Systems and Cognitive Neuroscience

NEU/PSY/MOL 502A

Professor: **Jonathan D. Cohen** *(jdc@princeton.edu)*

AI: Alex Ku *(alexku@princeton.edu)*

Systems and Cognitive Neuroscience

NEU/PSY/MOL 502A

Professor: Jonathan D. Cohen *(jdc@princeton.edu)*

AI: Alex Ku *(alexku@princeton.edu)*

Course Description:

A survey of experimental & theoretical approaches to understanding how cognition arises in the brain. This complements 501, focusing on the mechanisms responsible for perception, attention, decision making, memory, cognitive & motor control, and planning, with emphasis on the representations involved & their transformations in the service of cognitive function. Source material will span neuroscience, cognitive science, and work on artificial systems. Relevance to neurodegenerative and neuropsychiatric disorders will also be discussed.

Computational constructs will explored through "hands-on" modeling exercises carried in parallel in 502B

Systems and Cognitive Neuroscience

NEU/PSY/MOL 502A

Systems and Cognitive Neuroscience

NEU/PSY/MOL 502A

- **Lectures**
  - **Mondays and Thursdays, 2-4:30**
  - **Divided into 9 sections, each that will address a set of: cognitive phenomena/processes and computational/neural mechanisms:**

    1) Sensation And Perception — Inference And Constraint Satisfaction
    2) Decision Making — Integration
    3) Reinforcement Learning — Reward And Neuromodulation
    4) Semantic Memory — Statistical Learning And Distributed Representation
    5) Episodic Memory — Binding
    6) Attention, Working Memory And Cognitive Control — State Modulation
    7) Motor Function — Movement
    8) Development And Social Cognition — Interaction
    9) Disorders — Dysfunction

Systems and Cognitive Neuroscience

NEU/PSY/MOL 502A

- **Lectures**
  - Mondays and Thursdays, 2-4:30
  - Divided into 9 sections, each that will address a set of: **cognitive phenomena/processes** and **computational/neural mechanisms**:
    1) Sensation And Perception — Inference And Constraint Satisfaction
    2) Decision Making — Integration
    3) Reinforcement Learning — Reward And Neuromodulation
    4) Semantic Memory — Statistical Learning And Distributed Representation
    5) Episodic Memory — Binding
    6) Attention, Working Memory And Cognitive Control — State Modulation
    7) Motor Function — Movement
    8) Development And Social Cognition — Interaction
    9) Disorders — Dysfunction
  - Schema:
    - Monday: 1st half: overview lecture; 2nd half: deep dive — faculty guest lecture
    - Thursday: 1st half: overview lecture; 2nd half: student presentation

Systems and Cognitive Neuroscience

NEU/PSY/MOL 502A

Systems and Cognitive Neuroscience

NEU/PSY/MOL 502A

- **Readings**
  - **All are source materials; no official text** *(though see below)* **— that means class matters!**
  - **Lots are listed, all are available as PDFs**
  - **Asterisked readings are required;**
    others are meant primarily as a resource, to explore material covered in class in greater depth
  - **In addition, some good reference texts are:**
    - *Parallel distributed processing: Explorations in the microstructure of cognition*
      Rumelhart, Hinton & McClelland (1986)
    - *Computational explorations in cognitive neuroscience: Understanding the mind by simulating the brain*
      O'Reilly & Munakata (2000)
    - *Theoretical Neuroscience*
      Dayan and Abbott (2001)

