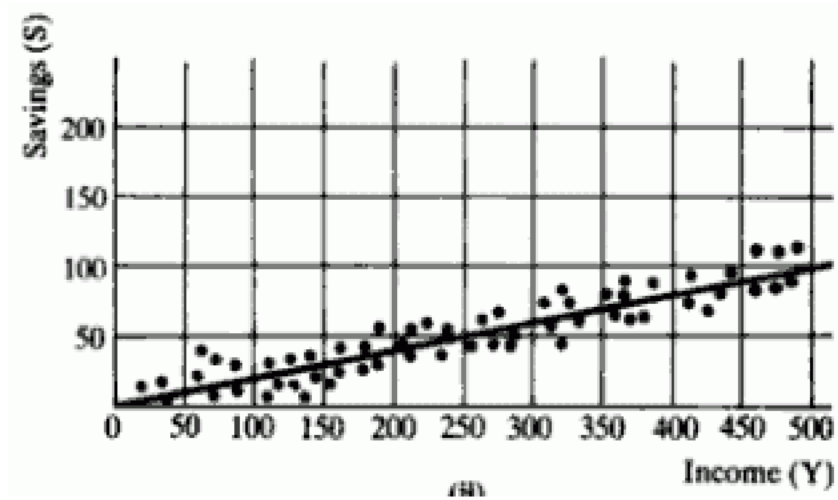
Computational Models of Human Cognition

Models

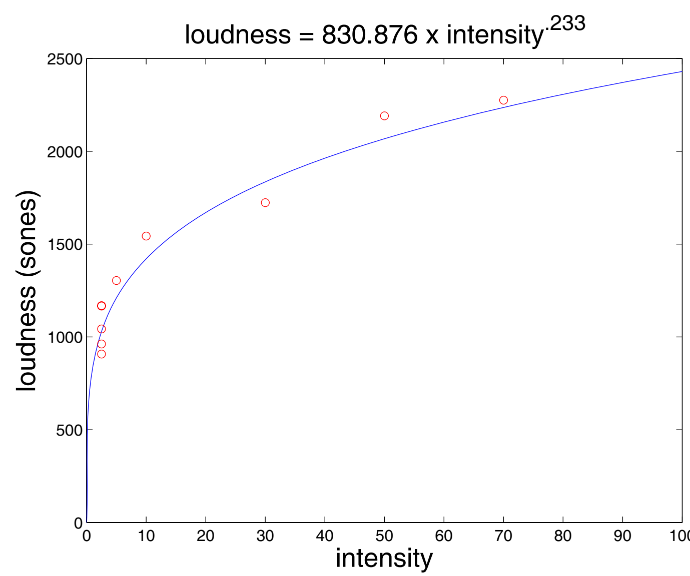
A model is a means of representing the structure or workings of a system or object.

e.g., model car

e.g., economic model

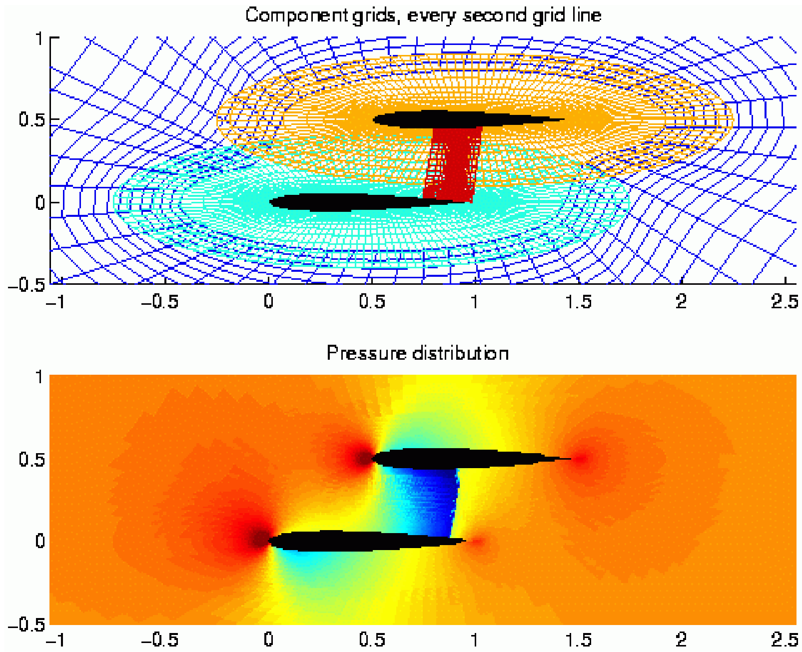


e.g., psychophysics model

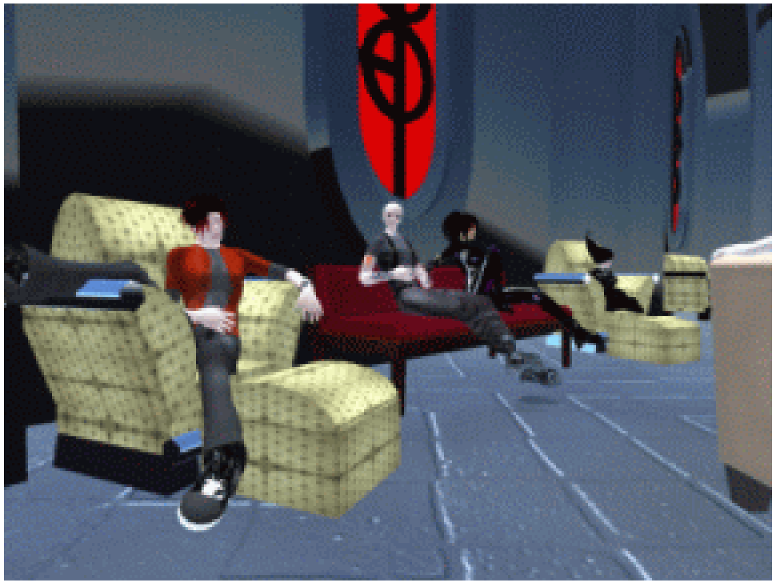


Computational Models

Models expressed as computer programs (sequence of instructions) or as complex mathematical equations that require simulation.



e.g., air flow



e.g., Second Life (3d virtual reality world)

Computational Models of Human Cognition

Computer simulation of neural and/or cognitive processes that underlie performance on a task

Goals of Computational Modeling in Cog Sci

- Understand mechanisms of information processing in the brain
- Explain behavioral, neuropsychological, and neuroscientific data
- Suggest techniques for remediation of cognitive deficits due to brain injury and developmental disorders
- Suggest techniques for facilitating learning in normal cognition
- Construct computer architectures to mimic human-like intelligence

Why Build Models?

- Forces you to be explicit about hypotheses and assumptions
- Provides a framework for integrating knowledge from various fields
- Allows you to observe complex interactions among hypotheses
- Provides ultimate in controlled experiment
- Leads to empirical predictions
- A mechanistic framework will ultimately be required to provide a unified theory of cortex.

