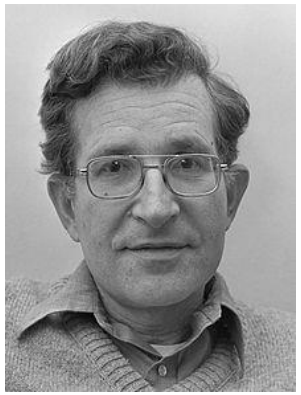
Neural Networks for Cognition and Cognitive Neuroscience: One Scientist's Perspective

Jay McClelland

Competing views of mind

- Rosenblatt: intuitive, statistical, experience driven
- Chomsky: formal, symbolic, innately structured
- Newell and Simon, Minsky: Discrete, symbolic, captured in a computer program

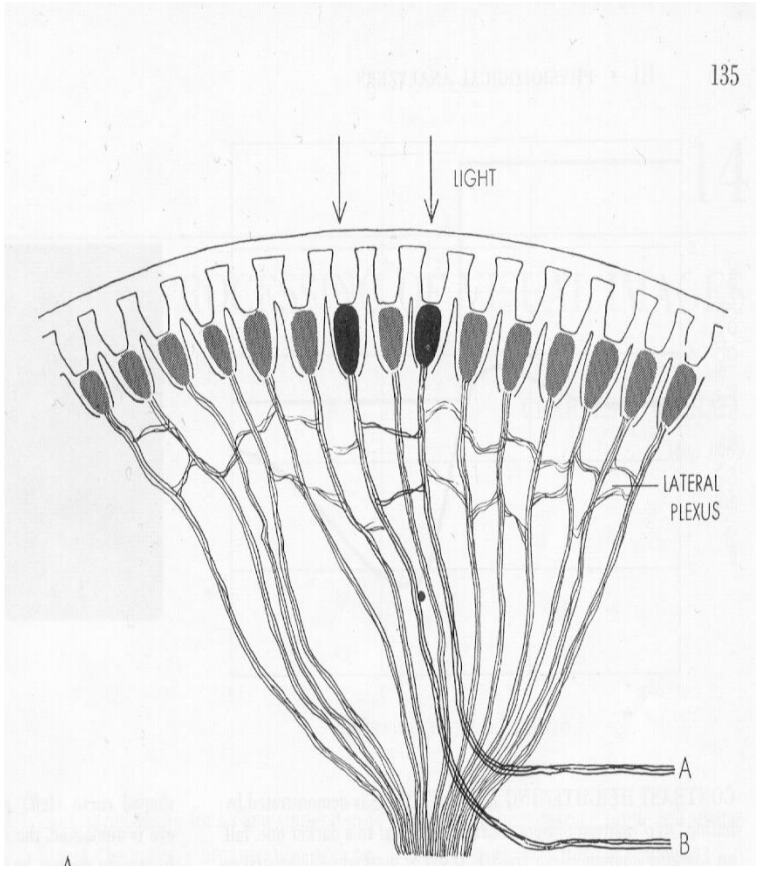


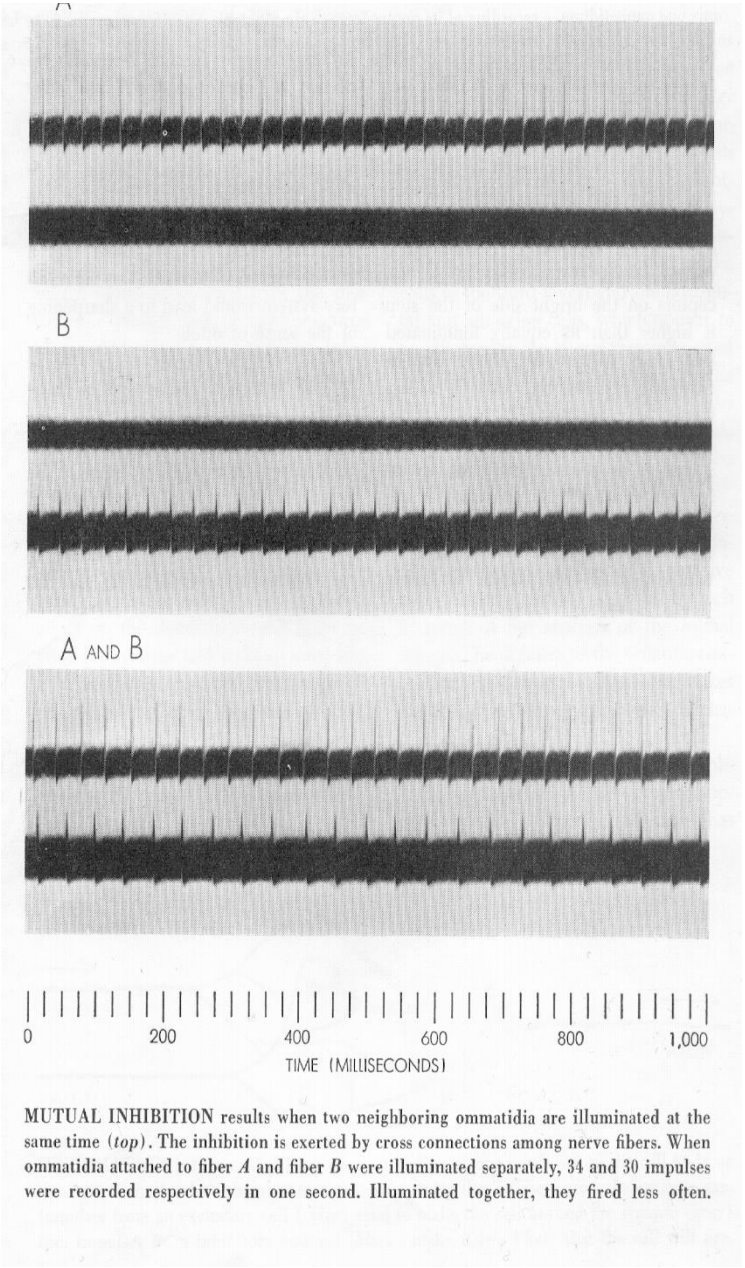
What is the relationship between brain and cognition?

- Neisser (1967), *Cognitive Psychology:* “Only of peripheral interest”

- Really?

Lateral Inhibition in Eye of Limulus (Horseshoe Crab)

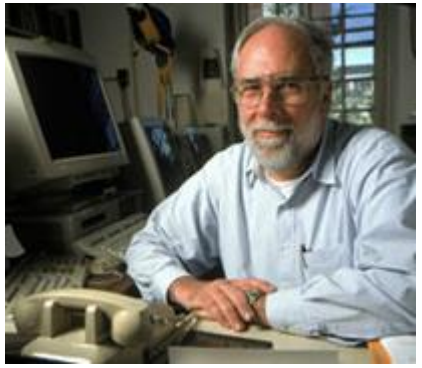




