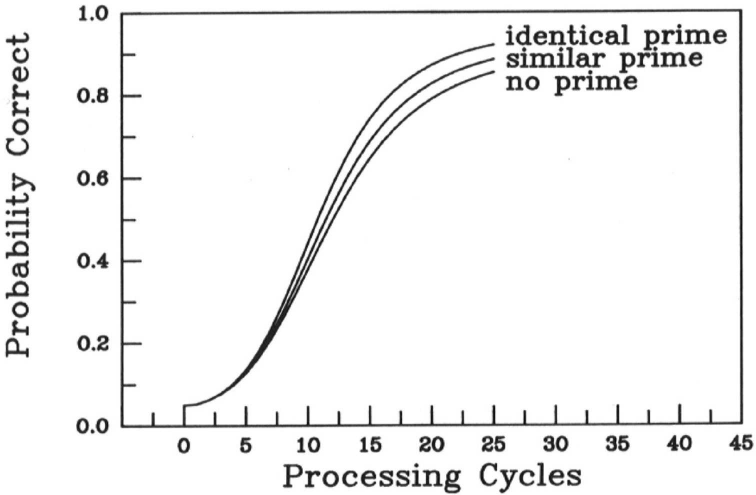
Effect of familiarity

- 10 training cycles on distortions (p=0.1) of 8 prototypes
- new distortions of familiar prototype produce stronger response than unfamiliar pattern



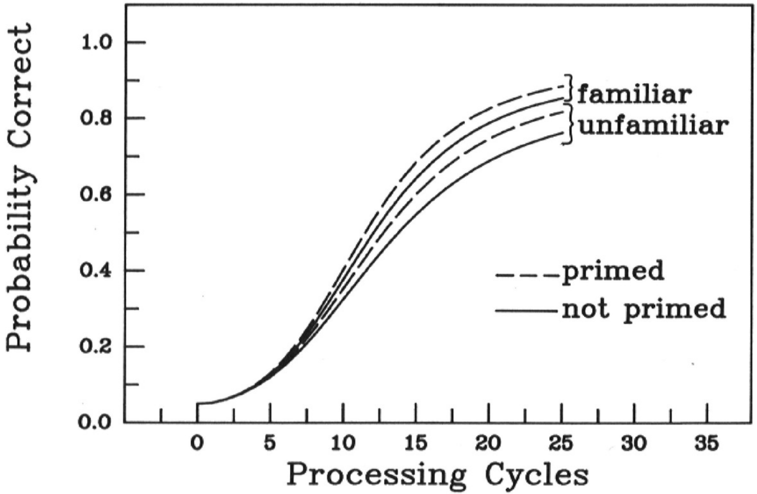
Priming: Effect of similarity

- primes: execute weight changes for pattern with no decay
- identical > similar > unrelated



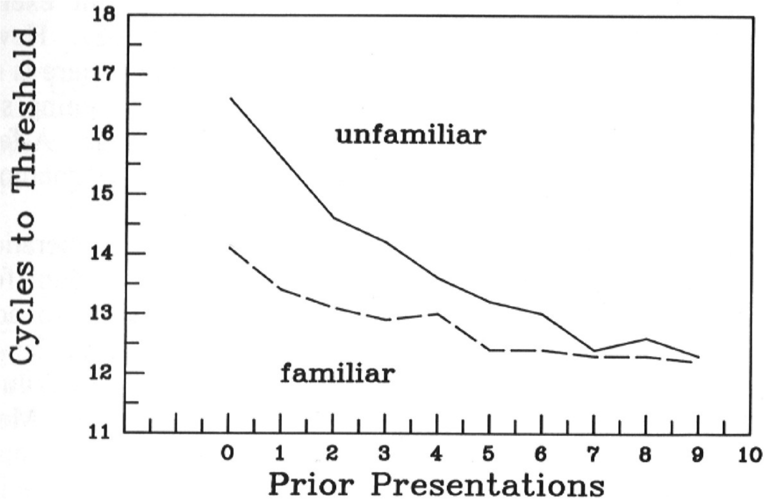
Priming: Effect of novelty on repetition priming

- Bigger priming effect for unfamiliar (novel) vs. familiar patterns



Priming: Effect of training

- Effect of repetition on settling time is larger for unfamiliar vs. familiar items