Levels of Modeling

Single cell

ion flow, membrane depolarization, neurotransmitter release, action potentials, neuromodulatory interactions

Network

neurophysiology and neuroanatomy of cortical regions, cell firing patterns, inhibitory interactions, mechanisms of learning

Functional

operation and interaction of cortical areas, transformation of representations

Computational

input-output behavior, mathematical characterization of computation

Overview

Computational modeling

Artificial neural networks

Modeling performance after brain damage

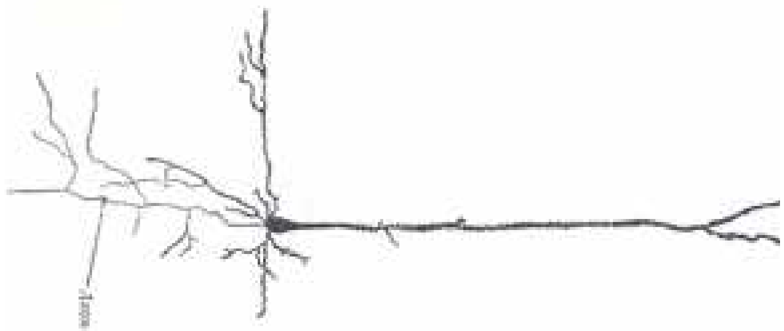
the neighborhood™ Jerry Van Amerongen

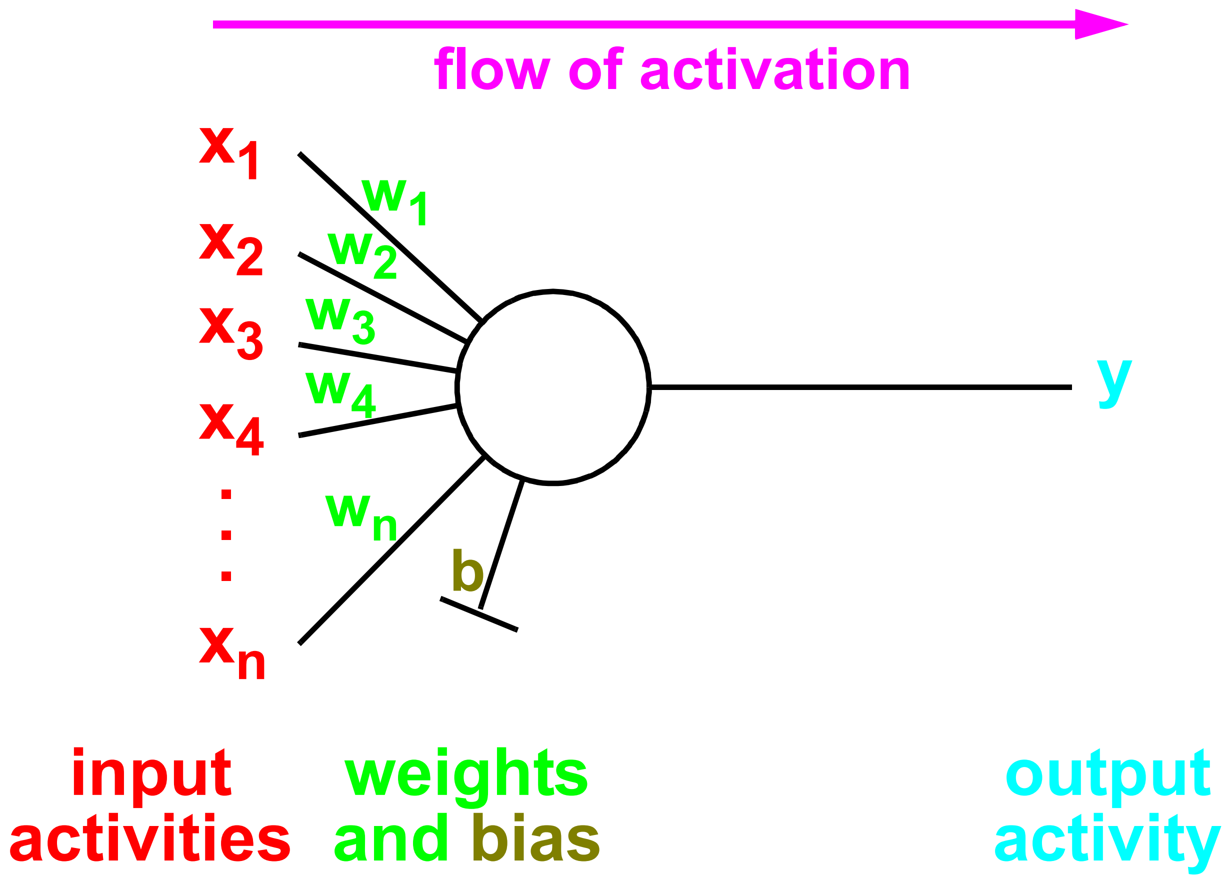
how the brain works

Key Features of Cortical Computation

- Neurons are slow ($10^{-3} - 10^{-2}$ propagation time)
- Large number of neurons ($10^{10} - 10^{11}$)
- No central controller (CPU)
- Neurons receive input from a large number of other neurons (10^4 fan-in and fan-out of cortical pyramidal cells)
- Communication via excitation and inhibition
- Statistical decision making (neurons that single-handedly turn on/off other neurons are rare)
- Learning involves modifying coupling strengths (the tendency of one cell to excite/inhibit another)
- Neural hardware is dedicated to particular tasks (vs. conventional computer memory)
- Information is conveyed by mean firing rate of neuron, a.k.a. *activation*