MUTUAL INHIBITION results when two neighboring ommatidia are illuminated at the same time (*top*). The inhibition is exerted by cross connections among nerve fibers. When ommatidia attached to fiber *A* and fiber *B* were illuminated separately, 34 and 30 impulses were recorded respectively in one second. Illuminated together, they fired less often.

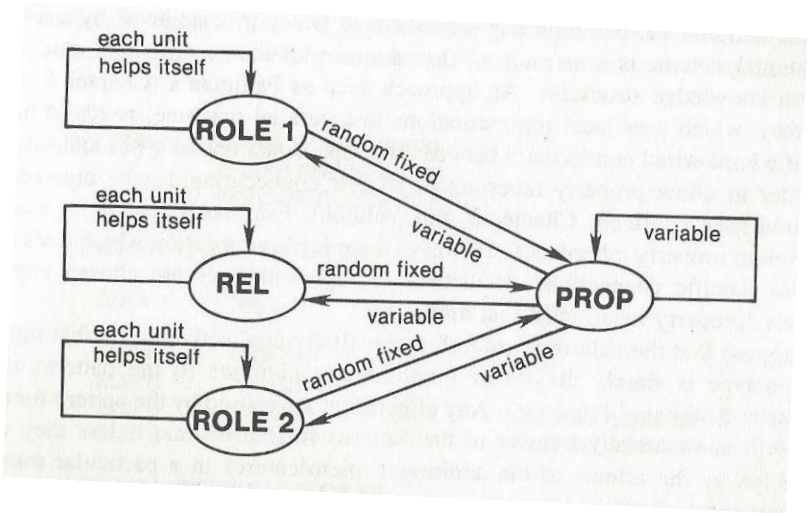
Grossberg and Anderson

- Grossberg:
  - Showed how competitive pools of neurons could capture many phenomena in perception
  - Introduced competitive learning
  - Envisioned completely neural models of mind
- Anderson:
  - Used vector/matrix approach to model
    - Attractor dynamics and learning