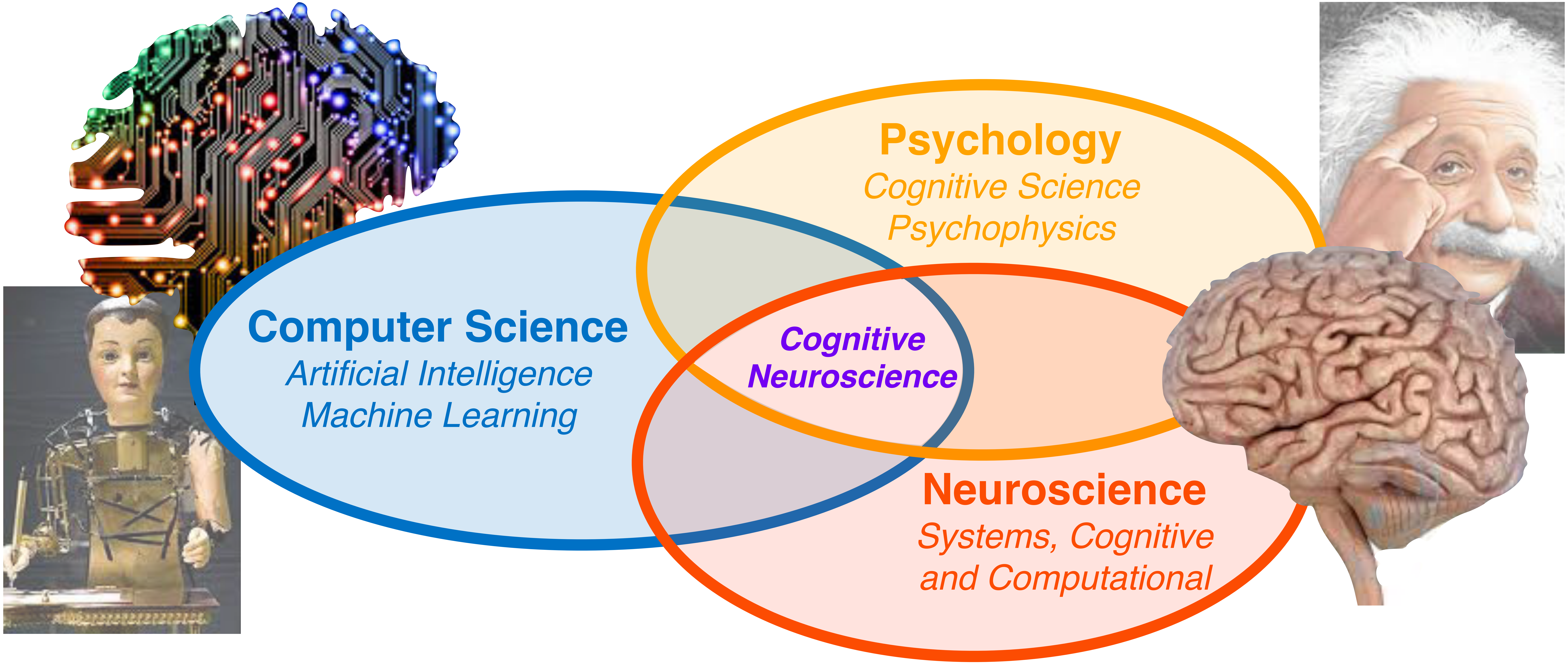
Systems and Cognitive Neuroscience

NEU/PSY/MOL 502A

- **Readings**
  - **All are source materials; no official text** *(though see below)* **— that means class matters!**
  - **Lots are listed, all are available as PDFs**
  - **Asterisked readings are required;**
    others are meant primarily as a resource, to explore material covered in class in greater depth
  - **In addition, some good reference texts are:**
    - *Parallel distributed processing: Explorations in the microstructure of cognition*
      Rumelhart, Hinton & McClelland (1986)
    - *Computational explorations in cognitive neuroscience: Understanding the mind by simulating the brain*
      O'Reilly & Munakata (2000)
    - *Theoretical Neuroscience*
      Dayan and Abbott (2001)
- **Course requirements and grading**
  - **Attend and participate in class (50%)**
  - **Paper presentation (50%)**