Modeling Individual Neurons



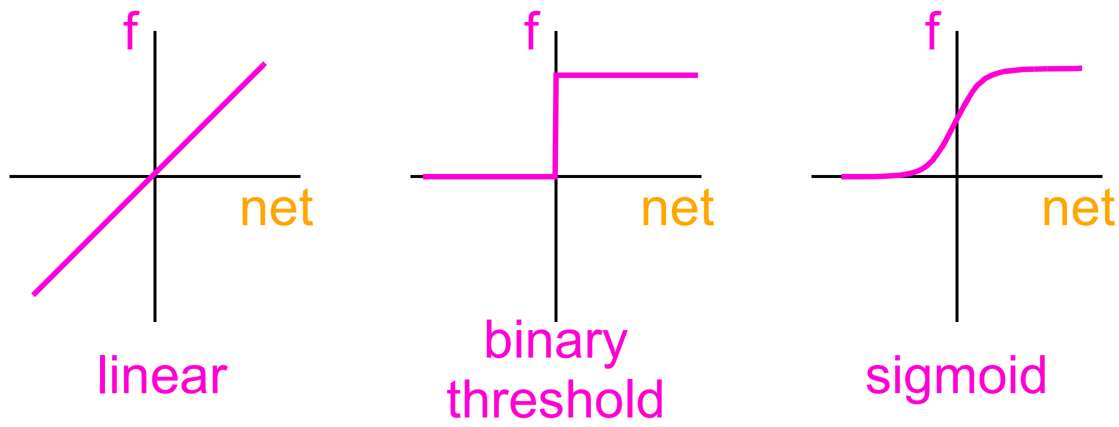


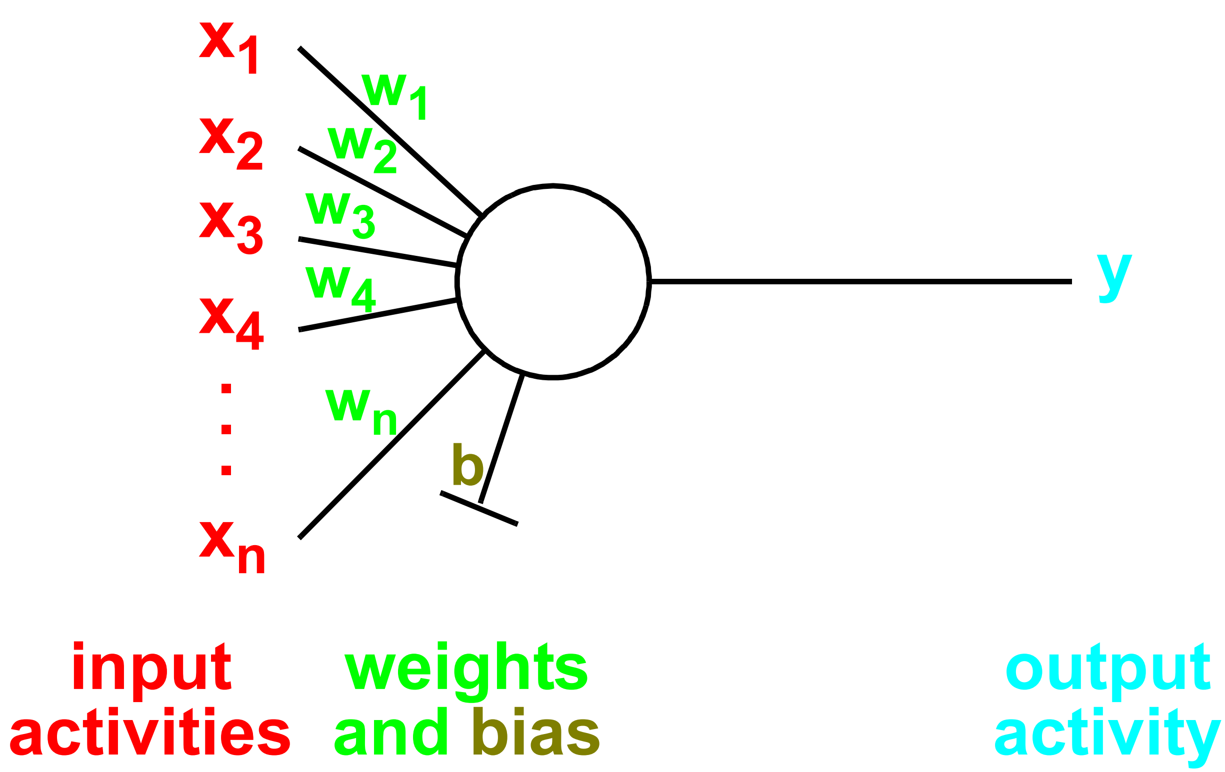
Modeling Individual Neurons

Activation function

$$net = \sum_i w_i x_i + b$$

$$y = f(net)$$

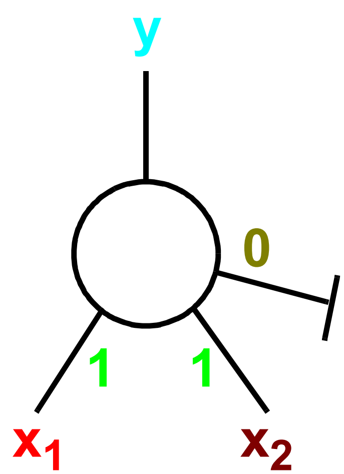




Computation With a Binary Threshold Unit

"Or" gate

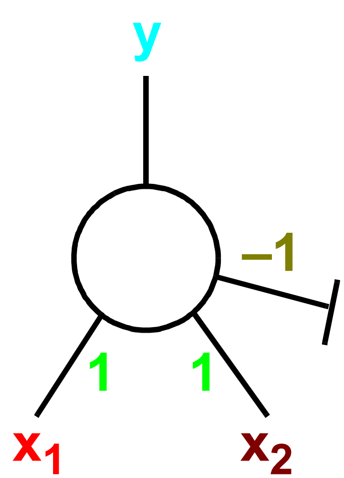
x1	x2	y
0	0	0
0	1	1
1	0	1
1	1	1



Computation With a Binary Threshold Unit

"And" gate

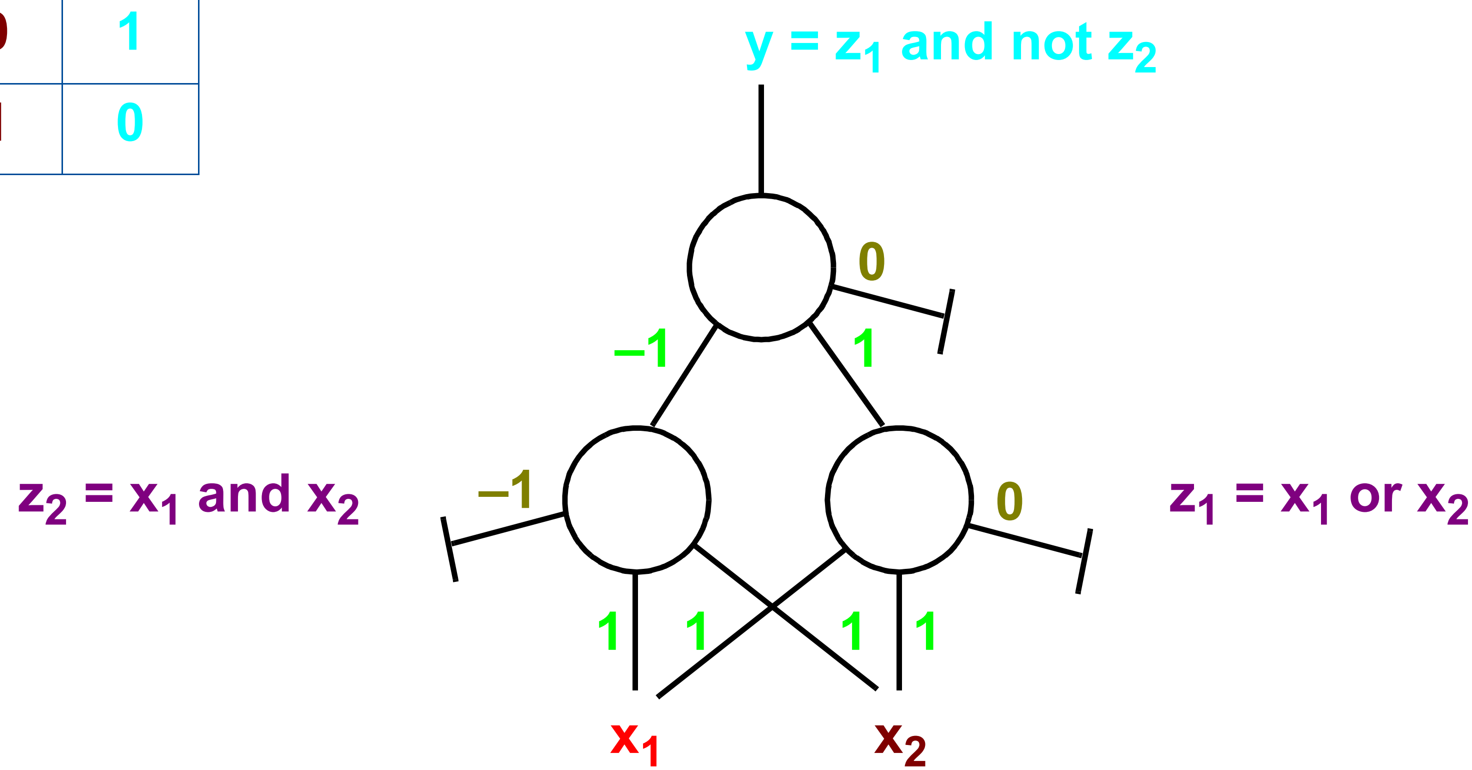
x1	x2	y
0	0	0
0	1	0
1	0	0
1	1	1



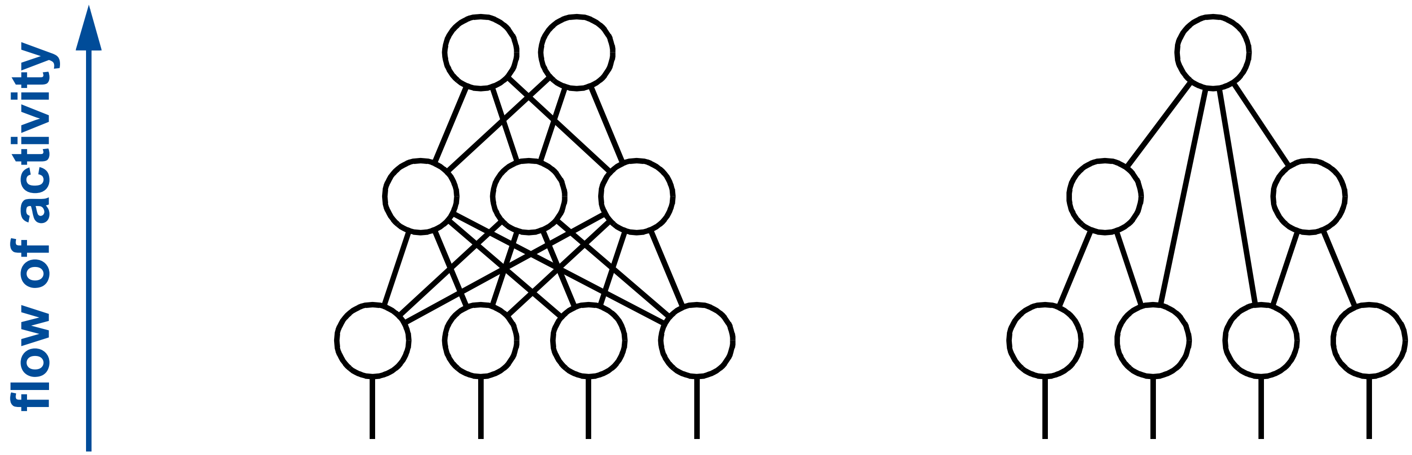
Computation With a Binary Threshold Unit

"Exclusive or" gate

x1	x2	y
0	0	0
0	1	1
1	0	1
1	1	0



Feedforward Architectures



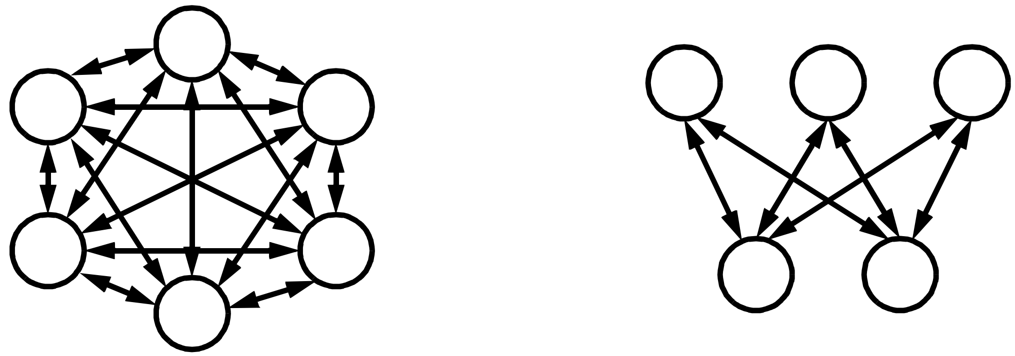
Activation flows in one direction; no closed loops

