Computational Modeling of Human Cognition

Professor Michael C. Mozer

CSCI 3202

Computational Modeling

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

Goals

- 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

Modeling human learning

Modeling performance after brain damage

Using Testing to Enhance Learning: A Comparison of Two Hypotheses

Michael C. Mozer
Michael Howe

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

Harold Pashler

Department of Psychology
University of California, San Diego

Fact Learning

E.g., foreign language vocabulary

French word for dog is *chien.*

E.g., history trivia

The University of Colorado was founded in 1876.

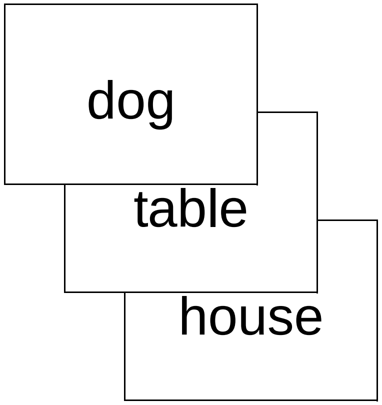
Facts can be framed as *cue – response* pairs.

e.g., dog – chien

e.g., Founding of University of Colorado – 1876

a.k.a. paired associate learning

Self Testing



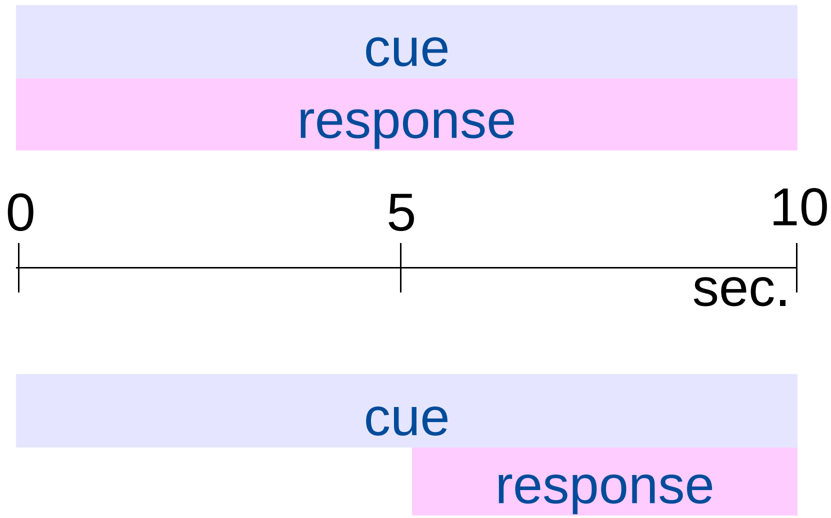
Does Self Testing Foster Learning?

Long history of empirical demonstrations, but many methodological difficulties.

Carrier and Pashler (1992)

Pure study (PS)

cue-response pair presented together for 10 sec.



Self testing (ST)

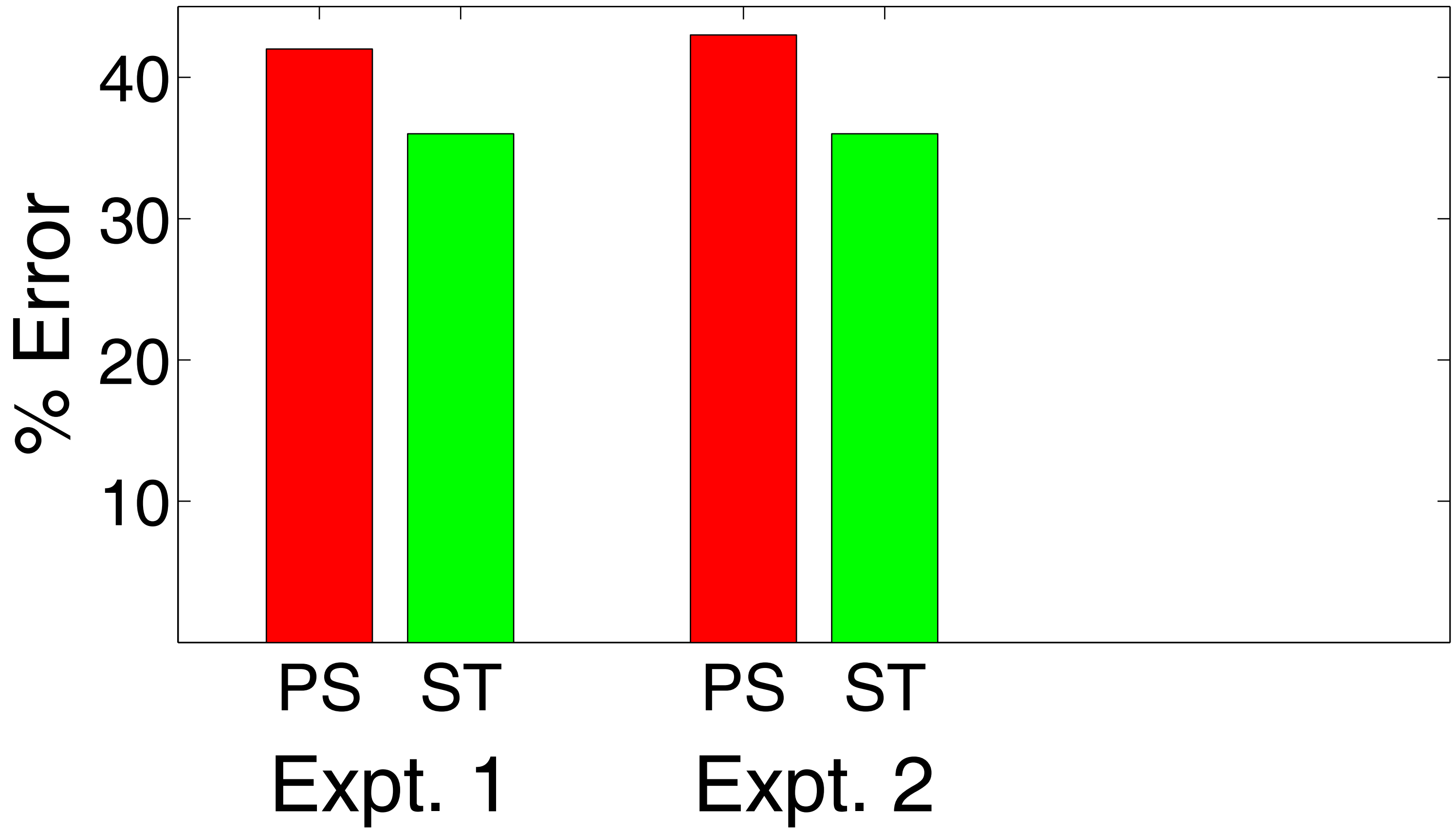
cue presented alone for 5 sec., during which response is to be retrieved