Artificial Intelligence

Natural Intelligence



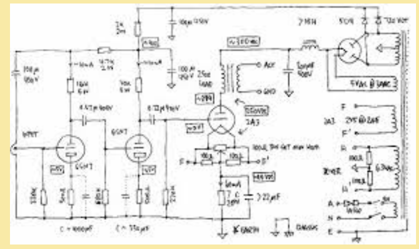
Long History of Interaction

Long History of Interaction

Neuroscience / Psychology

1940

Theories of Computation
von Neumann, Turing



Mathematics / Computer Science

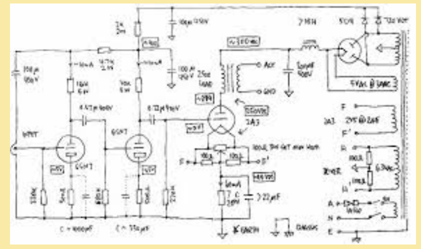
Long History of Interaction

Neuroscience / Psychology

Neural network computation
McCulloch, Hebb

1940 — *1950*

Theories of Computation
von Neumann, Turing

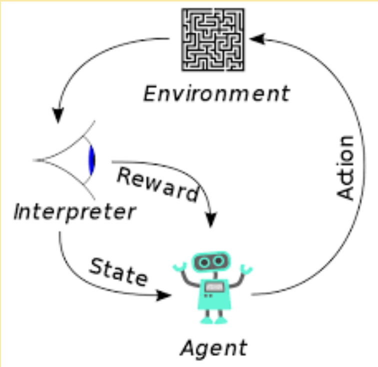


Mathematics / Computer Science

Long History of Interaction

Neuroscience / Psychology

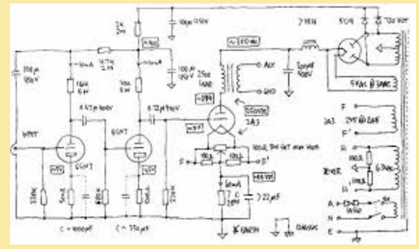
Neural network computation
McCulloch, Hebb



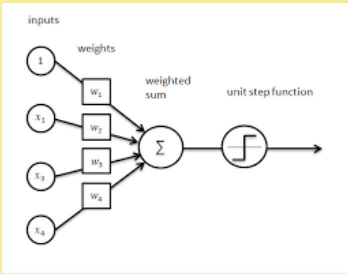
Reinforcement Learning
Recorla-Wagner

1940 — *1950* — *1960* — *1970*

Theories of Computation
von Neumann, Turing



Perceptron
Rosenblatt

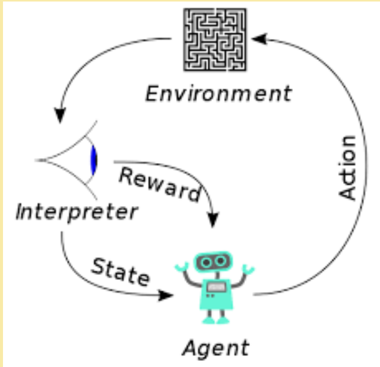


Mathematics / Computer Science

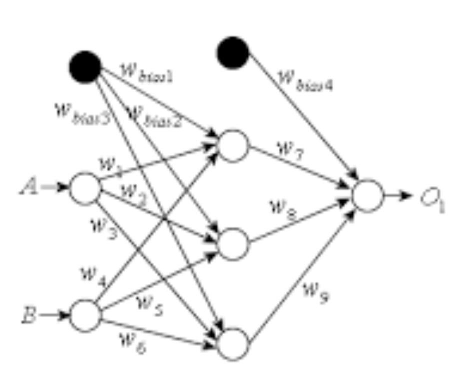
Long History of Interaction

Neuroscience / Psychology

Neural network computation
McCulloch, Hebb



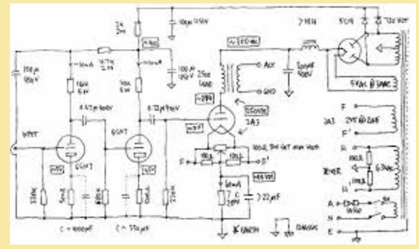
Reinforcement Learning
Recorla-Wagner



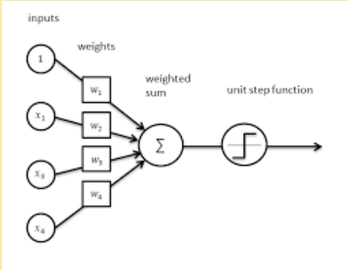
Backpropagation Learning Algorithm
Rumelhart, Hinton

1940 — 1950 — 1960 — 1970 — 1980

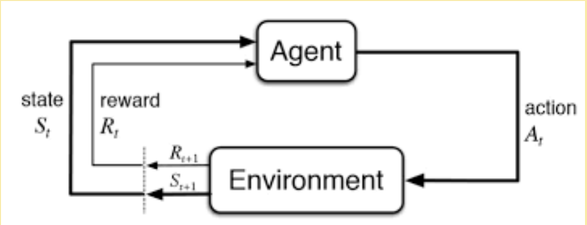
Theories of Computation
von Neumann, Turing