Performs association from input pattern to output pattern

big, hairy, stinky ➔ run away
small, round, orange ➔ eat
big, round, soft ➔ eat
small, orange, hairy ➔ run away
stinky, yellow ➔ eat

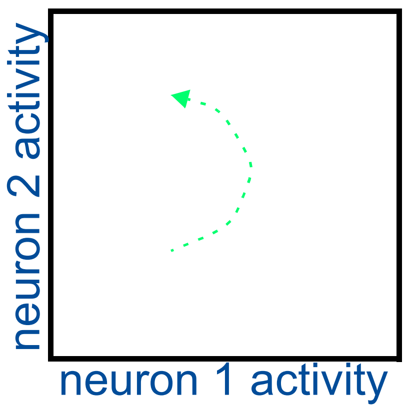
Learning: adjust connections to achieve input-output mapping

Recurrent Architectures

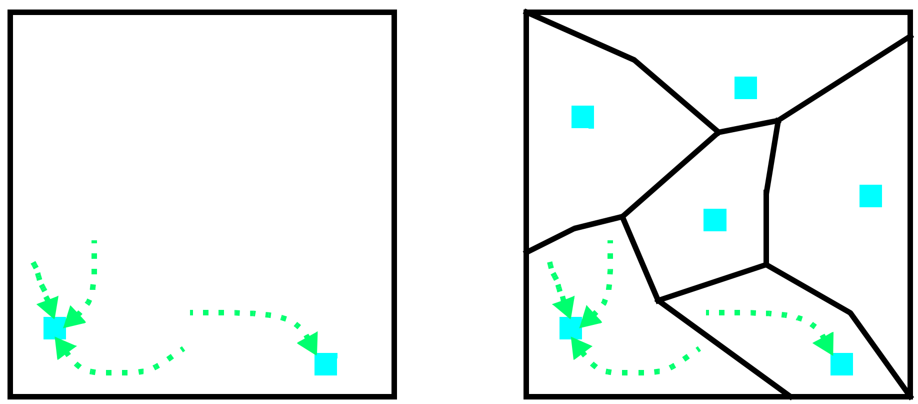


Achieves *best interpretation* of partial or noisy patterns, e.g., MAR – – M – LLOW

State space dynamics

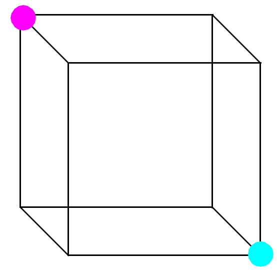


Attractor dynamics



Learning: establishes new attractors and shifts attractor basin boundaries

Necker Cube Example



Each vertex has two possible interpretations.

Interpretation of one vertex depends on interpretation of other vertices.