cue and response together for 5 sec., during which response is to be studied

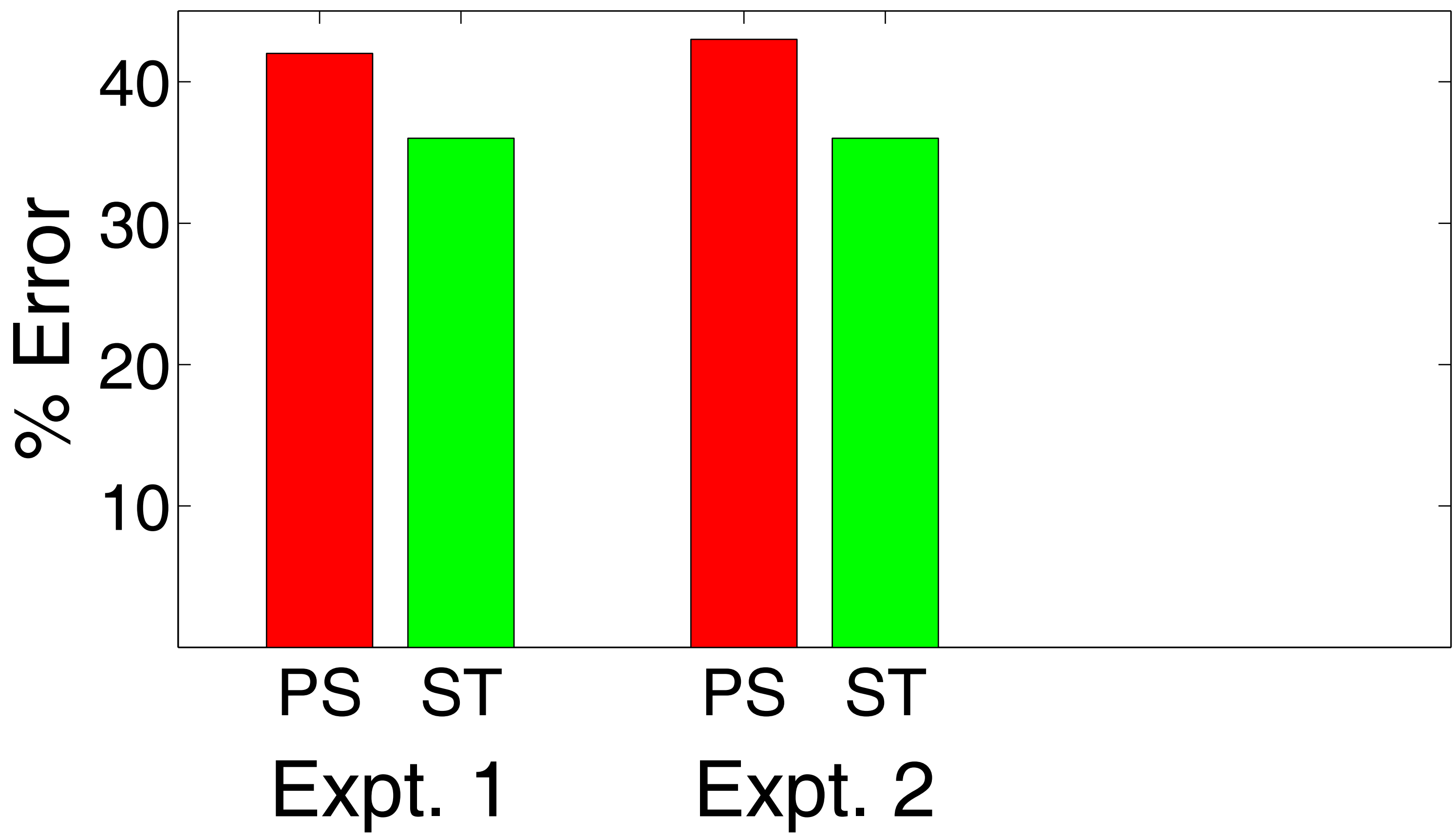
Design

- 20 items designated for PS, and 20 for ST
- 3 training blocks; all items studied in block 1
- final test phase; evaluation via cued recall
- Experiment 1: consonant trigrams – two-digit numbers
- Experiment 2: English – Siberian Eskimo Yupik word translation

Carrier and Pashler (1992)



Carrier and Pashler (1992)