Perceptron
Rosenblatt

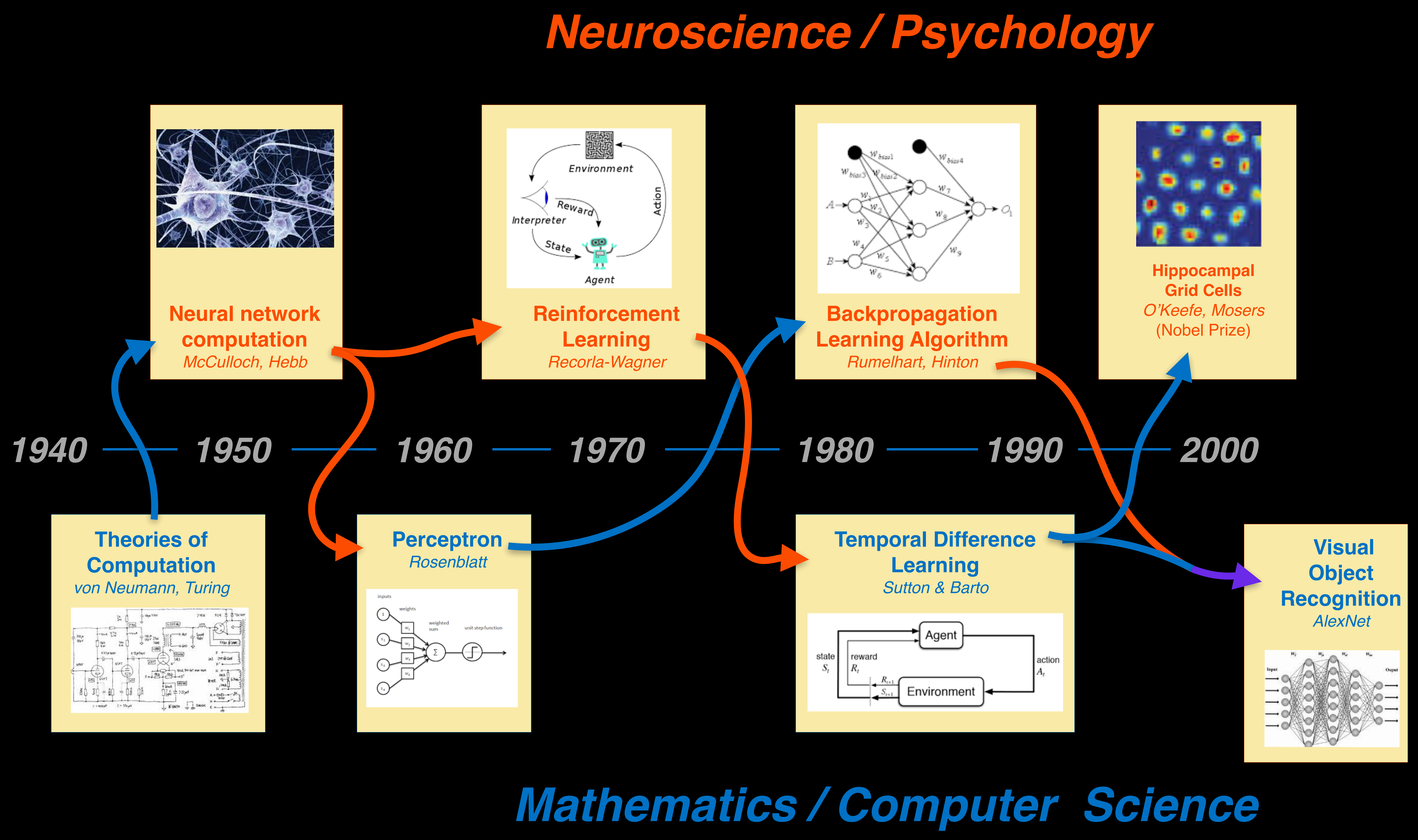


Temporal Difference Learning
Sutton & Barto

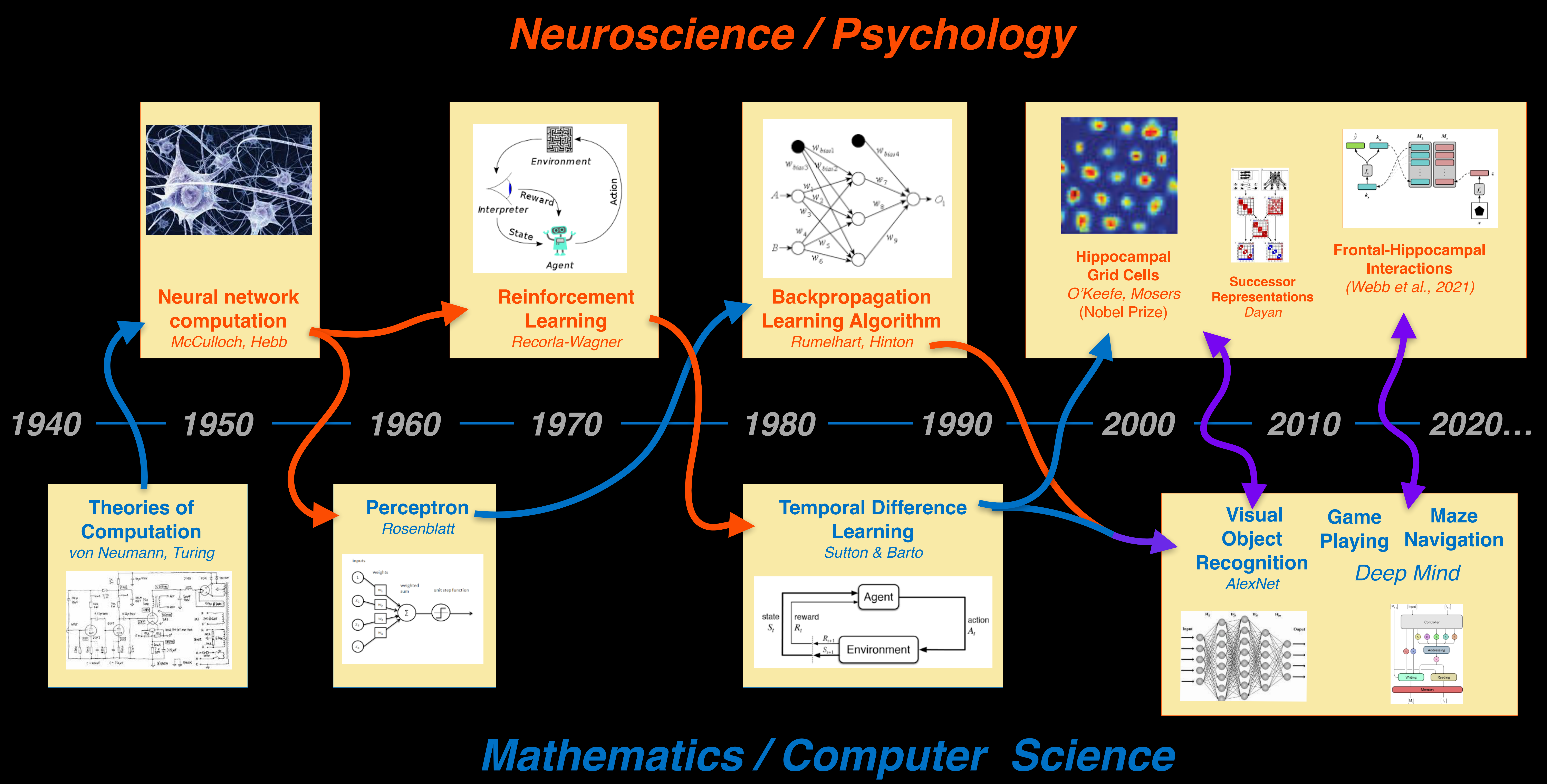


Mathematics / Computer Science

Long History of Interaction



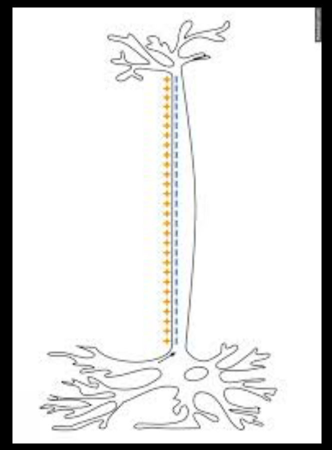