Trade-off of specific vs. general representation

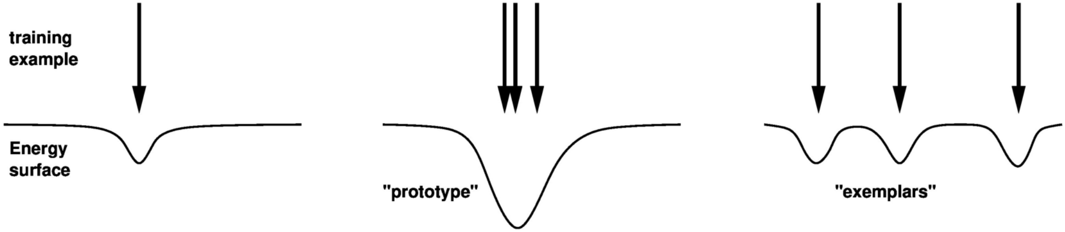
- Formation of prototype depends on collective similarity of exemplars
- Similarity (proximity) to a specific trained pattern is a strong determiner of perceptual performance
- But effects of specific exemplars break down when they are closer to the prototype
  - Then response to prototype is stronger than to any trained exemplar (even though it is "unfamiliar")
- **Intuition**: separate vs. converging gaussians

Relationship between training and energy/goodness

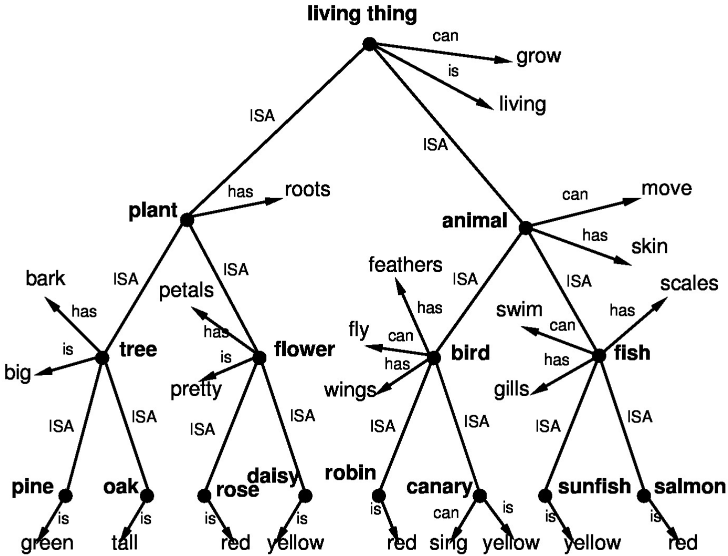
- Weight changes lower energy (raises goodness) of trained pattern <u>and those similar to it</u>

$$\text{Hebb:} \quad \triangle w_{ij} = \epsilon \; a_i \, a_j \qquad\qquad \text{Energy} = -\sum_{i<j} a_i \, a_j \, w_{ij}$$

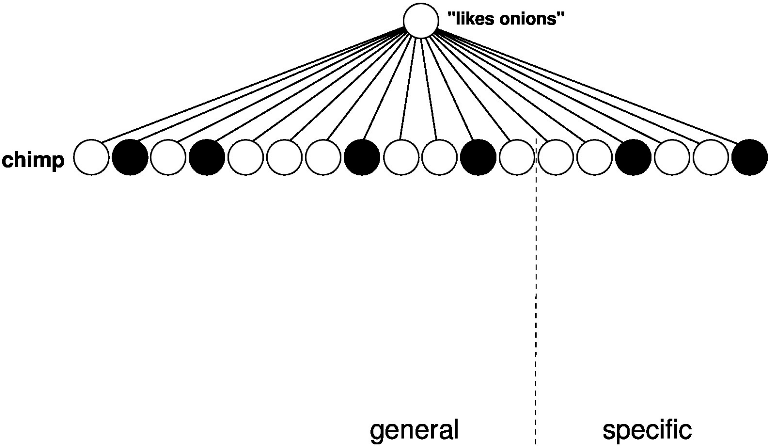
- Effects accumulate/combine if they are overlapping (over the same units) [**prototype-like**]
- Effects remain independent if they are non-overlapping [**exemplar-like**]