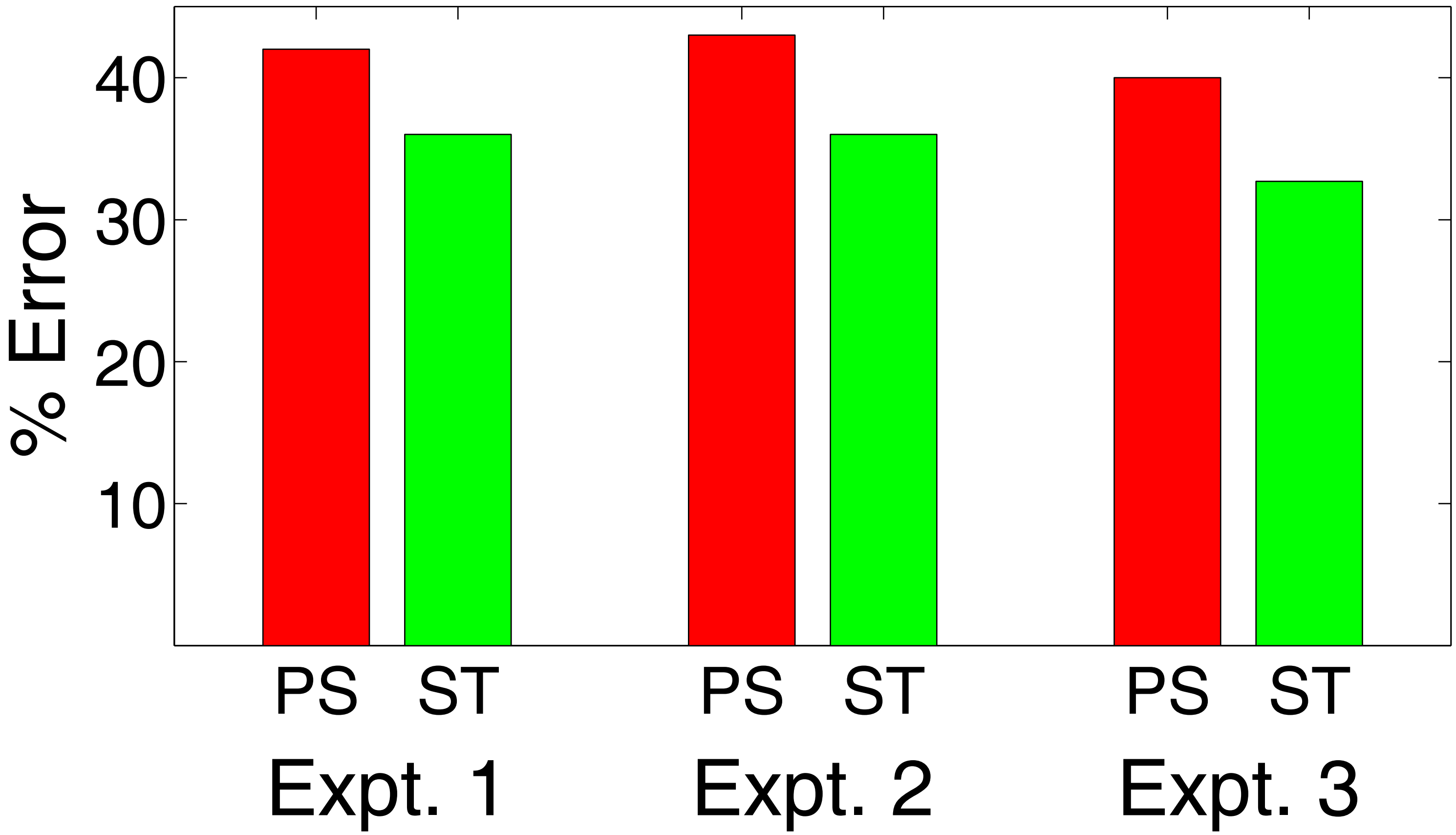


Possible explanation

Subjects used first self-test trial to assess item difficulty, and increased encoding effort on second self-test trial.

Expt. 3 same as Expt 2. except all items studied in first *two* training blocks

Carrier and Pashler (1992)



Some Explanations of Self-Testing Benefit

Landauer and Bjork (1978)

Retrieval attempts provide general boost to performance at a future time.

Incorrectly predicts that ST and PS items should benefit equally

Mandler (1979)

Cued recall strengthens structural, integrative information about an item.

Because cue and response are simultaneously activated for both ST and PS items, not clear why they wouldn’t both benefit.

Bjork (1975)

Act of retrieval strengthens existing retrieval routes to the response.

Consistent with data, but seems to require novel learning mechanisms

Basic Approach

Use a common, relatively noncontroversial architecture

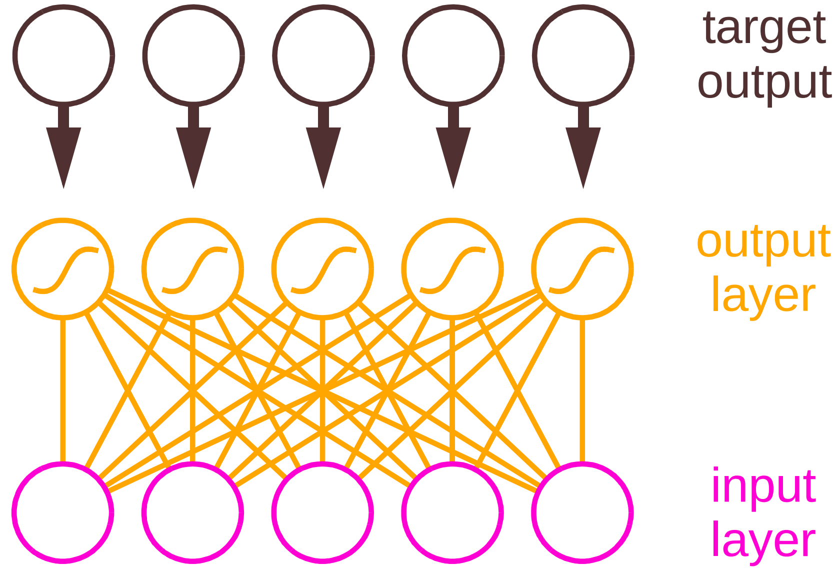
Feedforward neural network

Input layer connected to output layer

Standard sigmoidal activation function

Error correction learning

Widrow-Hoff (a.k.a. LMS)
network generates *actual* output
teacher provides *target* output
training depends on actual – target



Basic Approach

Use a common, relatively noncontroversial architecture

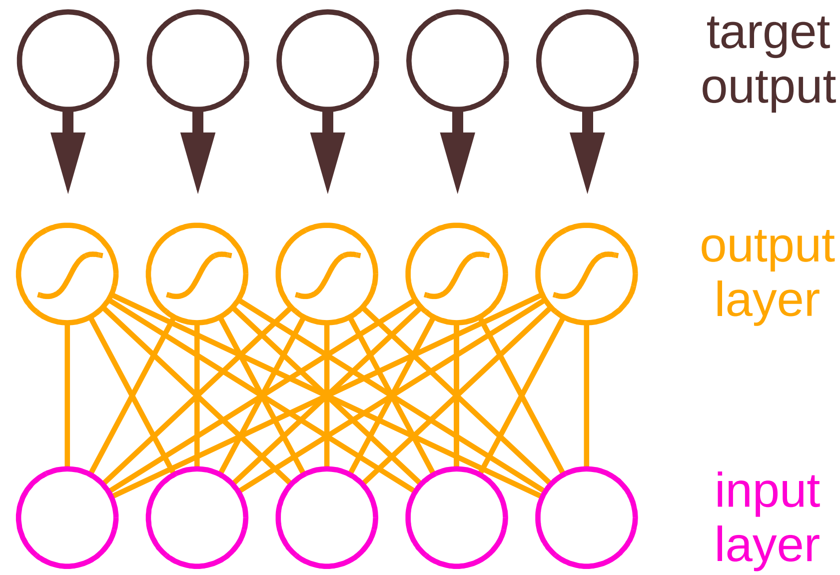
Feedforward neural network

Input layer connected to output layer

Standard sigmoidal activation function

Error correction learning

Widrow-Hoff (a.k.a. LMS)
network generates *actual* output
teacher provides *target* output
training depends on actual – target



Training of neural net often viewed as abstract procedure for loading knowledge into net.

Here, we make a stronger claim.

One training trial in neural net ~ one experimental trial

Hypothesis 1: Self-Generated Training Targets

Guthrie (1952)

One learns what one does.

ST item

Self test ⇨ candidate response ⇨ target for error-correction learning

Study ⇨ correct response ⇨ target for error-correction learning

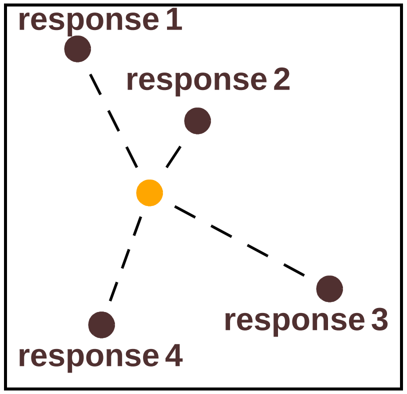
Both candidate and correct response are trained.

