Midterm check-in

- How’d the exam go?
- When will it be graded?

Recap

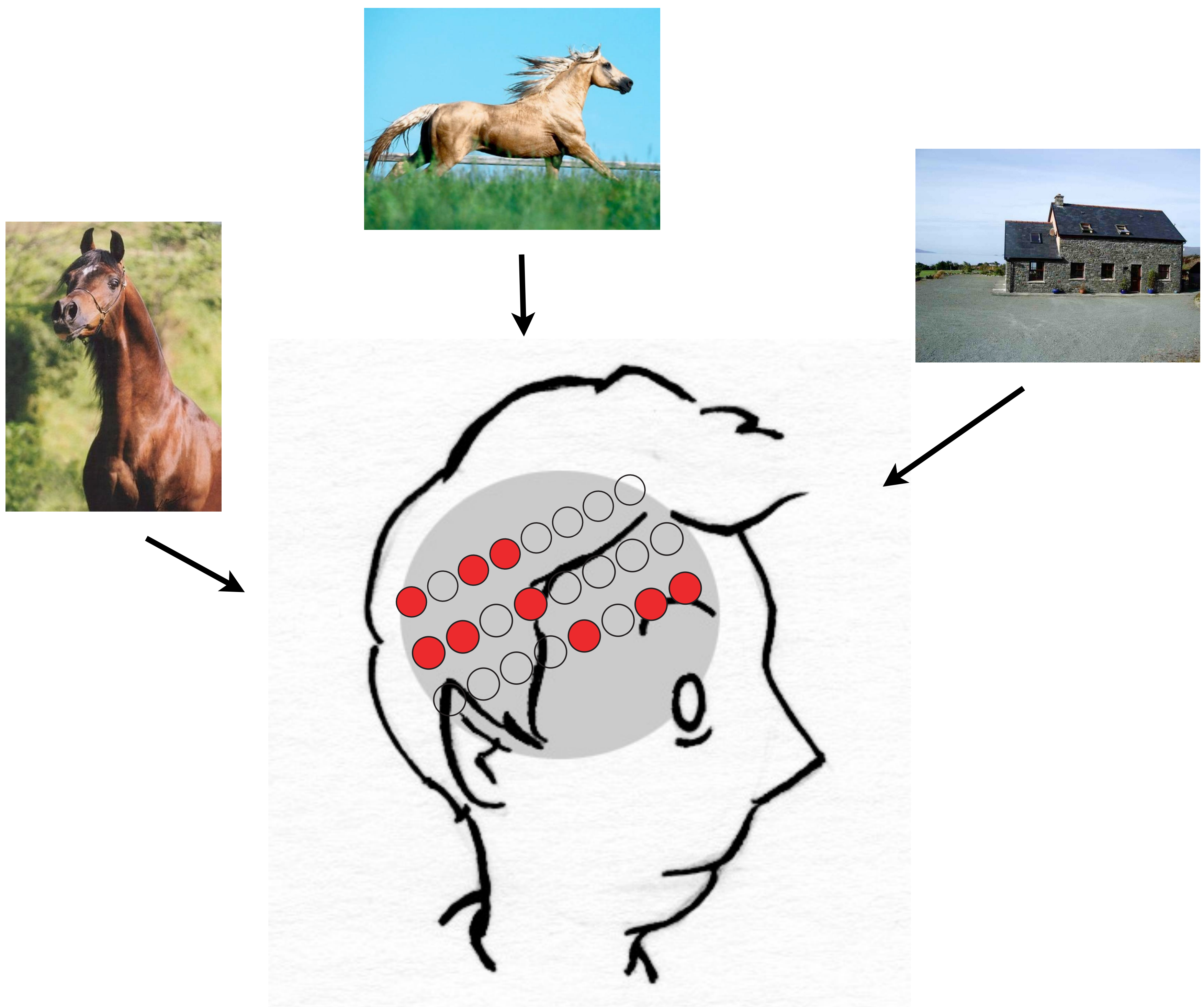
- Retrieval induced forgetting
- Attribute similarity models of recall

Models of association

PSYC 51.09: Human Memory
Spring 2021

Jeremy Manning
jeremy@dartmouth.edu

Multitrace Similarity Attribute Theory



From file cabinets to neural networks

- Our multiple-trace memory model helps us understand human behavior in a number of different situations/experiments
- But there are some important things that are left out by this model

Limitations

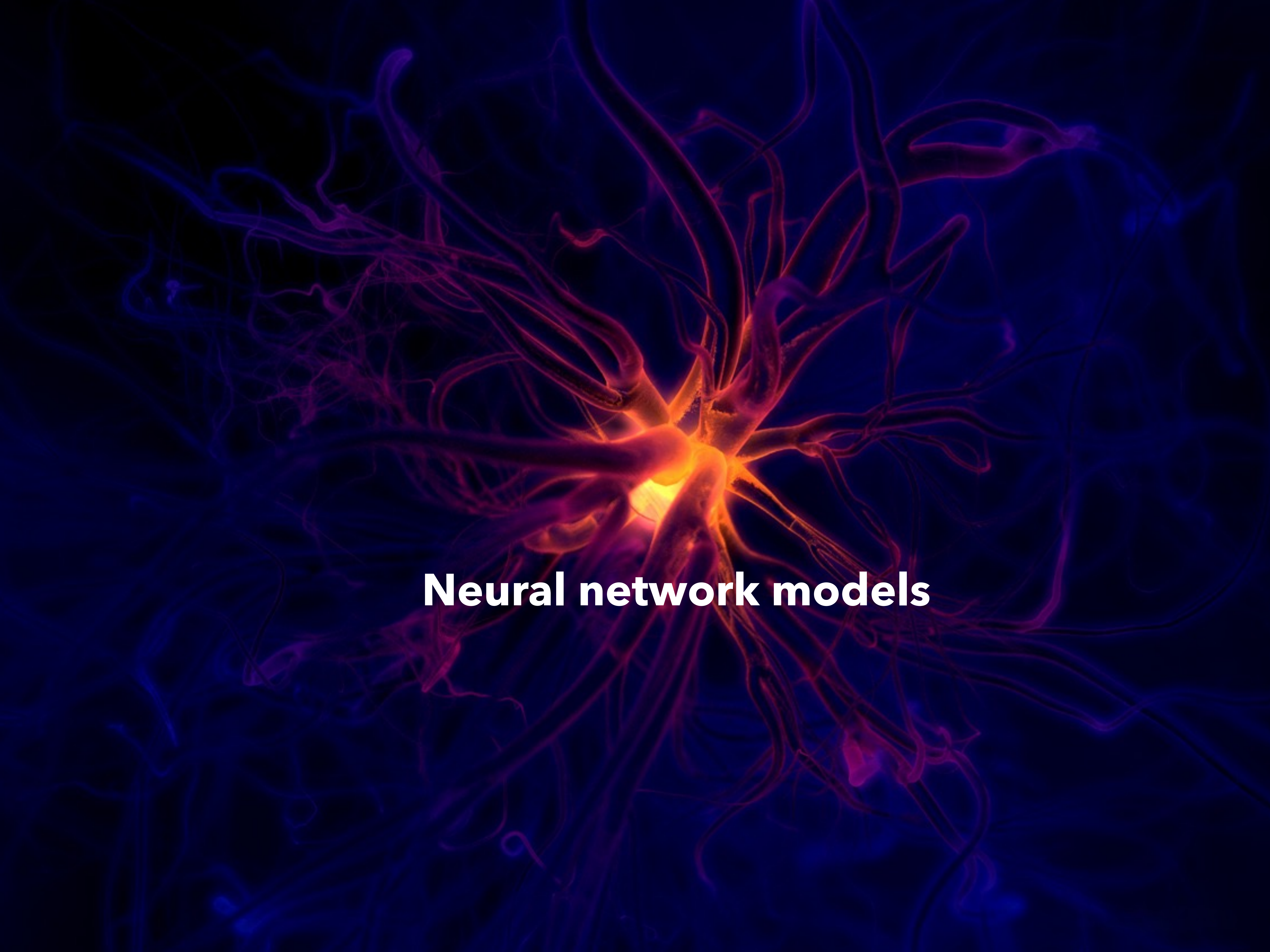
- We say that a given item has a particular representation (a set of attributes)…
- …but how does that representation form?
- …and what holds it together?