Rumelhart & Hinton

- Rumelhart became dissatisfied with symbolic AI
  - Argued that perception and comprehension require graded constraint satisfaction
  - Couldn't easily capture that using Lisp-like representations of knowledge and processing
- Hinton explored neural network models and applied them to semantics
  - Elephant – Size -- ??



Moore’s law in perspective

- 1957: x flops/sec
- 1977: 1,000x flops/sec
- 1997: 1,000,000x flops/sec
- 2017: 1,000,000,000x flops/sec
- 2037: 1,000,000,000x flops/sec
- Rosenblatt was given meagre allotments of precious 1957 computer time (one night, for one critical experiment!)
- In 1977, we got our first desktop terminals into a multi-user computer and could use the computer 24/7

The Interactive Activation Model

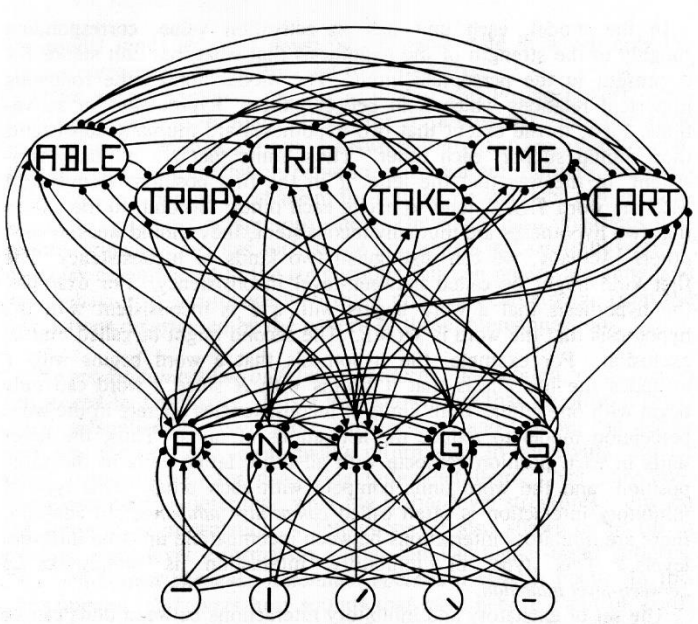


FIGURE 7. The unit for the letter *T* in the first position of a four-letter array and some of its neighbors. Note that the feature and letter units stand only for the first position; in a complete picture of the units needed from processing four-letter displays, there would be four full sets of feature detectors and four full sets of letter detectors. (From "An Interactive Activation Model of Context Effects in Letter Perception: Part 1. An Account of Basic Findings" by J. L. McClelland and D. E. Rumelhart, 1981, *Psychological Review, 88,* p. 380. Copyright 1981 by the American Psychological Association. Reprinted by permission.)