Constraint satisfaction problem (suitable for attractor net)

Necker Cube Demo

See http://www.cs.cf.ac.uk/Dave/JAVA/boltzman/Necker.html

Supervised Learning in Neural Networks

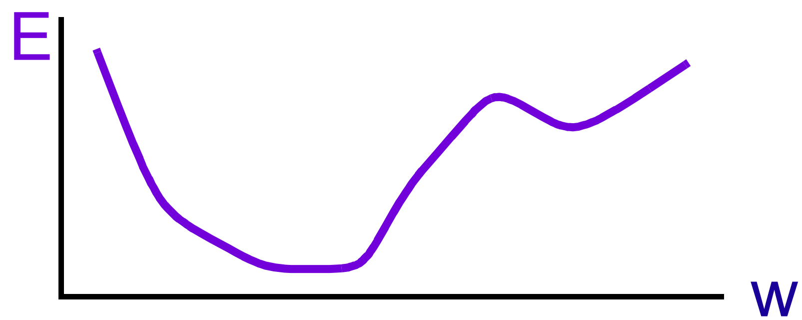
1. Assume a set of training examples, $\{\mathbf{x}^i, \mathbf{d}^i\}$

e.g., MAR – – M – LLOW ➔ MARSHMALLOW

e.g., big, hairy, stinky ➔ run away

2. Define a measure of network performance, e.g.,

$$E = \sum_i \left\| \mathbf{d}^i - \mathbf{y}^i \right\|^2$$



3. Make small incremental changes to weights to decrease error (*gradient descent*), i.e.,

$$\Delta w_{ji} \sim -\partial E / \partial w_{ji}$$

For multilayered sigmoidal neural networks, gradient descent update has a simple *local* form (depends on activity of neuron i and error associated with neuron j)

Modeling Neuropsychological Phenomena

Michael C. Mozer
Mark Sitton

Department of Computer Science and
Institute of Cognitive Science
University of Colorado, Boulder

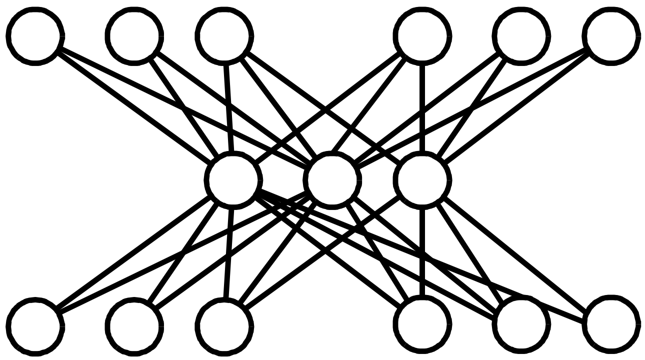
Martha Farah

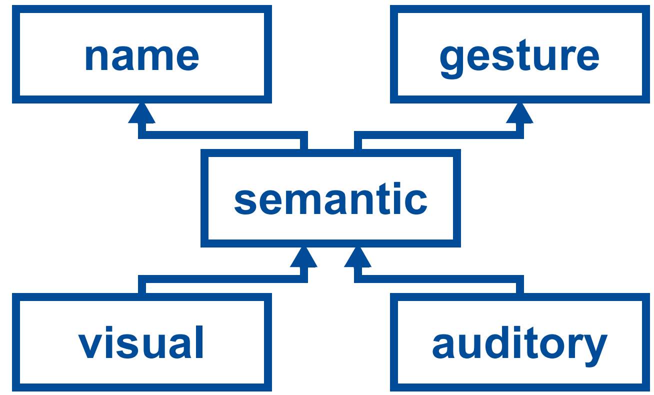
Department of Psychology
University of Pennsylvania

Optic Aphasia

- **Neuropathology: unilateral left posterior lesions**
- **Deficit in naming visually presented objects, in the absence of visual agnosia and general anomia**
  - Nonverbal indications of recognition: sorting, gesturing
  - Naming possible given verbal definition, tactile stimulation, object sounds

Modeling Naming and Gesturing

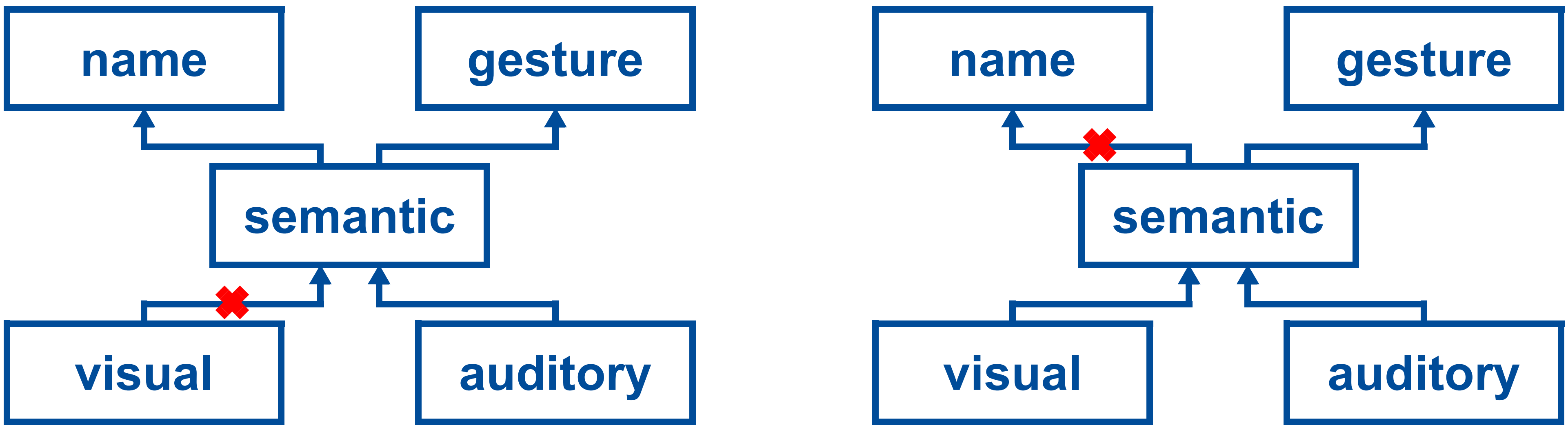




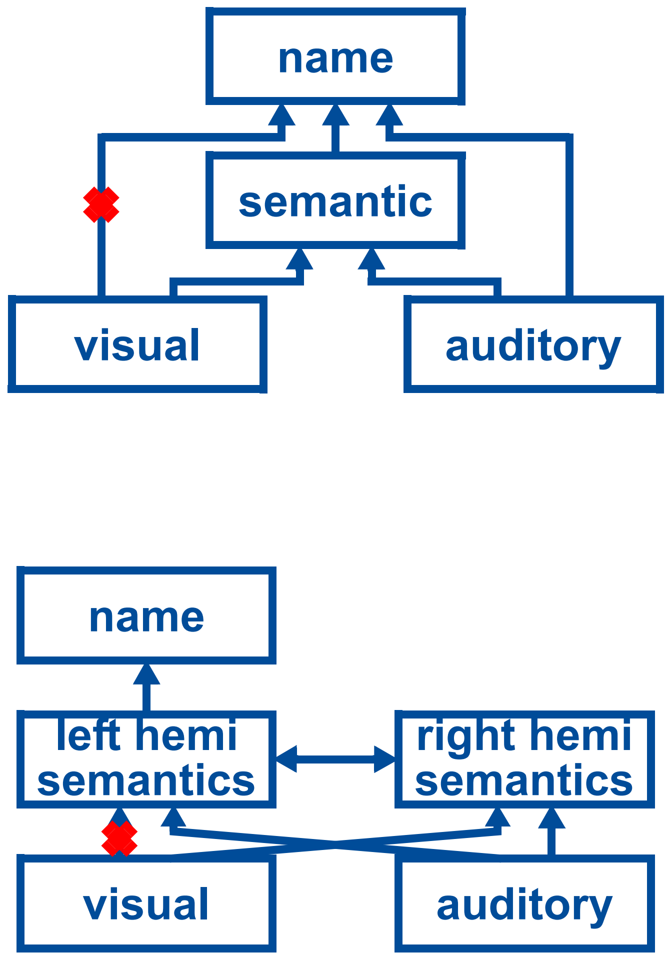
Each arrow represents a processing pathway (neural net)