Limitations

- Each new experience causes a new memory trace to be formed…is this reasonable?
  - Can an old memory trace be altered, adjusted, damaged?
  - Can storing a new memory damage an old memory?
  - Do we ever run out of storage space?

Limitations

- Search Problem
  - The probe is compared to every memory in the system, but we haven't said how this happens
  - In cued recall, we reactivate the memory that best matches the cue, but we haven't said how this happens, either!



Neural network models

Neural network models

- A representation is an activity pattern across a network of neurons
- These neurons are all connected to one another by synapses, which specify how strongly one neuron influences the activity of any other neuron
- Learning involves adjusting the connection strengths between neurons

Neural network models

- representation = (feature) vector = pattern = state
- element = feature = neuron = node
- synapse = weight = connection

Neural network models

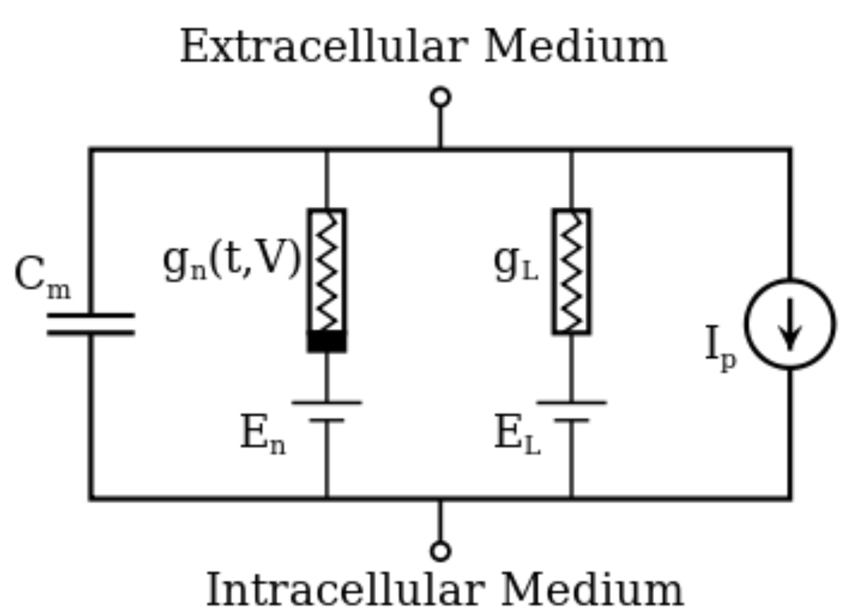
- With just a handful of rules, we can create a memory system that can:
  - explain what it means for a representation to be stable
  - explain how storing a new memory can damage other memories
  - estimate storage capacity and reaction time
  - …and more!

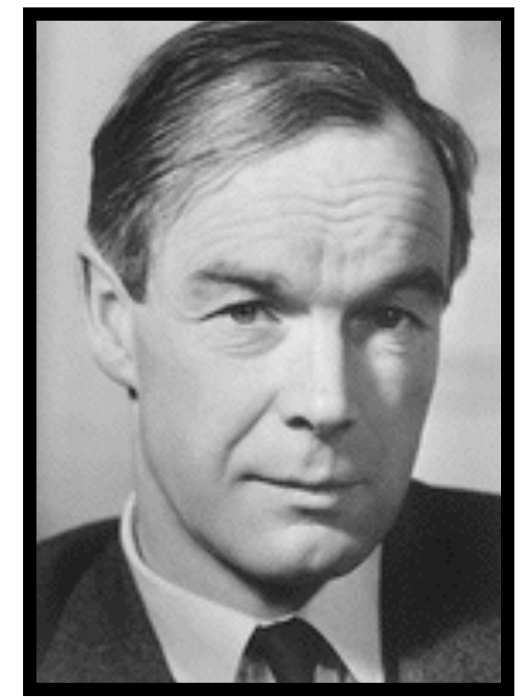
Road map

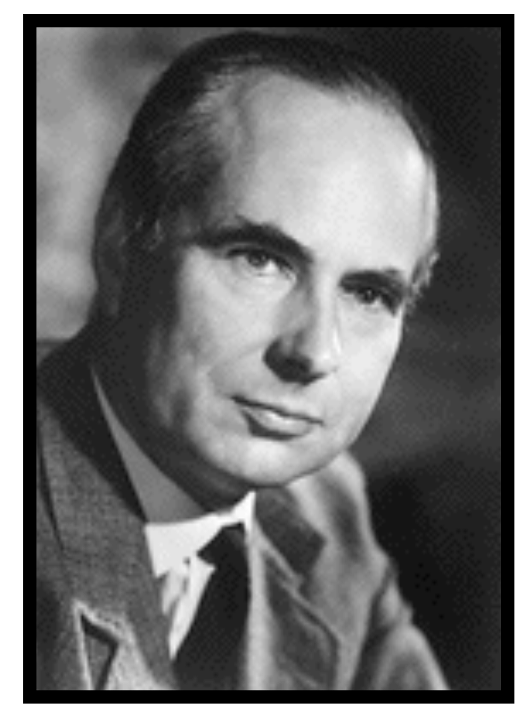
- The basics of neural network models
- Hopfield model and pattern completion
- Neural dynamics in the human brain

Biophysical models

- Hodgkin & Huxley (1952) developed a mathematical model describing how shifts in ionic currents alter the electrical potential of the cell, giving rise to an action potential.
- Nobel prize for this work (1963)
- It is possible to make very detailed models of the biophysics of individual neurons