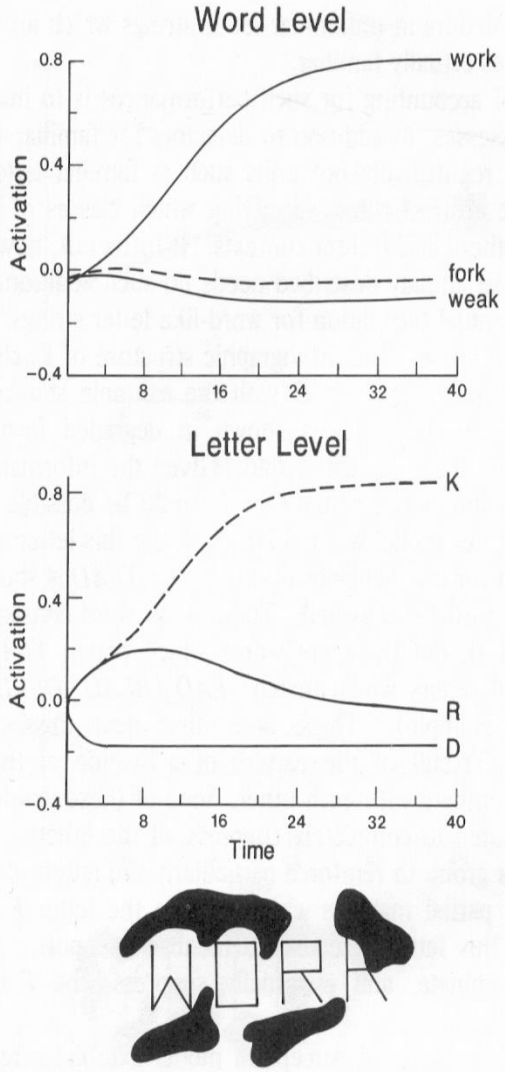
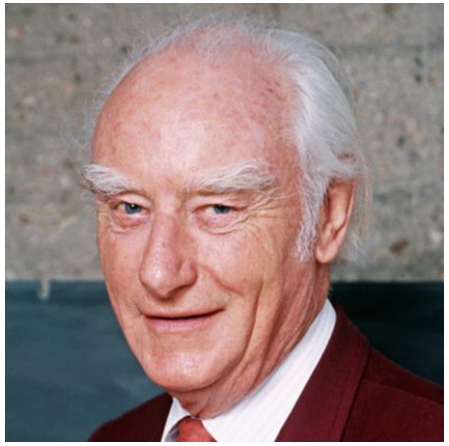


FIGURE 8. A possible display which might be presented to the interactive activation model of word recognition, and the resulting activations of selected letter and word units. The letter units are for the letters indicated in the fourth position of a four-letter display.

The PDP Research Group

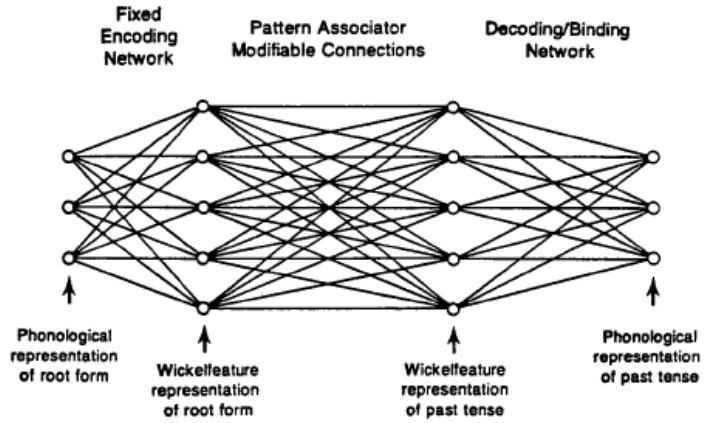
- Brought together mathematical, computational, psychological, and neuroscientific perspectives
- Explored models of cognition based on neural networks
- Boltzmann machines and back-propagation were developed along the way

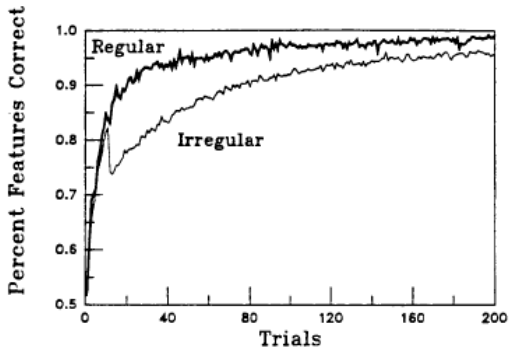



Are rules a thing of the past?