PS item

Only correct response is trained.

Choosing candidate response

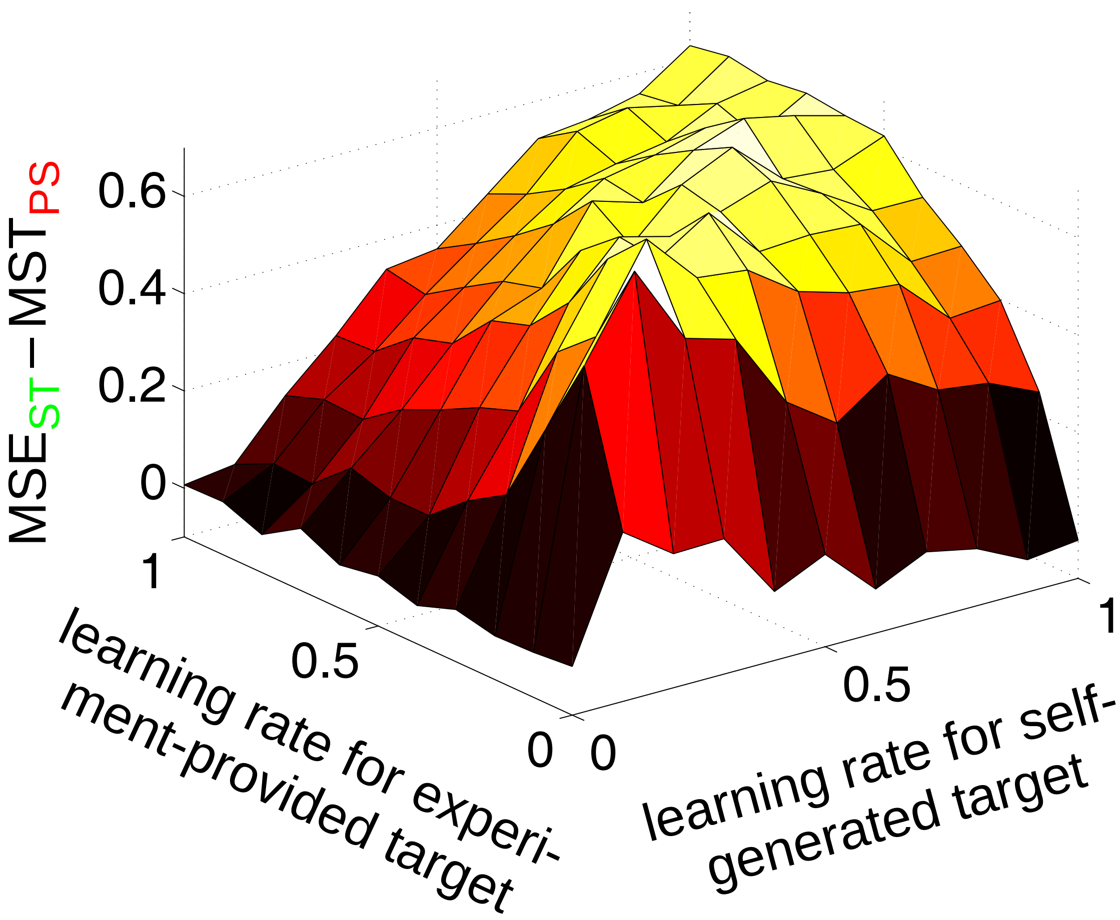
Probabilistic selection with Luce Choice Rule (a.k.a. Boltzmann distribution)



Hypothesis 1: Simulation Result

No parameter settings found that yield an enhancement of learning by testing.

In final test, mean-squared error (MSE) significantly higher for ST than PS items.



Hypothesis 2: Interruption of Cue Processing

Carrier and Pashler (1992)

Presentation of the response simultaneously with cue interrupts processing of the cue, making learning less efficient.

Hypothesis 2: Interruption of Cue Processing

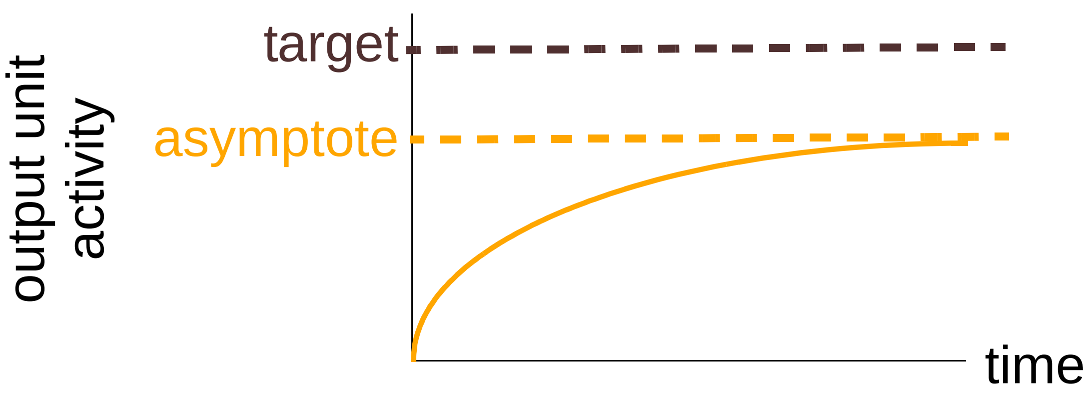
Carrier and Pashler (1992)

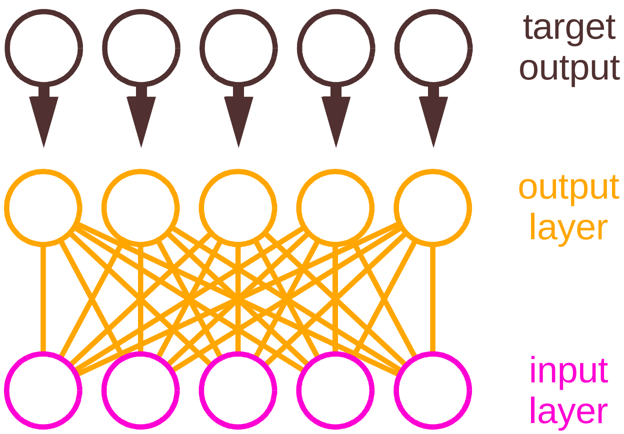
Presentation of the response simultaneously with cue interrupts processing of the cue, making learning less efficient.

Our interpretation

Units in neural net have temporal dynamics.