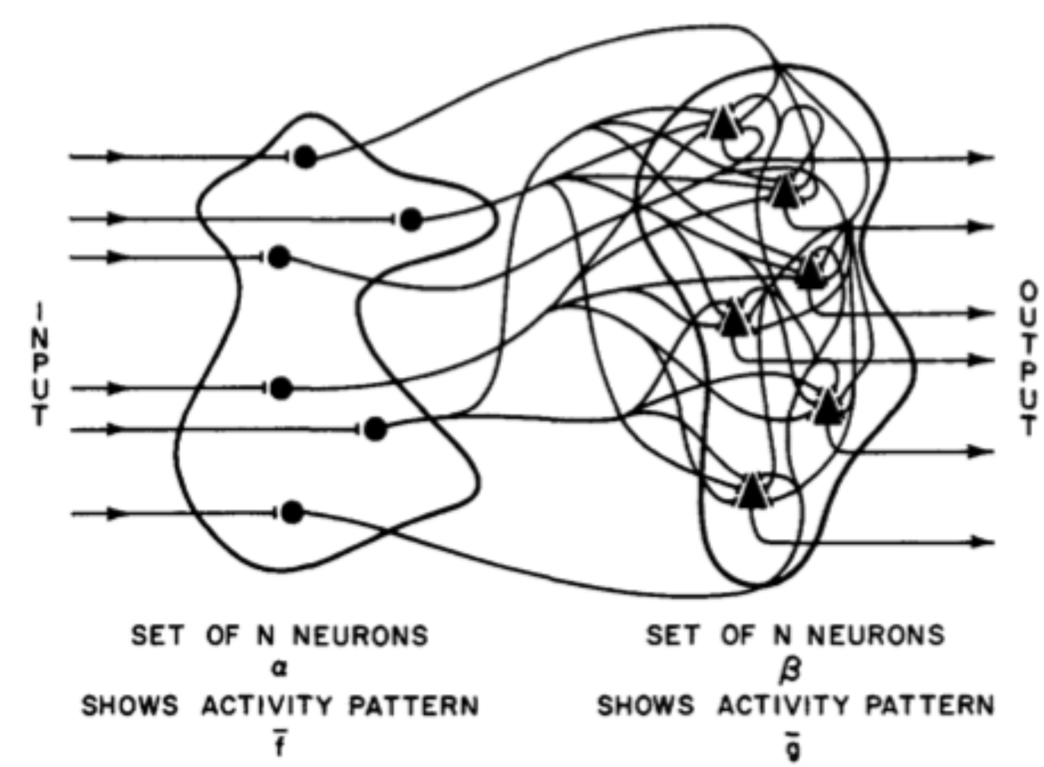






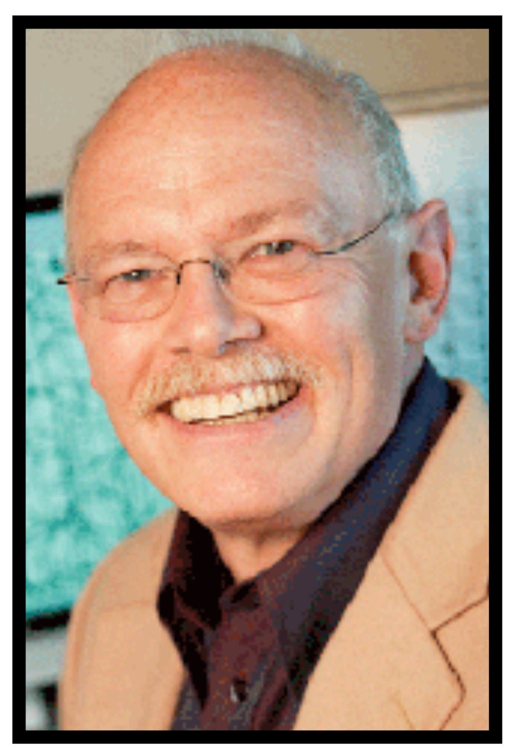
Neuro-cognitive models

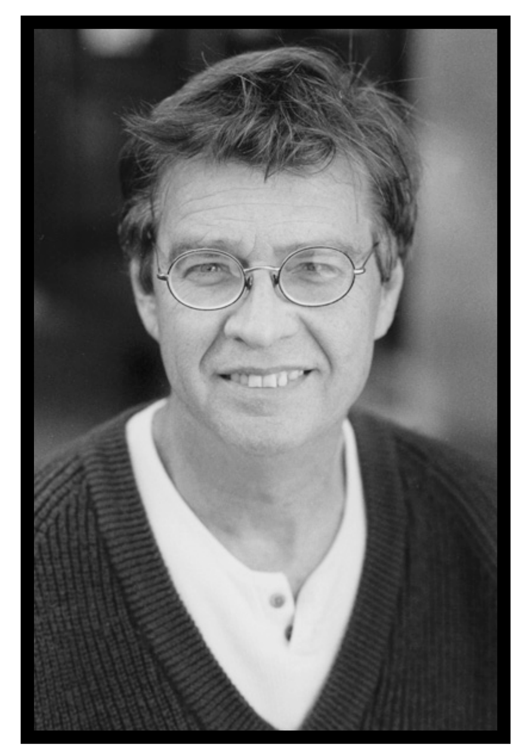
- It is also possible to work with very abstract versions of neurons, which lets us focus on the computational properties of the broader system (how lots of neurons interact)
- **Linear associators:** Leon Cooper (1973, Nobel Lecture). Major development by Jim Anderson, in Estes Lab.



Neuro-cognitive models

- **Other pioneers:** Grossberg, Rumelhart, McClelland, Hinton, Sejnowski, and Kohonen
- These researchers have created a computational foundation for cognitive neuroscientific theory, establishing the possible mechanisms used by the brain to perceive, attend, learn, and act!
- We will return to these models in Chapter 7…





Link to statistical physics

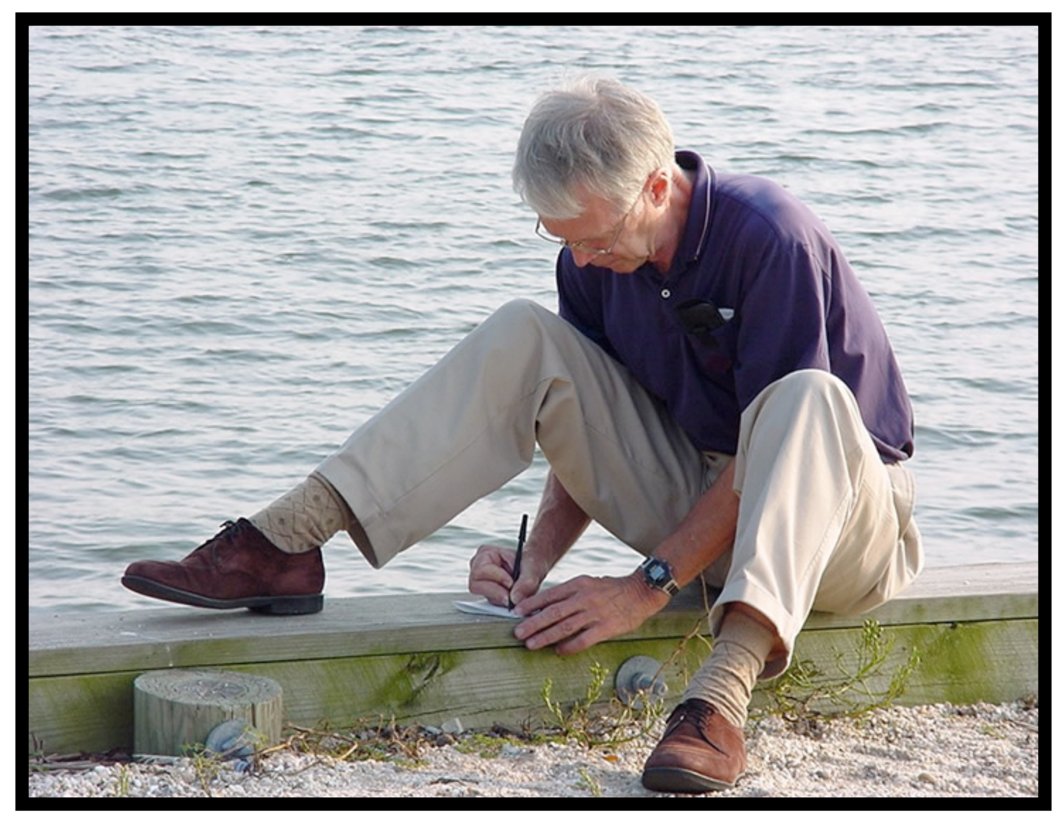
- **Hopfield networks:** In 1982, John Hopfield developed the link between physical models of magnetic systems and biophysical models of neural networks.