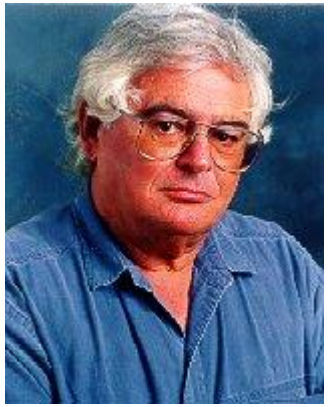
- Learned from pairs like:
  - Like-liked, love-loved, bake-baked
  - Go-went, give-gave, take-took
- Produced over-regularization errors
  - Take-taked
- Generalized to novel strings
  - Catch-catched
- Captured the quasi-regularity in exceptions and extended it to novel forms
  - Weep-wept
  - Bid-bid





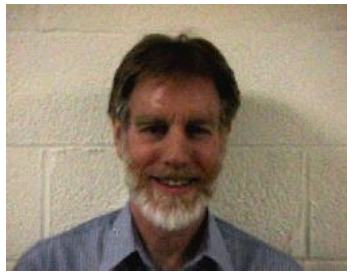
Pinker & Fodor

- Pinker thought PDP captured processing of irregulars, but argued that we still need a system of ‘algebra-like rules’ governed by strict linguistic categories
  - ‘No mere mortal ever flew out to left field’
  - Child learns the past tense rule in a ‘Eureka moment’
- Fodor argued that neural networks could not capture the systematic aspects of cognition
  - Only an absolute system of structure sensitive words can allow language to express the full range of expressible propositions