Pathway act as associative memories

Simple Lesion Cannot Explain Optic Aphasia



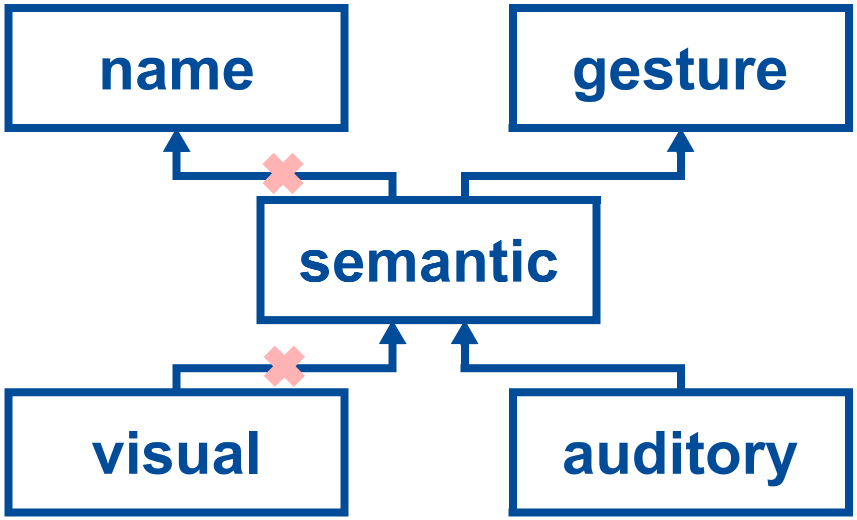
More Complex Architectures Are Unparsimonious



Alternative Explanation

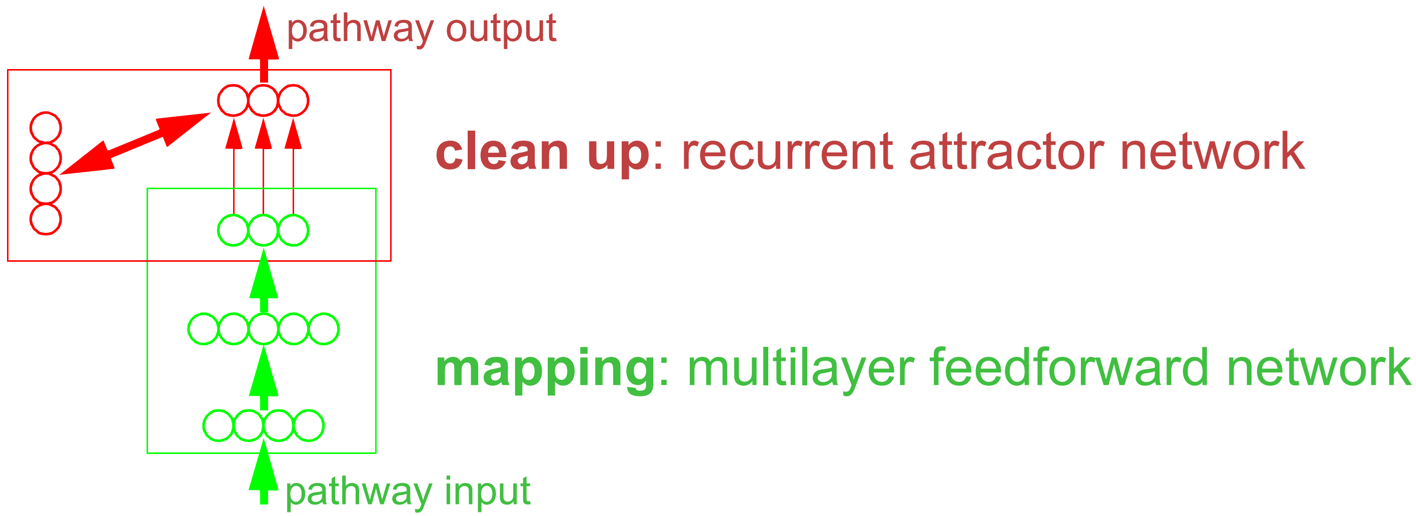
Partial damage to two systems (Farah, 1990)

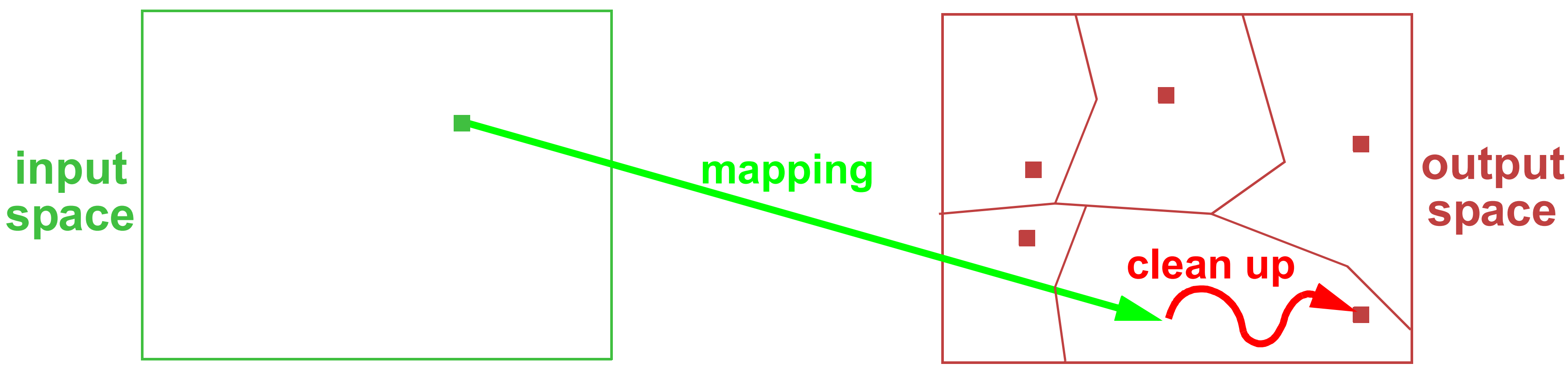
superadditive effect of damage



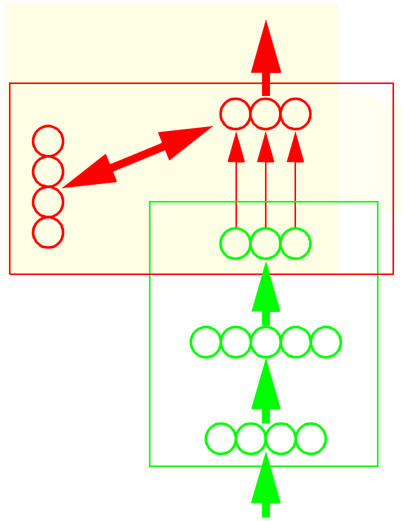
Simulation by Sitton, Mozer, and Farah (2000)