This framework was very attractive to mathematicians, because it was possible to develop formal proofs regarding network dynamics

Link to statistical physics

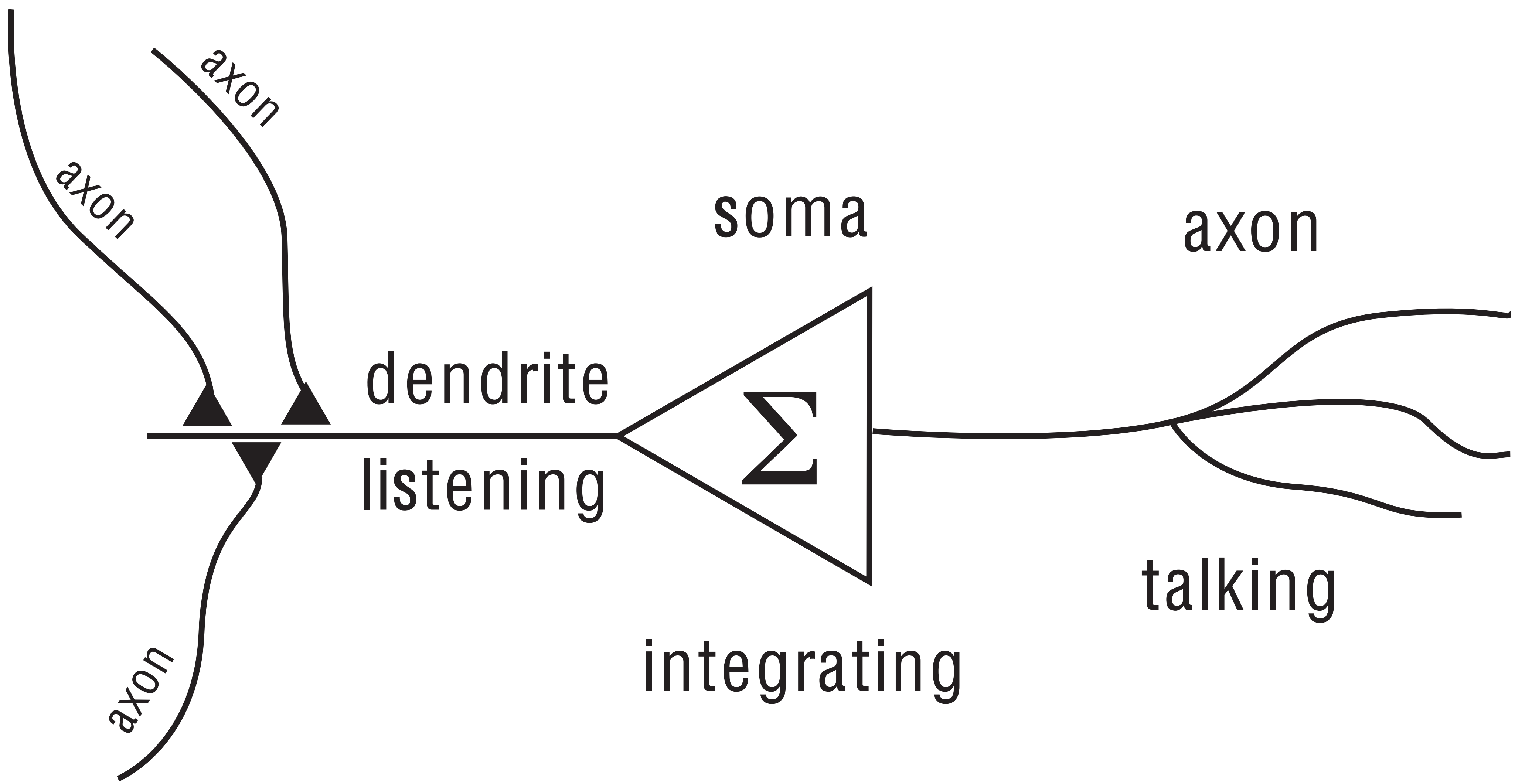
- **Hopfield networks:** In 1982, John Hopfield developed the link between physical models of magnetic systems and biophysical models of neural networks.

Also, it is a super-simplified version of a neural network, which allows us to work with it without a computer, and gain some intuitions about how these models work!