Long History of Interaction



Brief Historical Review

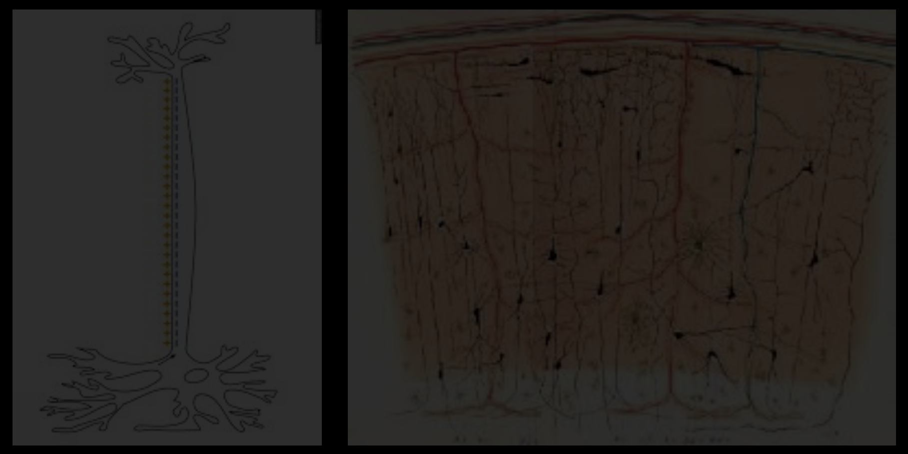
Brief Historical Review

- Spiking Neurons and Neural Assemblies
  - McCulloch & Pitts (neuroscience)
  - Hebb (psychology)
  - Norbert Weiner (mathematics)