Leaky integrator model: $y_i(t) = \tau y_i(t-1) + (1-\tau)f(net_i(t))$





Hypothesis 2: Interruption of Cue Processing

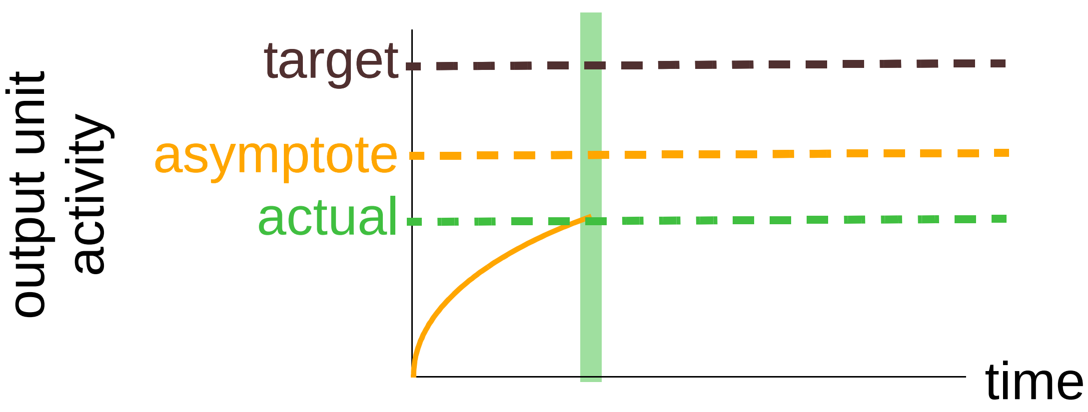
Carrier and Pashler (1992)

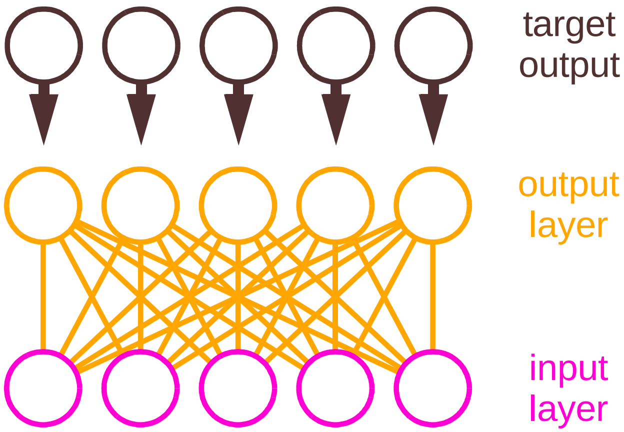
Presentation of the response simultaneously with cue interrupts processing of the cue, making learning less efficient.

Our interpretation

Units in neural net have temporal dynamics.

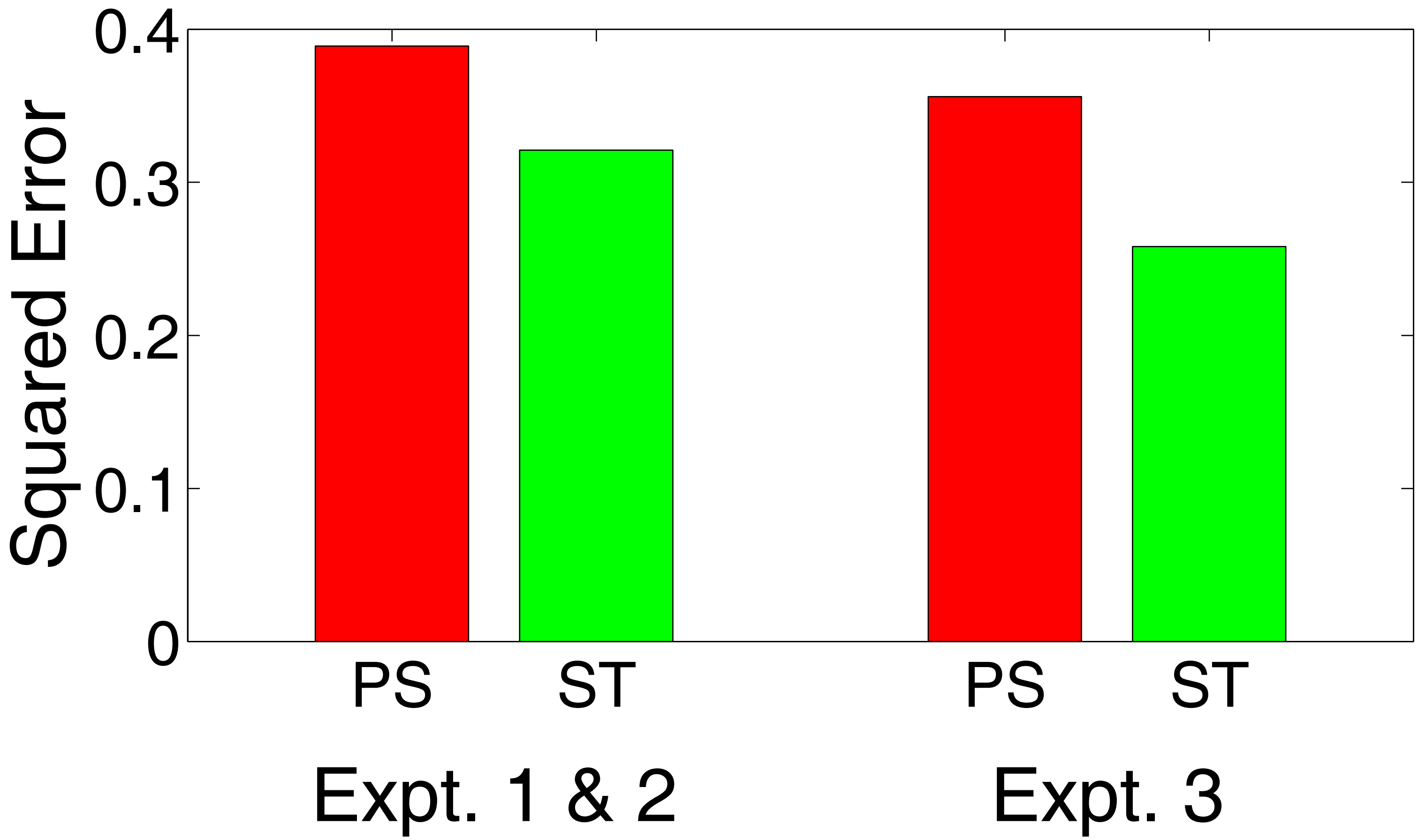
Leaky integrator model: $y_i(t) = \tau y_i(t-1) + (1-\tau)f(net_i(t))$





Presentation of correct response ⇨premature termination of processing ⇨ incorrect output ⇨ incorrect error signal

Hypothesis 2: Simulation Result



Summary

Goal

Explain the enhancement of learning through testing

Approach

In the context of a simple neural network model, we explored two alternative hypotheses.