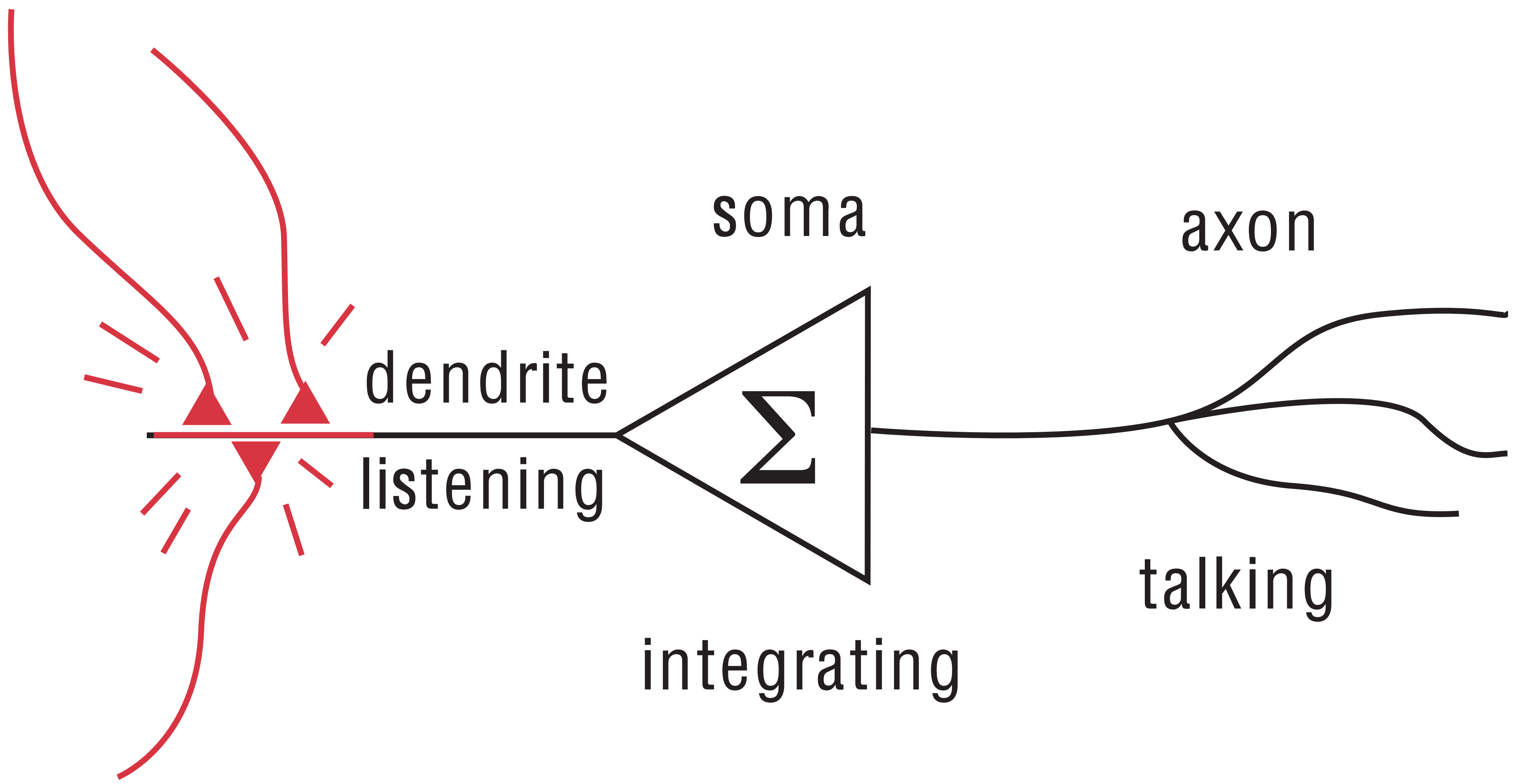


Boiling neurons down to the basics

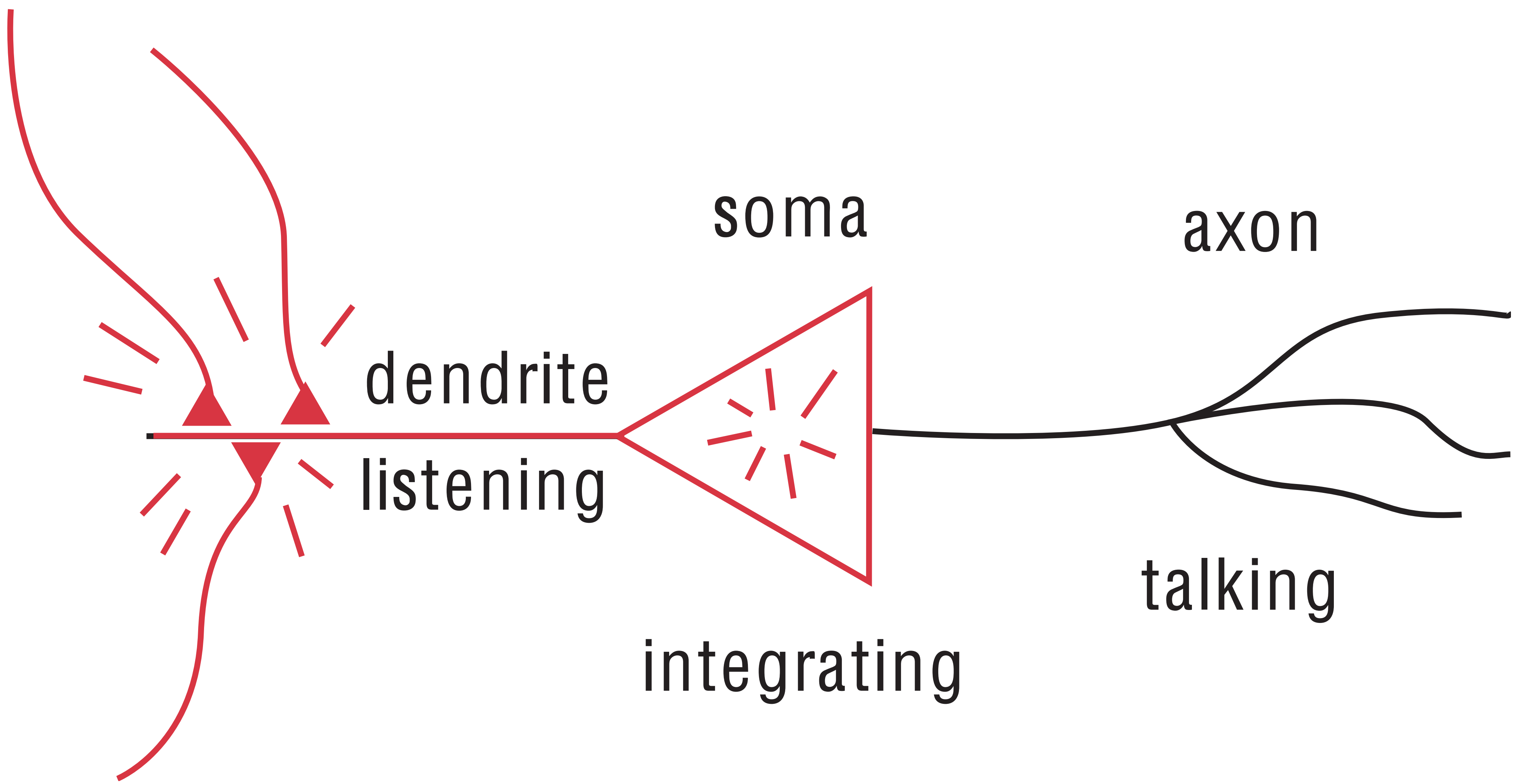
A schematic neuron



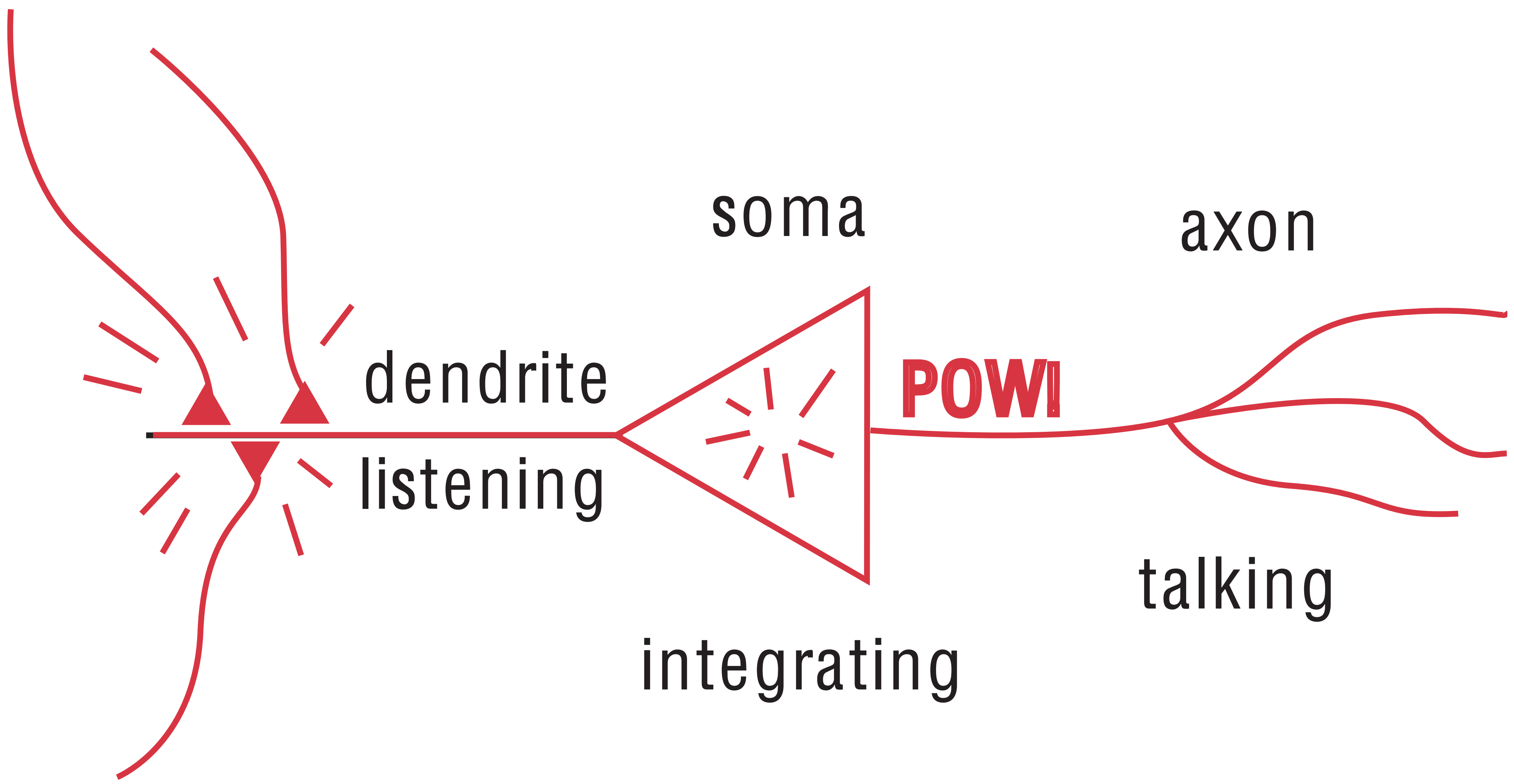
A schematic neuron



A schematic neuron



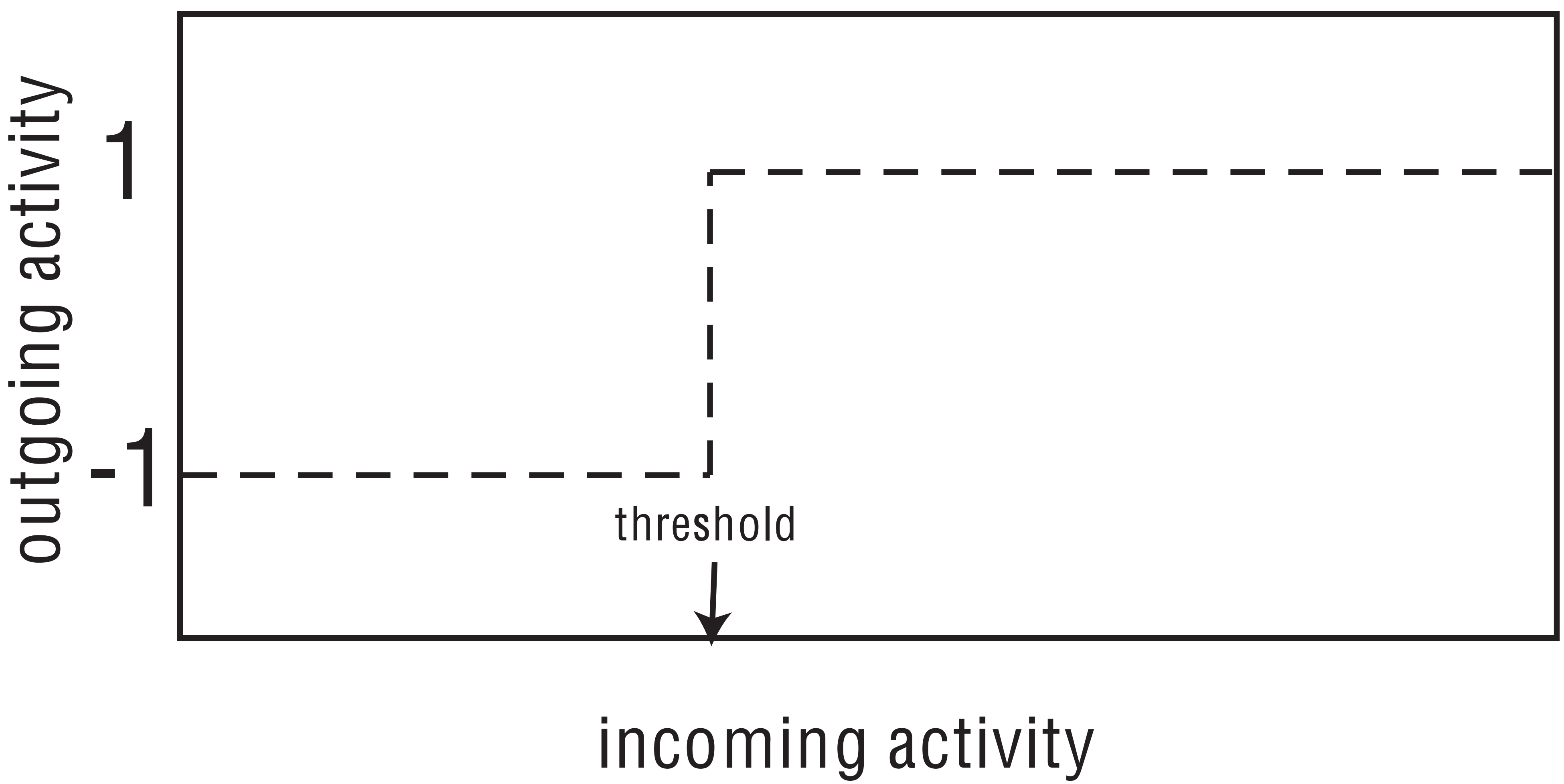
A schematic neuron



How do neurons talk?

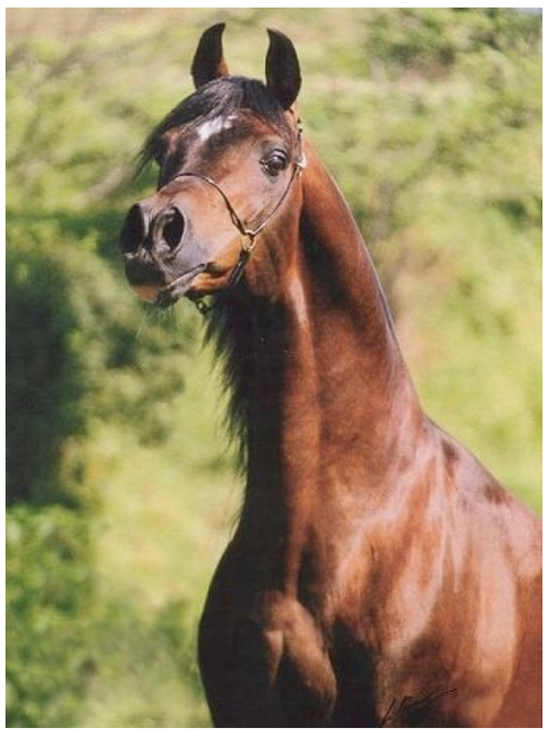
- In real neural networks, there are many different kinds of neurons