Similarity and generalization

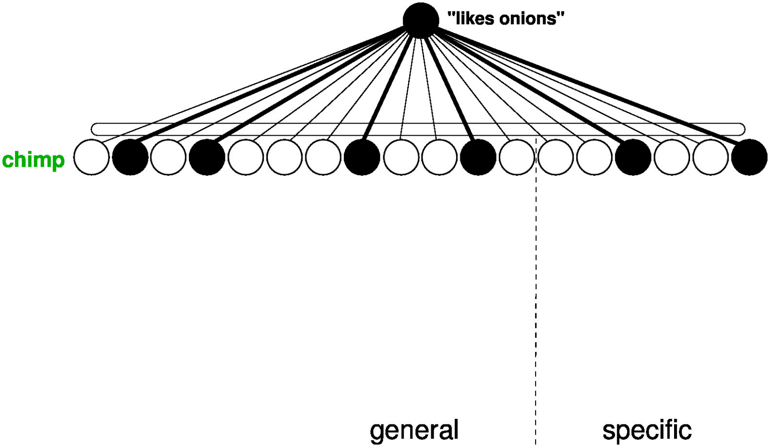


Do chimps like onions?



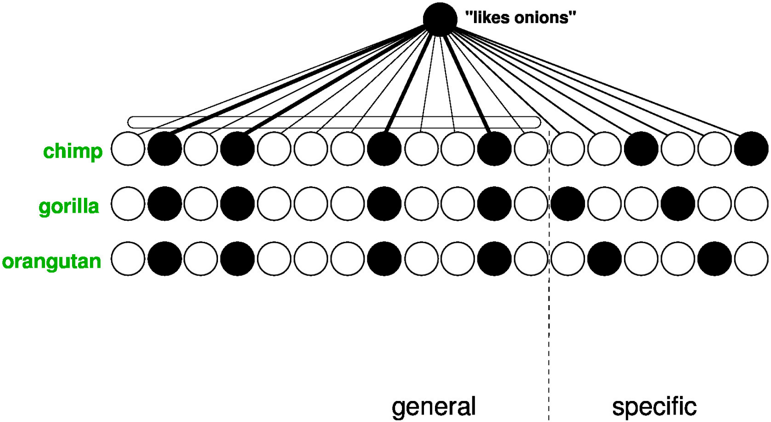
- Input is distributed representation of entity (chimp)
- Output is observed features (just considering "likes onions" here)
- Weights are updated with Hebbian learning

Chimps like onions



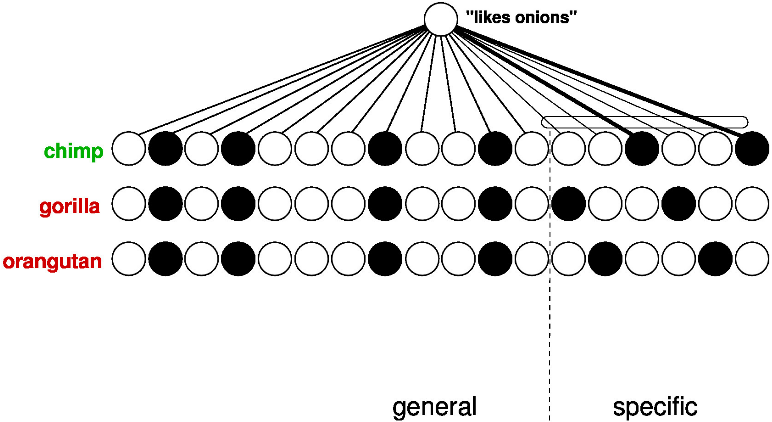
- Weights build up from all active input units due to correlation with output