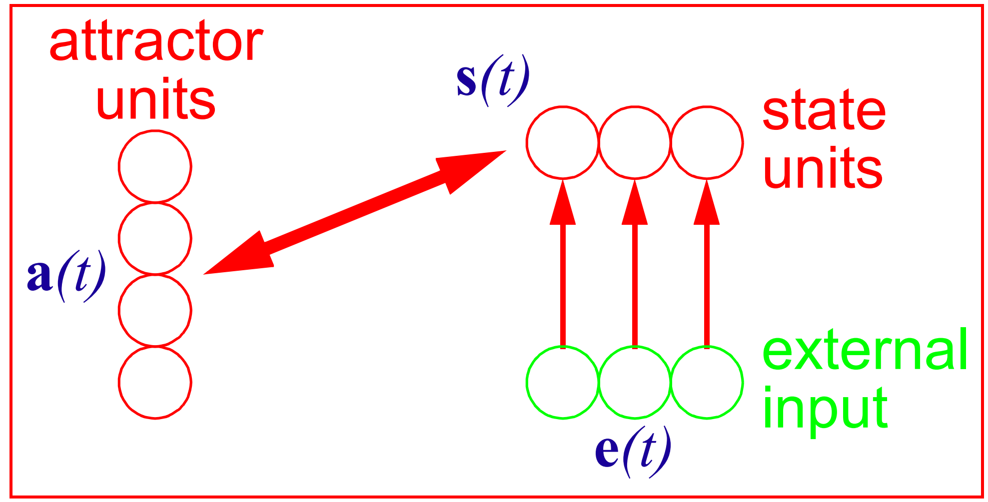
Neural Network Implementation of Pathway





Model Dynamics





μ_j: center of attractor *j*

β_j: width of attractor *j*

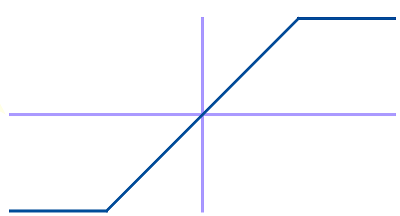
attractor unit update equation:

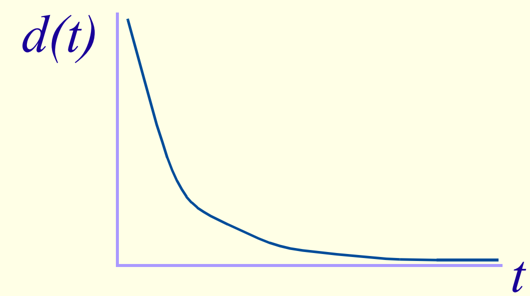
$$\hat{a}_j(t) = \exp(-\|\mathbf{s}(t)-\mu_j\|^2/\beta_j)$$

$$a_j(t) = \hat{a}_j(t)/\sum \hat{a}_i(t)$$

state unit update equation:

$$s_i(t) = h\left[d_i(t)e_i(t) + (1-d_i(t))\sum_j a_j(t-1)\mu_{ji}\right]$$





$$d_i(t) = 1 - \bar{e}_i(t-1)/e_i(t)$$

$$\bar{e}_i(t) = \alpha e_i(t) + (1-\alpha)\bar{e}_i(t-1)$$

Key Properties of Neural Network Pathway

Gradual convergence of pathway output on best interpretation over time

Continuous availability of information from other pathways

Simulation Methodology

Define neural activity patterns in visual, auditory, semantic, name, and gesture spaces

Pair patterns randomly

Train the four pathways to produce correct associations

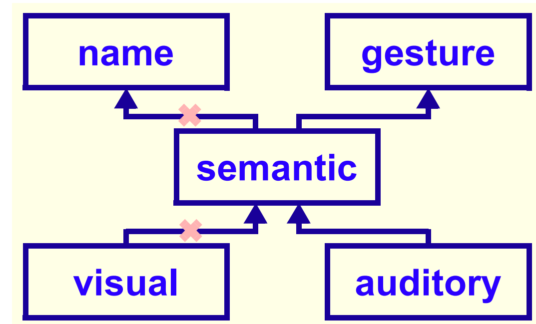
Lesion model

- Remove 30% of connections in V→S and S→N pathways

Evaluate lesioned model performance

Error Rates by Task

task	*error rate*	*damaged pathways*
A→G	0.0%	
A→N	0.5%	S→N
V→G	8.7%	V→S
V→N	36.8%	V→S, S→N



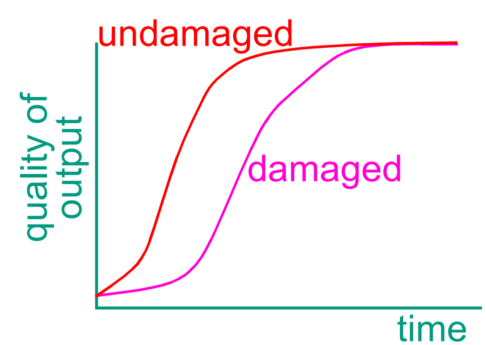
A→N: clean up compensates for S→N pathway damage

V→G: clean up compensates for V→S pathway damage

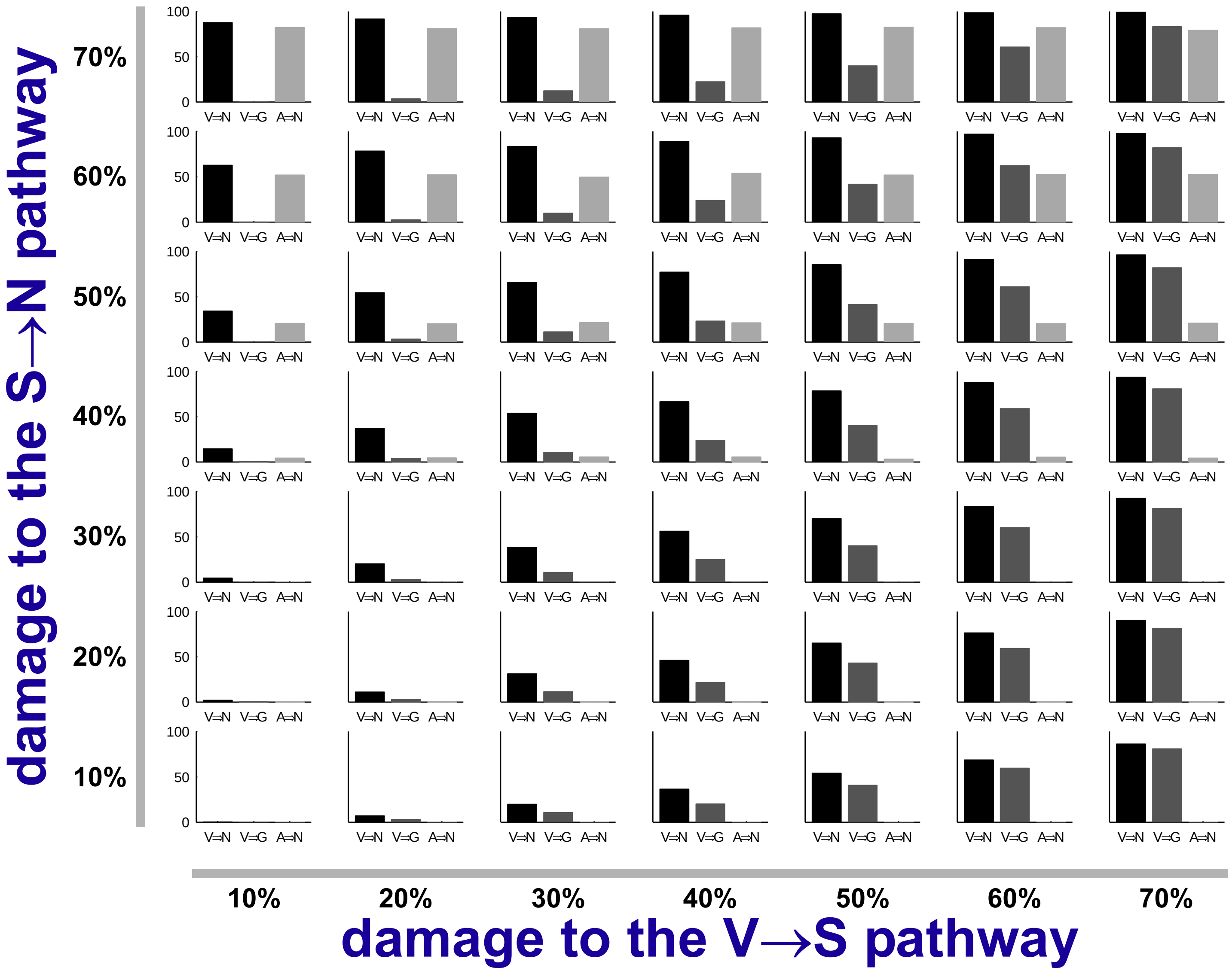
V→N: effects of damage to V→S and S→N pathways interact