- Some are excitatory (they activate other neurons) and some are inhibitory (they turn off or suppress other neurons)
- In the Hopfield model, there is only one kind of neuron, but it can send out both excitatory and inhibitory signals

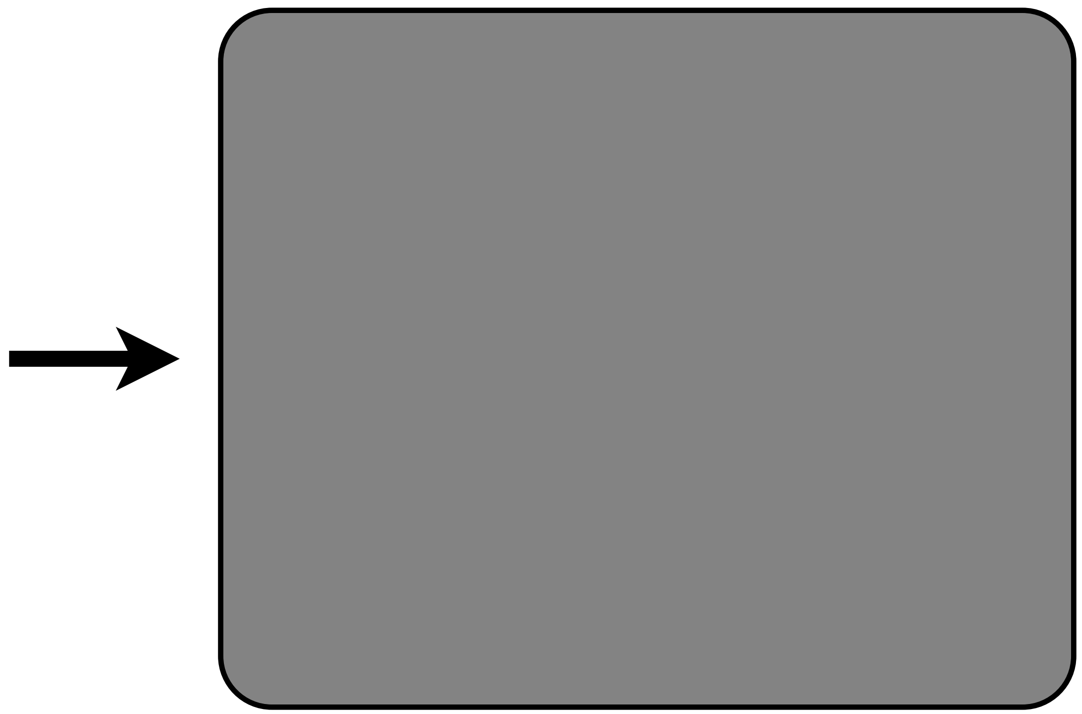
Activation Function

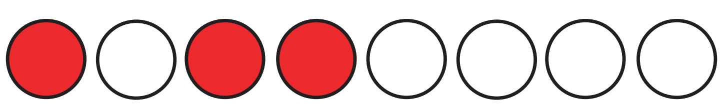


How do neurons learn?