All primates like onions



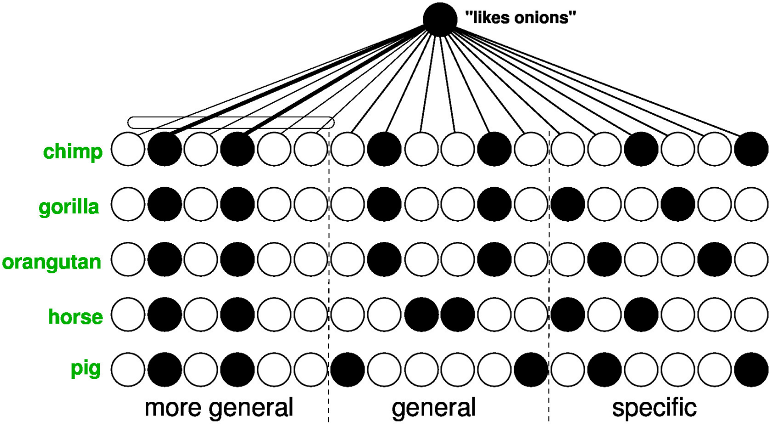
- "General" weights build up due to correlations
- "Specific" weights don't—no correlation because inputs vary when output is active

Other primates don't like onions



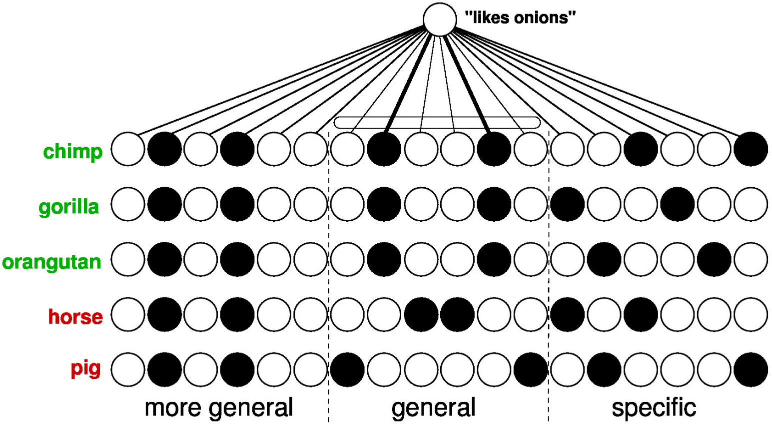
- "Specific" weights build up due to correlations
- No correlation for "general" weights because input is stable but output varies

All animals like onions



- Only the "more general" units are correlated with the output

Other animals don't like onions (but primates do)



- No correlation for "specific" weights because inputs vary
- No correlation for "more general" weights because output varies
- Only intermediate "general" weights build up due to correlations

Learning new concepts

- Localist: must **create** new unit and its connections or **find** such a pre-existing unit
  - Learning of concept is **discrete** event
- Distributed: make new pattern stable by weight adjustment
  - Concept emerges **gradually** over time
- Descriptors/"microfeatures" (units) constitute **language** for describing concepts
  - Called *microfeatures* to emphasize that they need not correspond to verbalizable "features"
  - 2^n potential concepts for n units (subject to similarity constraints)
- How is an appropriate pattern chosen for a new concept?
  - Pattern requiring minimal weight changes to become stable and have the required effects
  - Should incorporate general/specific relationships as just described
  - In general, **learned** in the context of particular tasks/behavior
    - Requires training internal (**"hidden"**) units that mediate between inputs and outputs

Performance of Two Patients With Impaired Knowledge of Living Things on Various Semantic Memory Tasks

Case	Living thing	Nonliving thing
	Picture identification	
JBR	6%	90%
SBY	0%	75%
	Spoken word definition	
JBR	8%	79%
SBY	0%	52%
	Examples of definitions	
JBR	Parrot: don't know	Tent: temporary outhouse, living home
	Daffodil: plant	Briefcase: small case used by students to carry papers
	Snail: an insect animal	Compass: tools for telling direction you are going
	Eel: not well	Torch: hand-held light
	Ostrich: unusual	Dustbin: bin for putting rubbish in
SBY	Duck: an animal	Wheelbarrow: object used by people to take material about
	Wasp: bird that flies	Towel: material used to dry people
	Crocus: rubbish material	Pram: used to carry people, with wheels and a thing to sit on
	Holly: what you drink	Submarine: ship that goes underneath the sea
	Spider: a person looking for things, he was a spider for his nation or country	Umbrella: object used to protect you from water that comes

Table 2
Performance of Two Patients With Impaired Knowledge of Nonliving Things on Various Semantic Memory Tasks