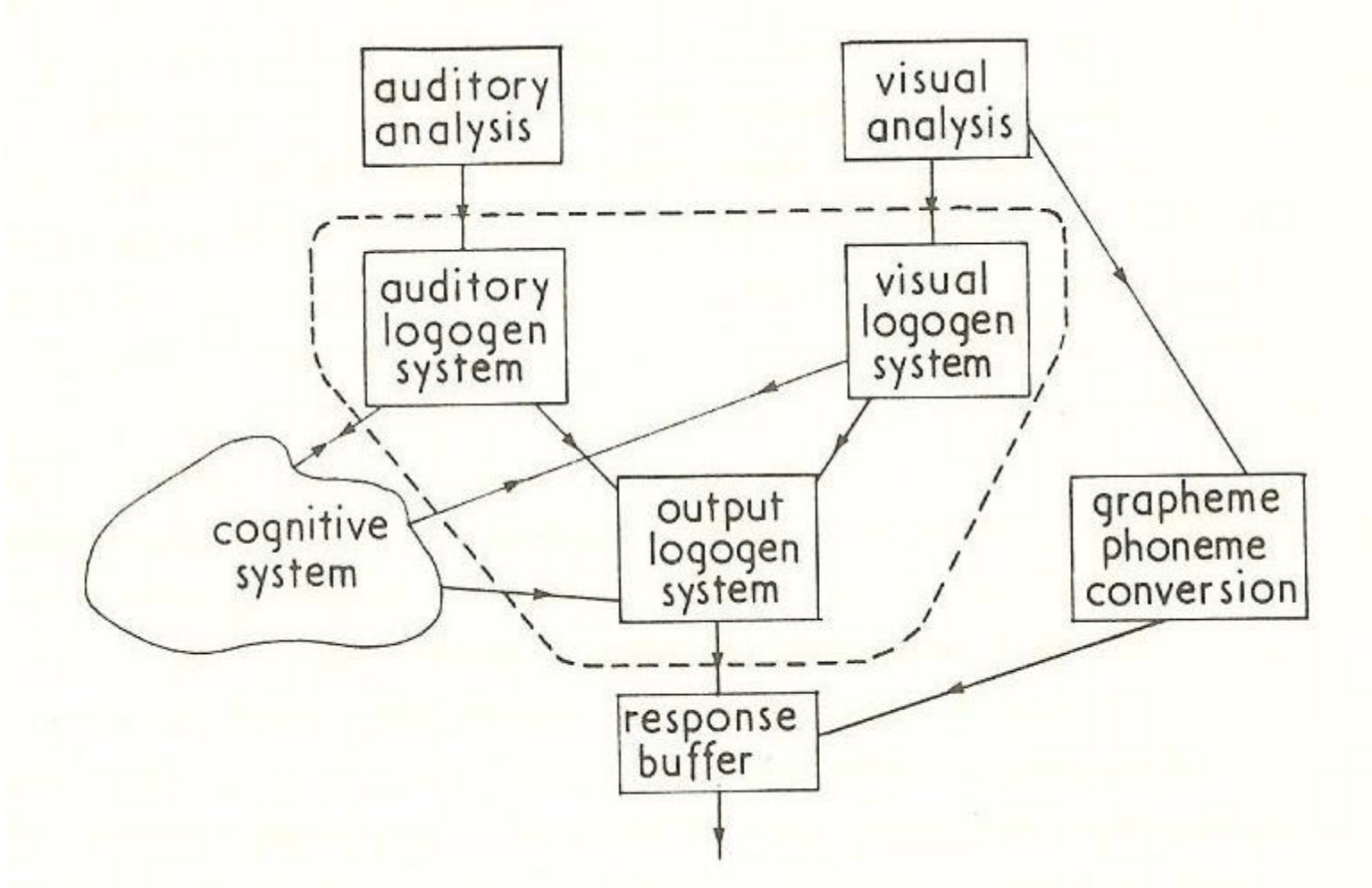
Backpropagation

- Very popular in the late 1980's and early 1990's
- Extended in many different ways
- Rejected as an approach to learning by many neurobiologists
- But was used to challenge past beliefs about what we should expect to see when we record from a neuron in the brain
  - Units in a neural network that solves a problem aren't necessarily that easily interpretable
  - Why should it be so easy to interpret neurons in the brain?



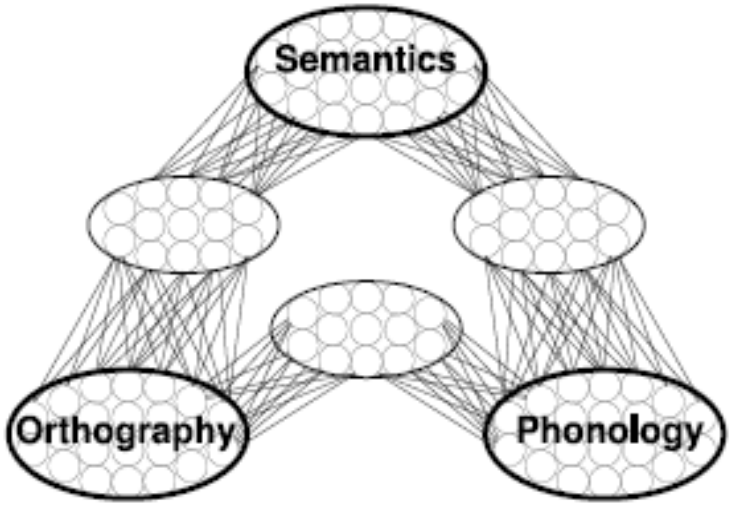
Cognitive Neuropsychology (1970's)

- Geshwind's disconnection syndromes:
  - Conduction Aphasia
    - Patient can understand and produce spoken language but cannot repeat sentences or nonwords
  - Alexia without Agraphia
- Deep and surface dyslexia (1970's):
  - Deep dyslexics can't read non-words (e.g. VINT), make semantic errors in reading words (PEACH -> 'apricot')
  - Surface dyslexics can read non-words, and regular words (e.g. MINT) but often regularize exceptions (PINT).
- Work leads to 'box-and-arrow' models, reminiscent of flow-charts

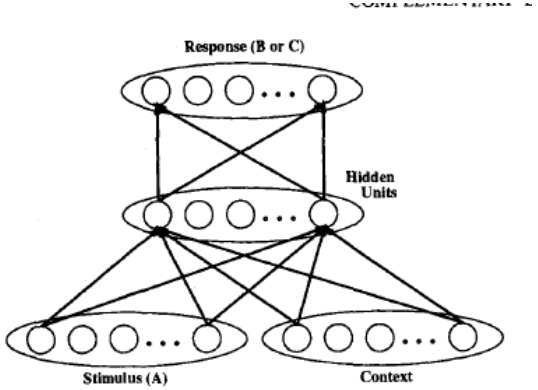

auditory analysis
visual analysis
auditory logogen system
visual logogen system
cognitive system
output logogen system
grapheme phoneme conversion
response buffer

Graceful Degradation in Neuropsychology

- Patient deficits are seldom all or none
- And error patterns are far from random:
  - Visual and semantic errors in deep dyslexia suggest degradation, rather than loss of a module or disconnection
  - Regularization errors depend on a word's frequency, and how many other exceptions there are that are like it
- Effects of lesions to units and connections in distributed connectionist models nicely capture both of these features of neuropsychological deficits.