noisy input + internal damage to S→N pathway

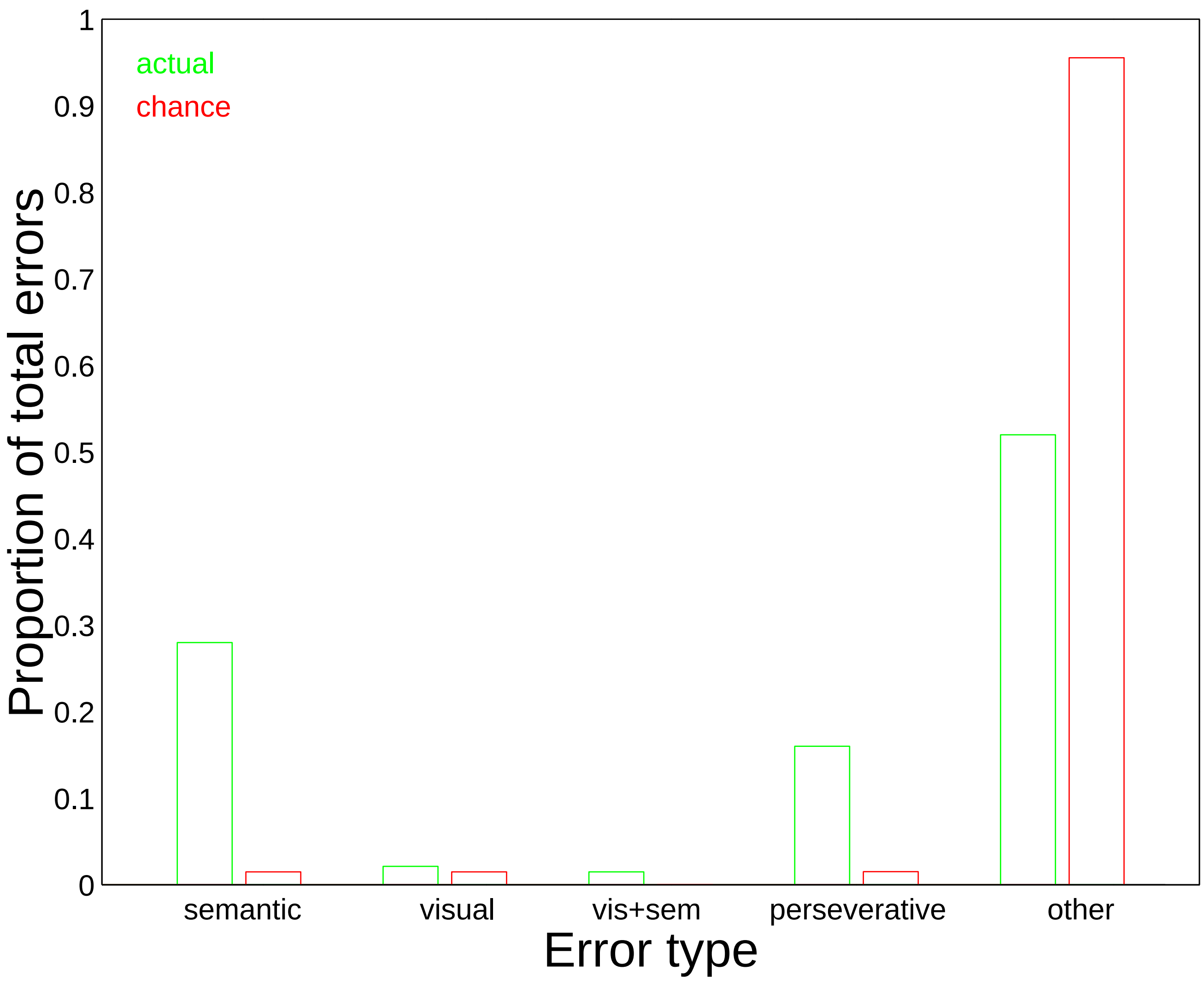
Interaction would not occur if
(a) pathways operated sequentially, or
(b) pathways showed no hysteresis



Error Rates Based on Relative Damage

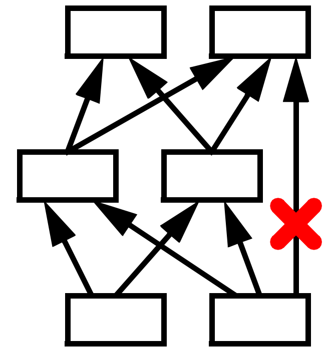


Distribution of Errors for Visual Object Naming

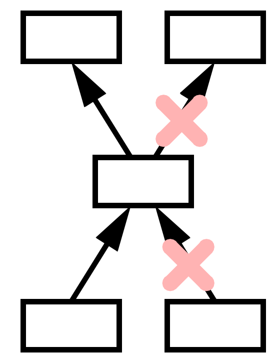


Interactivity in Brain Damage

Neuropsychological disorders have traditionally been explained by a focal lesion to a single processing pathway.



Farah (1990) argued that certain highly-selective deficits might have a parsimonious account in terms of multiple lesions with *interactive* effects.



The model illustrates the viability of this account.

Value of the Model

Past accounts have claimed the cognitive architecture is complex and unparsimonious.

multiple semantics systems or multiple functional pathways to naming

Instead, model can explain optic aphasia via a simple cognitive architecture and multiple lesions (each with a single dimension of selectivity).

Model can explain other aspects of phenomenon

e.g., naming errors tend to be semantic or perseverative, not visual

Model might be useful for understanding severity of lesions.