Case	Category		
	Animal	Flower	Object
	Spoken word–picture matching		
VER	86%	96%	63%
YOT	86%	86%	67%
	Picture–picture matching		
YOT	100%	—	69%

Farah & McClelland (1991) model

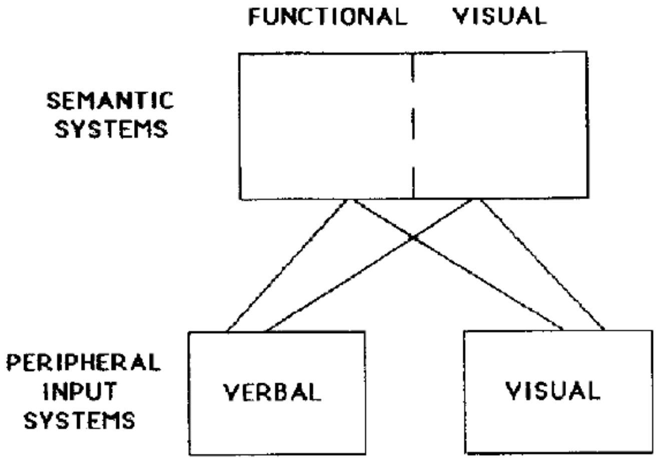
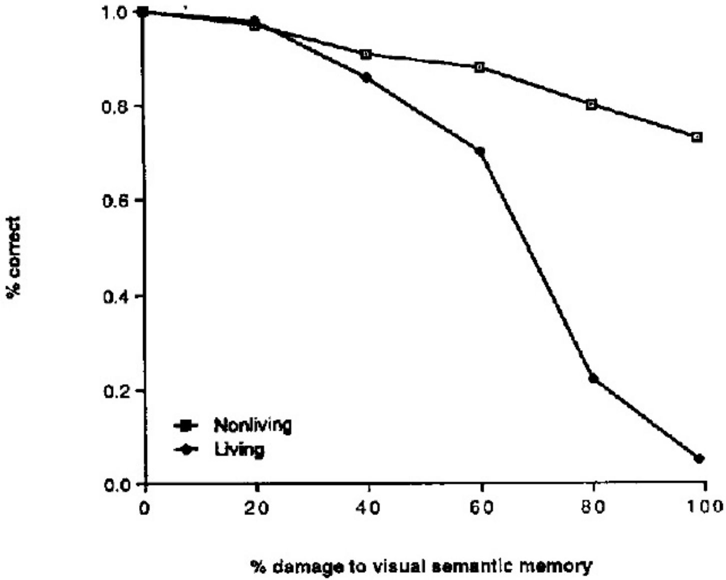


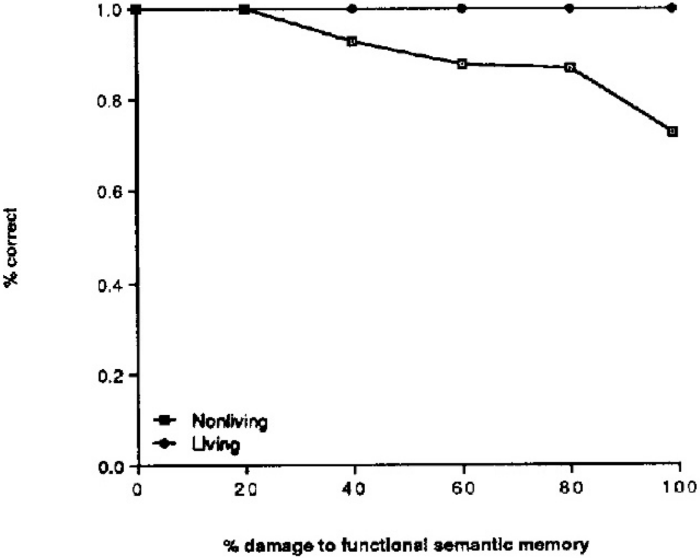
Figure 1. Schematic diagram of the parallel distributed processing model of semantic memory.

- Trained on name-picture associations (bidirectionally)
- Nonliving things have equal numbers of visual and functional semantic features
- Living things have 3 times more visual than functional semantic features

Visual semantic damage



Functional semantic damage



Functional impairment from visual semantic damage