Creating a stable representation

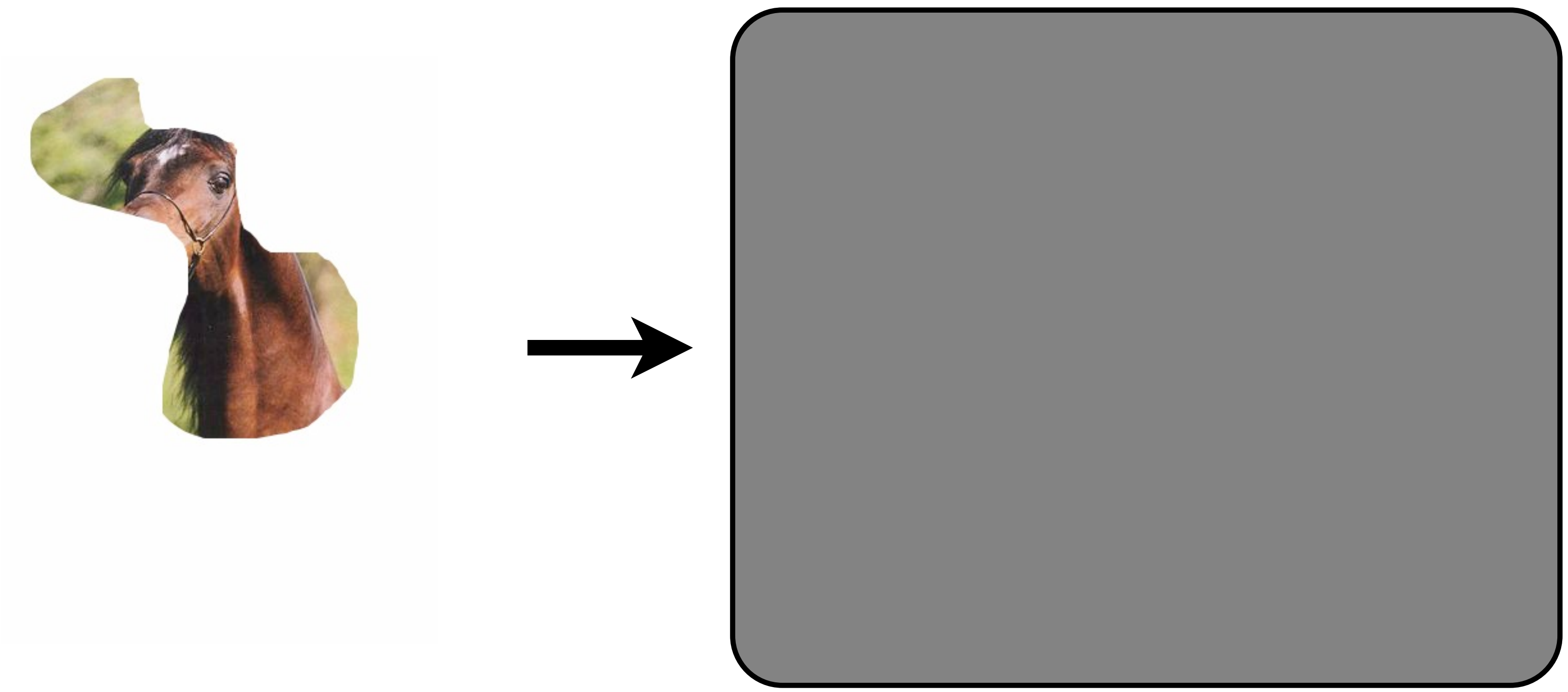






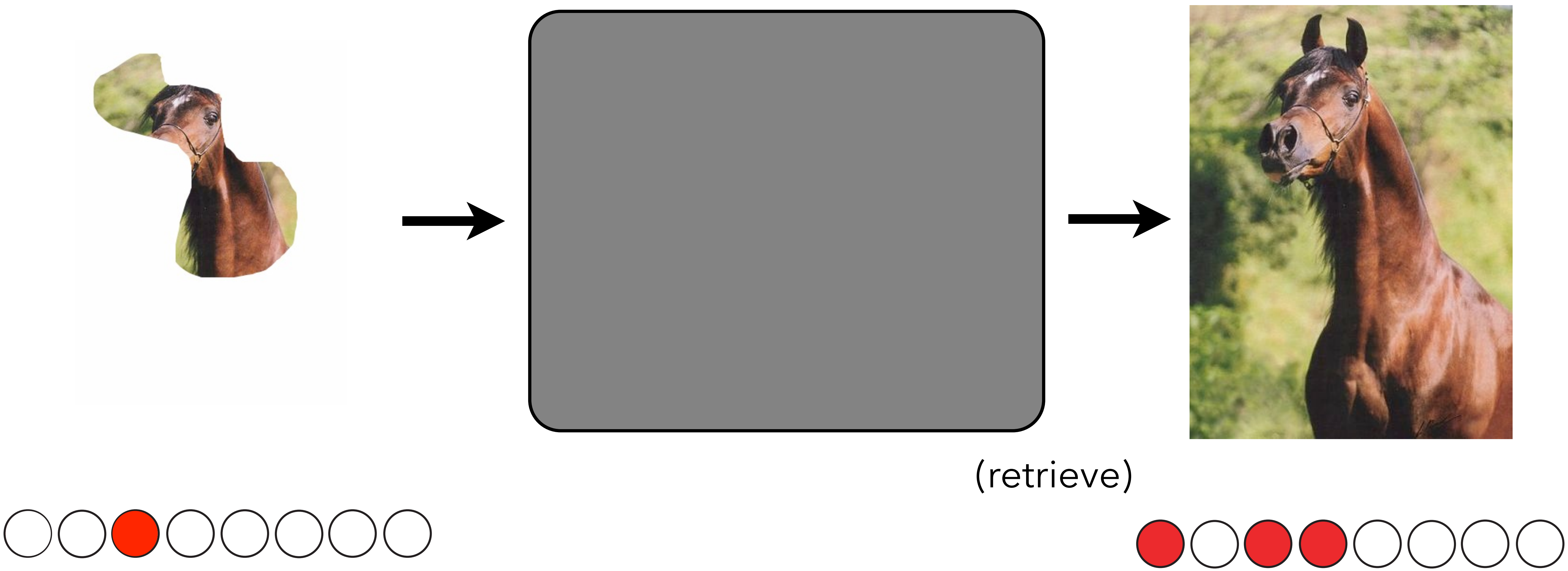
Learning

Probing memory

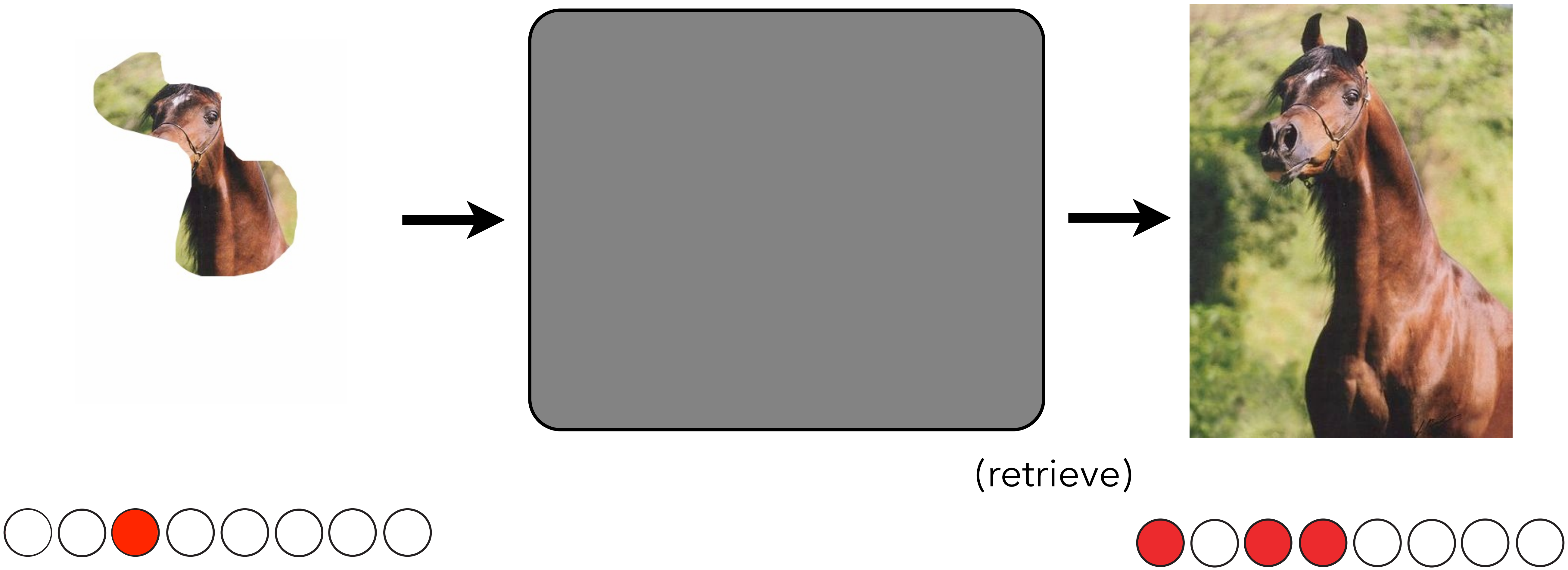


(retrieve)

Pattern completion



Pattern completion



To create a neural network that can do pattern completion, we need to come up with rules that determine (1) whether a given neuron should be active, and (2) how the connections change strength during learning.

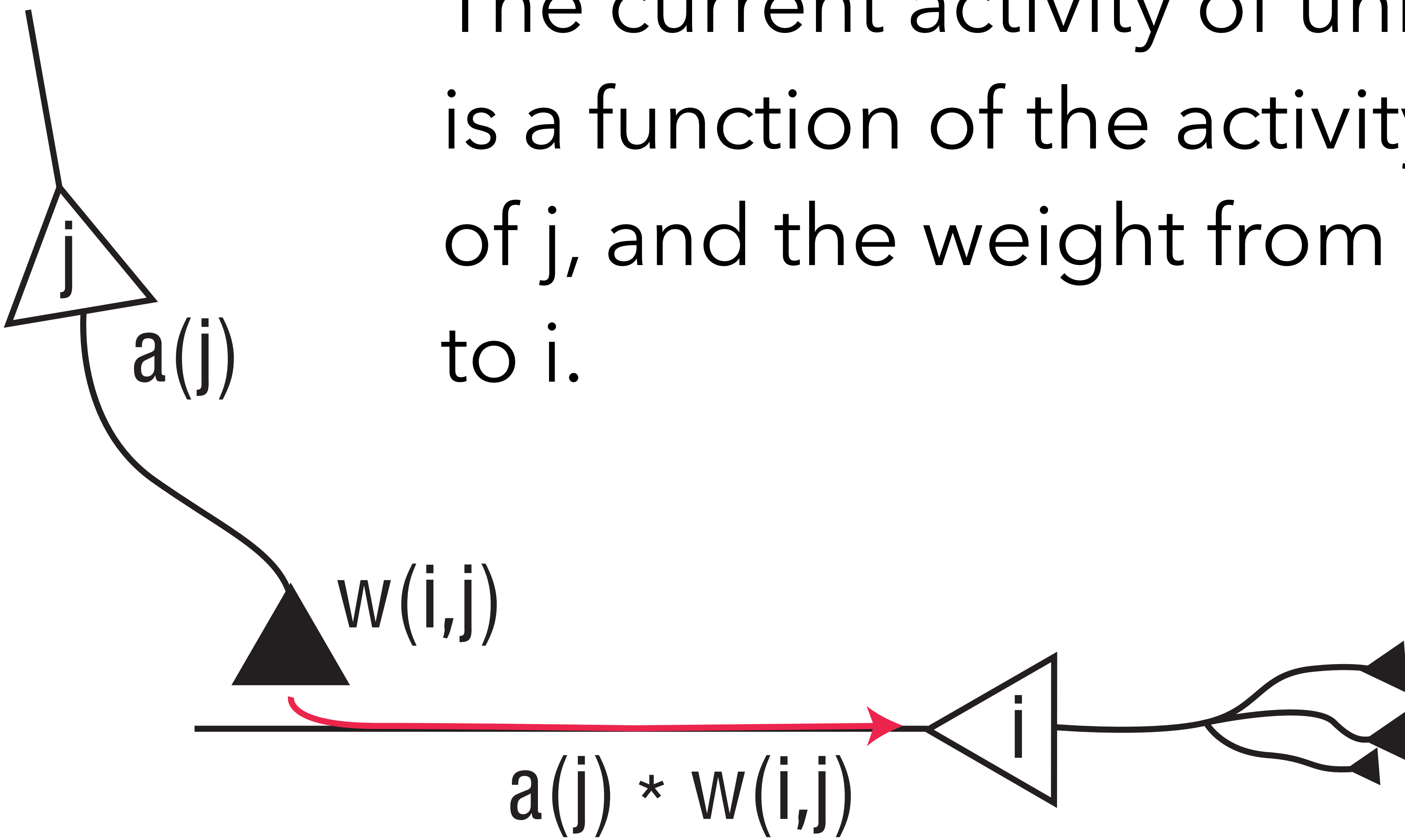
Some terminology

- We can think of the neural network as being represented by a vector of attribute values
- If there are 5 elements in the vector, there are 5 neurons in our network
- At a given moment (time *t)*, only one pattern can be active
- The currently active pattern is the active state of the network

Some terminology

- We can specify the connection strength between all pairs of neurons; these connections are meant to represent synapses
- Multiple memories can be stored in these connections
- Learning = modification of connection weights
- Nodes (neurons) interact with each other to recall memories (carrying out pattern completion)

The current activity of unit i is a function of the activity of j, and the weight from j to i.



Learning rule

- Tells us how to encode memories in the network

$$w(i, j) = \sum_{k=1}^{L} a_k(i) a_k(j)$$

Learning multiple memories (example)

$$w(i,j) = \sum_{k=1}^{L} a(i)a(j)$$

$$\mathbf{a}_1 = \begin{pmatrix} +1 \\ -1 \\ -1 \\ +1 \\ +1 \end{pmatrix} \quad \mathbf{a}_2 = \begin{pmatrix} -1 \\ +1 \\ -1 \\ -1 \\ +1 \end{pmatrix} \quad \mathbf{a}_3 = \begin{pmatrix} +1 \\ +1 \\ +1 \\ -1 \\ -1 \end{pmatrix}$$