(1) Testing obtains a candidate response whose association to the cue is strengthened, making the association less vulnerable to decay or interference.

(2) Error-correction learning requires comparison of the correct response to a candidate response. Testing forces a candidate response to be generated, whereas pure study does not.

Result

Simulations supported hypothesis 2, not hypothesis 1.

Modeling Neuropsychological Phenomena

Michael C. Mozer
Mark Sitton

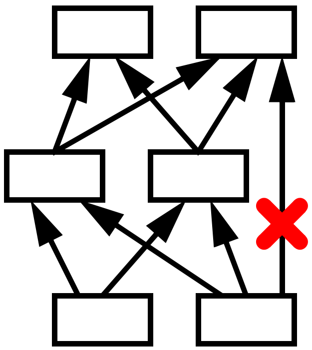
Department of Computer Science and
Institute of Cognitive Science
University of Colorado, Boulder

Martha Farah

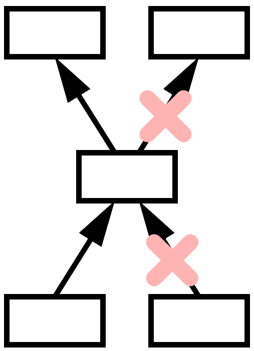
Department of Psychology
University of Pennsylvania

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.

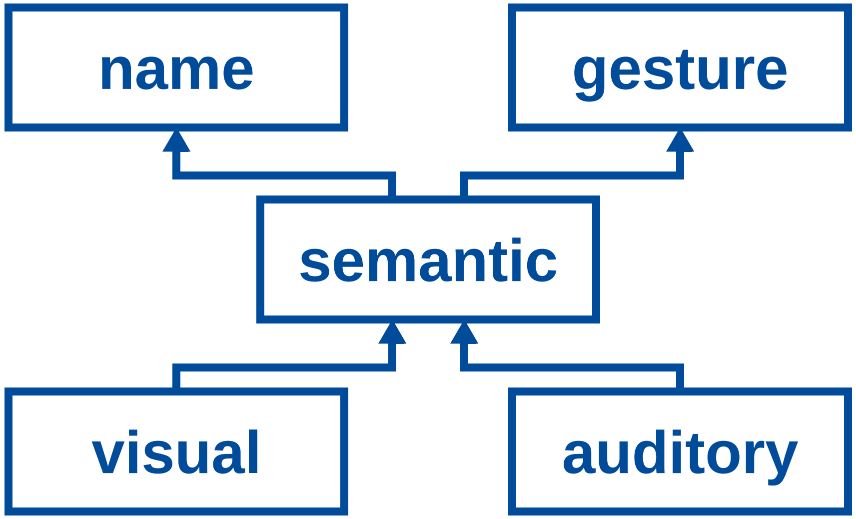


We illustrated the viability of this account via a neural network model of *optic aphasia*.

Optic Aphasia

- **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
- **Visual system roughly intact**
  - Insensitivity to visual quality; can copy drawings; normal interaction with world
- **No prosopagnosia**
- **Alexia**
- **Neuropathology: unilateral left posterior lesions, including occipital cortex and white matter**
- **Past accounts have postulated multiple semantics systems or multiple functional pathways to naming.**