Catastrophic interference and complementary learning systems



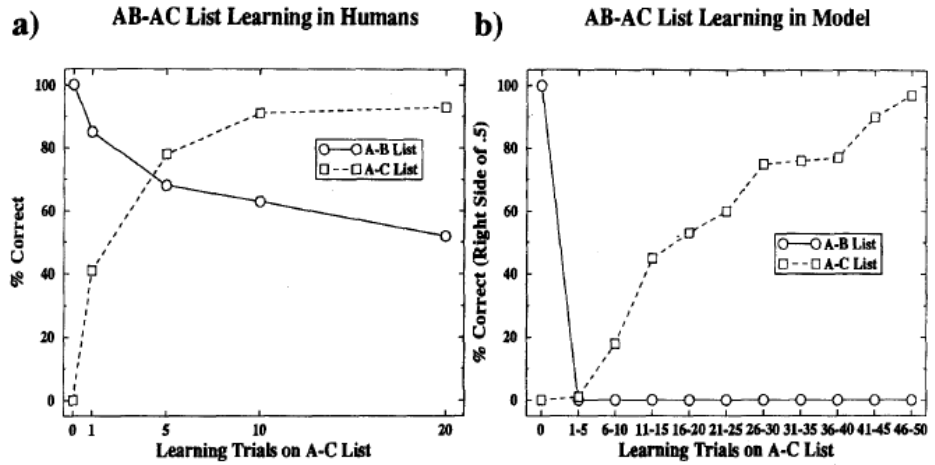
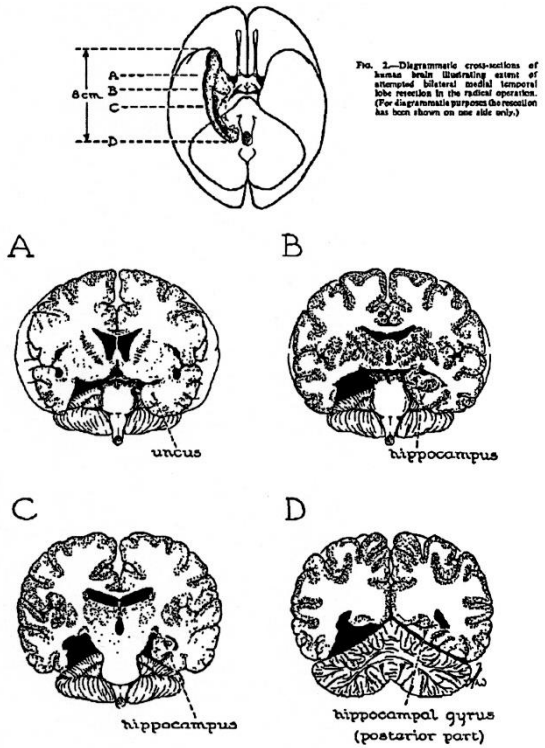
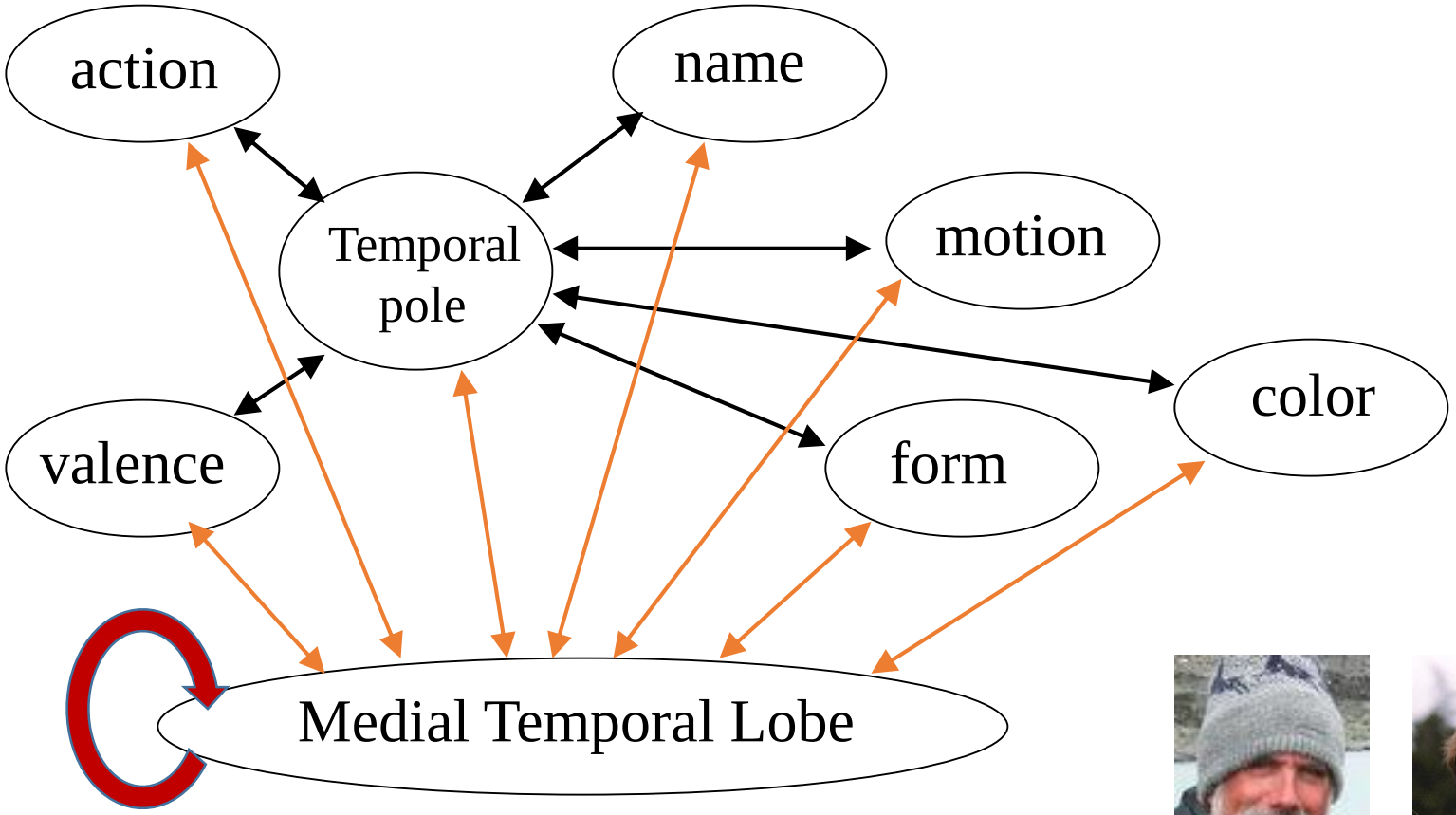