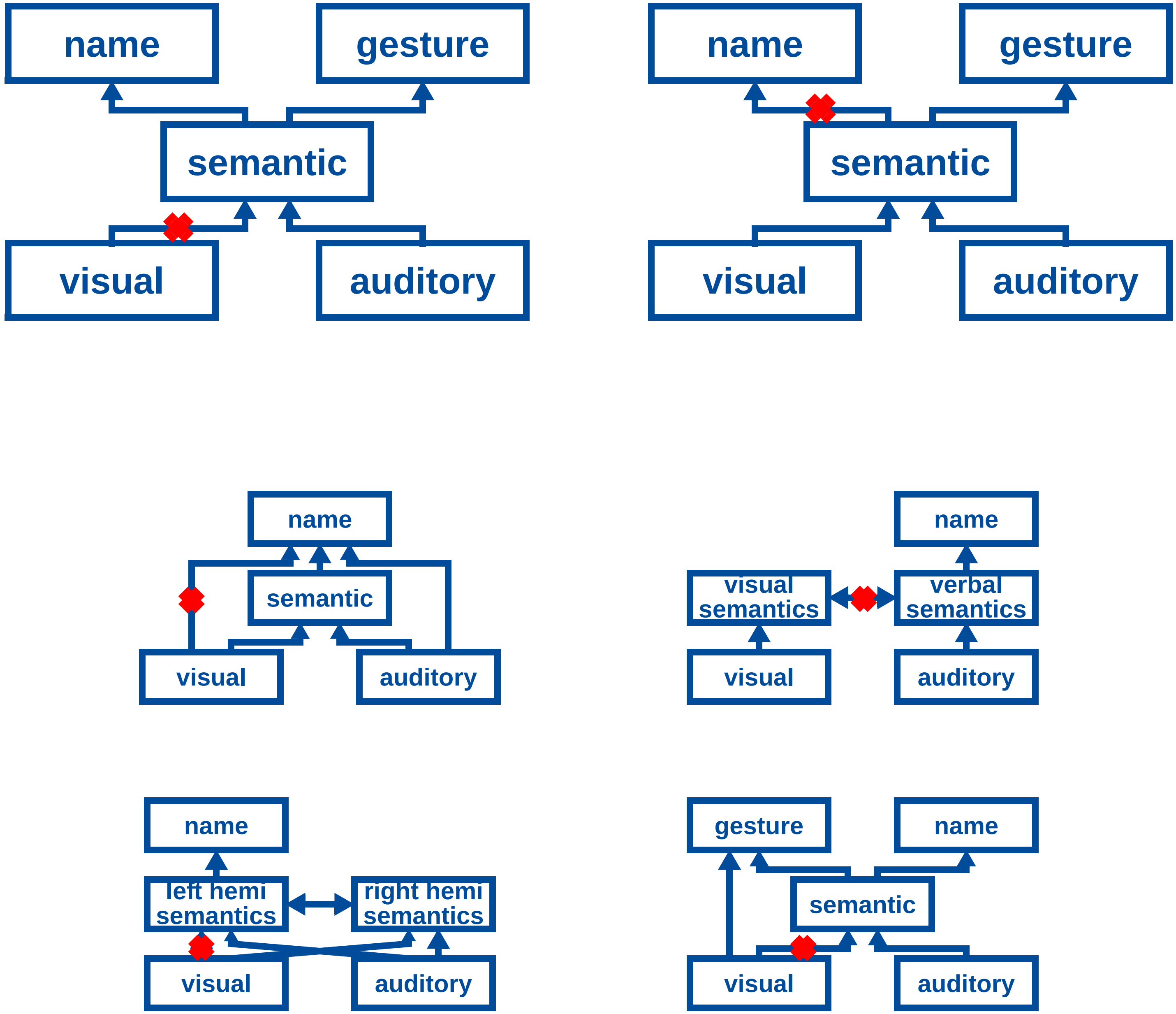
Cognitive Architecture



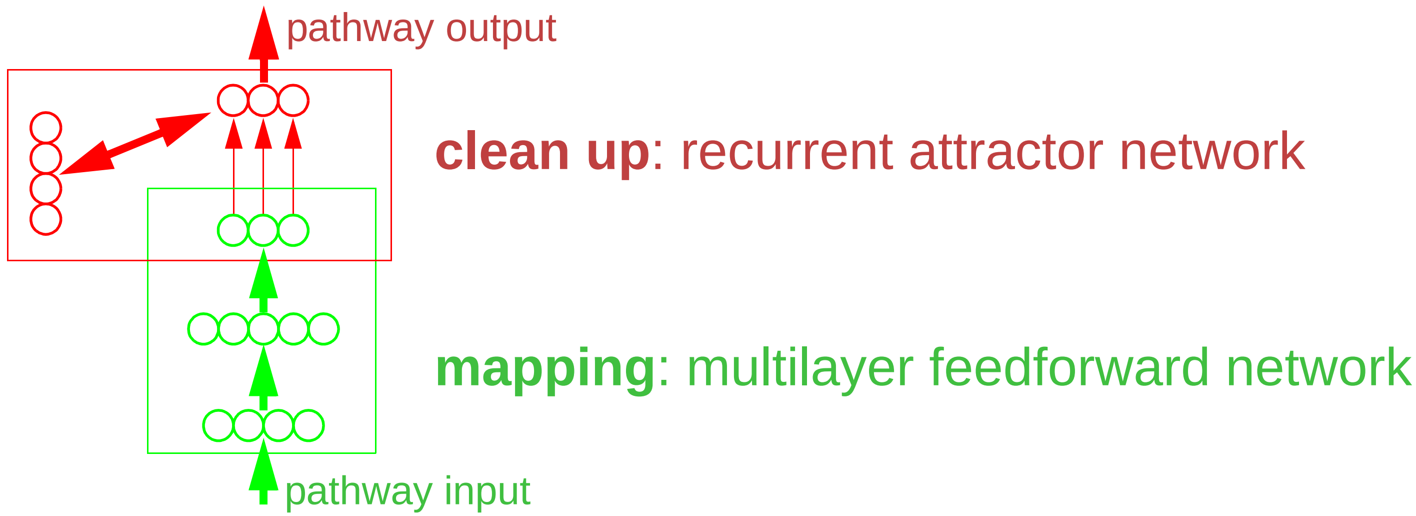
Each arrow represents a processing pathway (neural net)

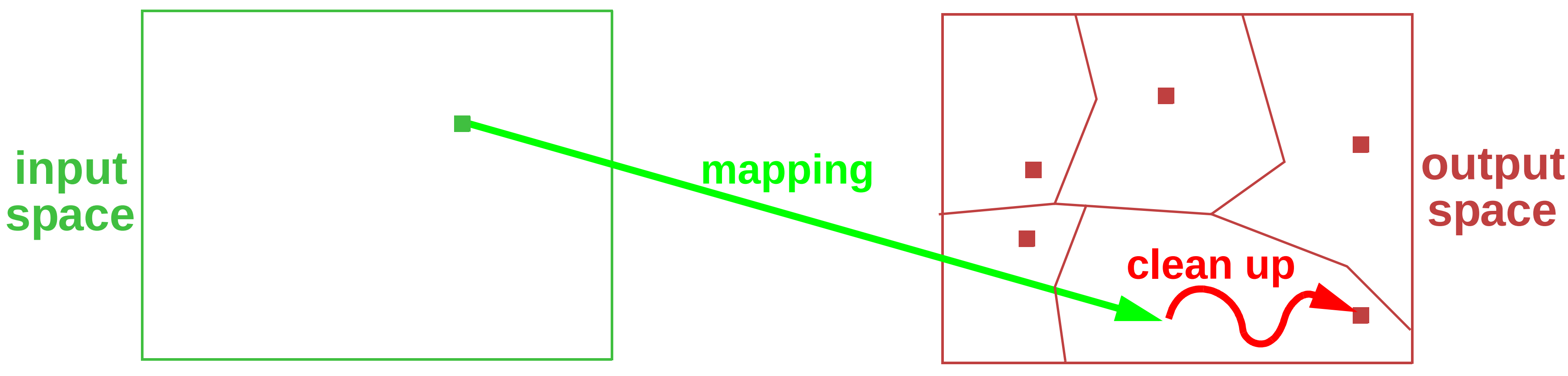
Pathway act as associative memories

Proposed Explanations are Unparsimonious

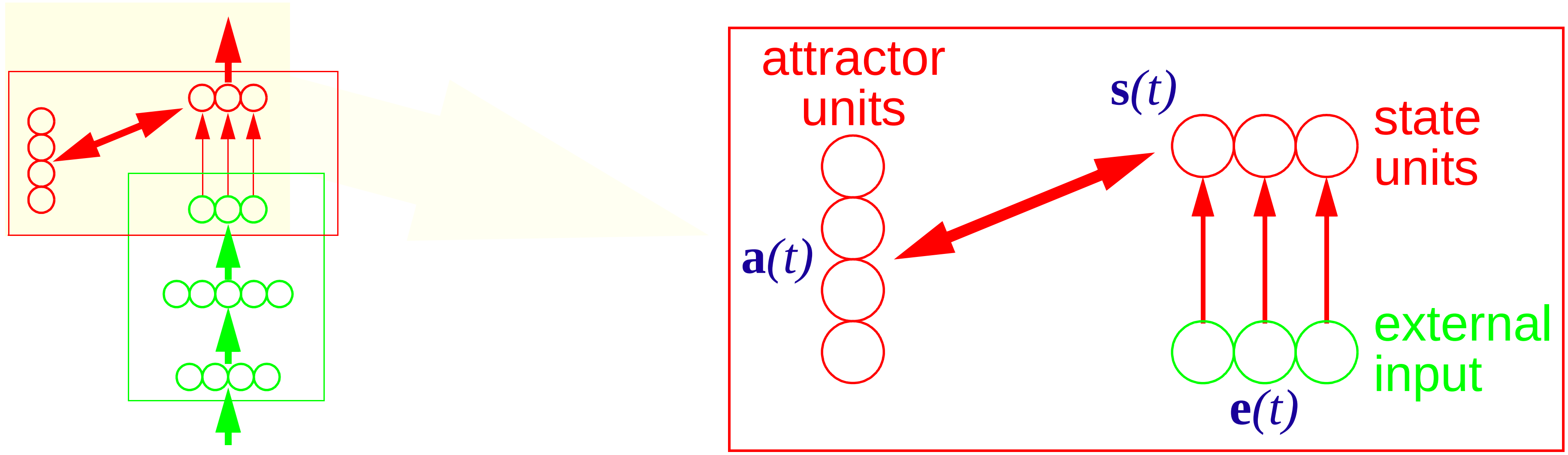


Neural Network Implementation of Pathway





Model Dynamics



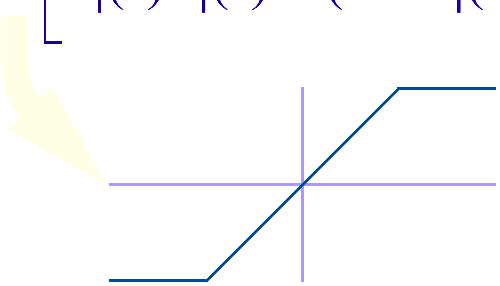
attractor unit update equation:

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

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

state unit update equation:

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



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

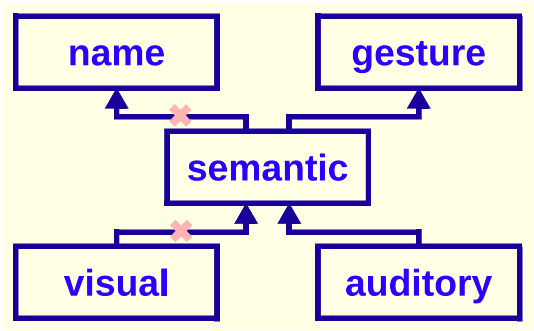
Lesion model

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

Evaluate lesioned model performance

Results of Partial Damage to Two Pathways

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**

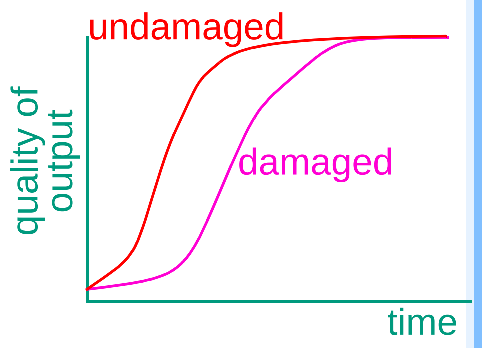


A→N, V→G: clean up compensates for damage

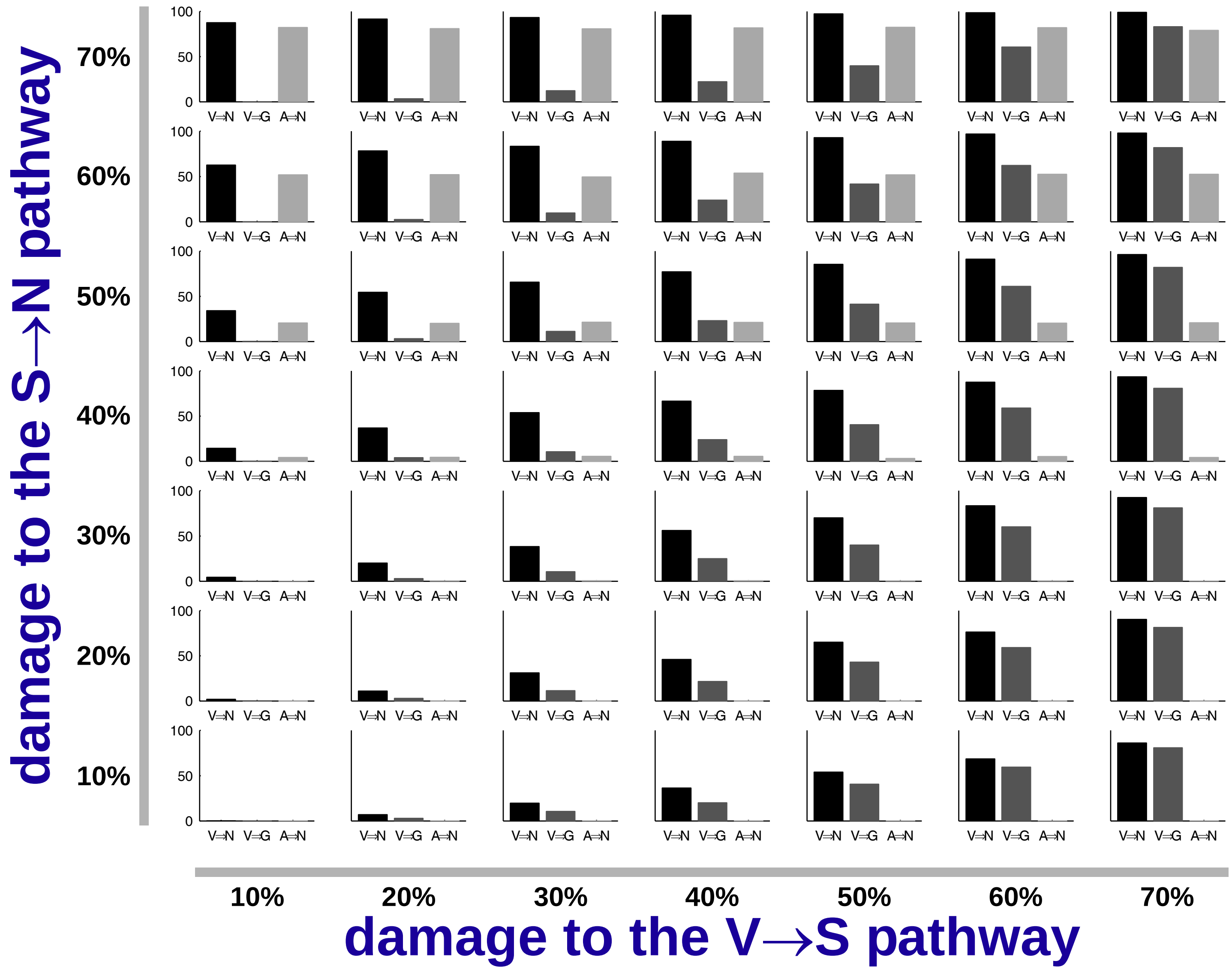
V→N: effects of damage to V→S and S→N pathways interact

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



Summary

Optic aphasia model can explain other aspects of phenomenon

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

Framework may be useful for explaining other highly selective cognitive impairments

Deficit in verb naming and reading aloud, versus deficit in writing responses involving nouns

Unilateral neglect to just faces, human bodies, number, or words

By hypothesizing multiple lesions, each with a single dimension of selectivity, we can account for highly selective deficits without positing implausible, counterintuitive, and unparsimonious cognitive architectures.