Figure 10. Panel a: Experimental data showing mild interference in humans in the *AB–AC* paradigm (Barnes & Underwood, 1959). Panel b: Simulation results demonstrating catastrophic interference. *Note.* From "Catastrophic Interference in Connectionist Networks: The Sequential Learning Problem," pp. 125 and 129, by M. McCloskey and N. J. Cohen, in *The Psychology of Learning and Motivation,* edited by G. H. Bower, 1989, New York: Academic Press. Copyright 1989 by Academic Press, Inc. Reprinted with permission.



Complementary Learning Systems Theory (McClelland, McNaughton & O'Reilly, 1995)



A second neural network winter?

- For AI, early promise of neural networks didn’t seem to be paying off by the late 1990’s
- It was widely believed that depth was a curse, not a blessing
  - Vanishing gradient problem
- Explicit Bayesian approaches became popular
- Stanford CS stopped teaching back-propagation

What changed?

- Another factor of 1,000 in computer size and speed
- Huge data sets
- Effective utilization of underappreciated innovations
  - LSTMs
- A new generation of brilliant minds, inspired by the promising results of others

Reflections

- People always seem to think that the limitations of what we have today are principled and insurmountable
- It is always rash predict what will happen in the future…

- But perhaps it is even more rash to try to rule anything out!