Brief Historical Review

- Spiking Neurons and Neural Assemblies
  - McCulloch & Pitts (neuroscience)
  - Hebb (psychology)
  - Norbert Weiner (mathematics)



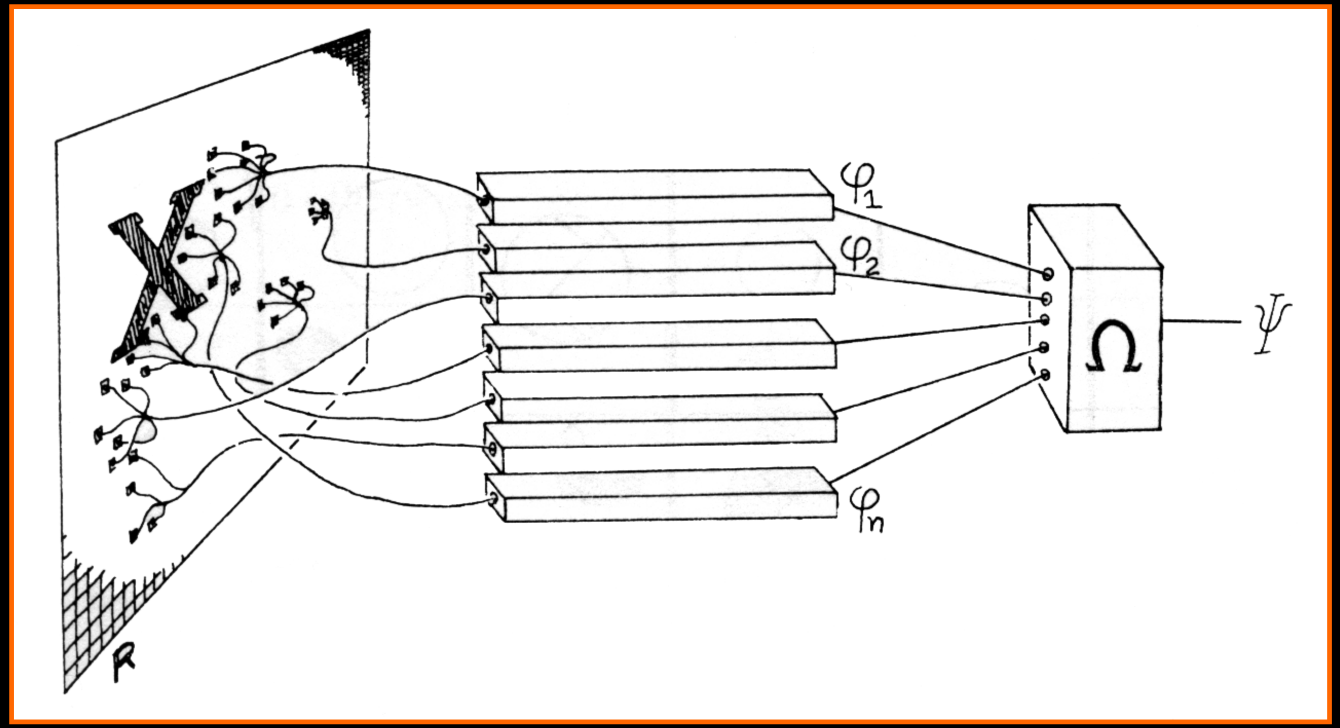
- Early neural network models…

The Perceptron

Frank Rosenblatt, 1957

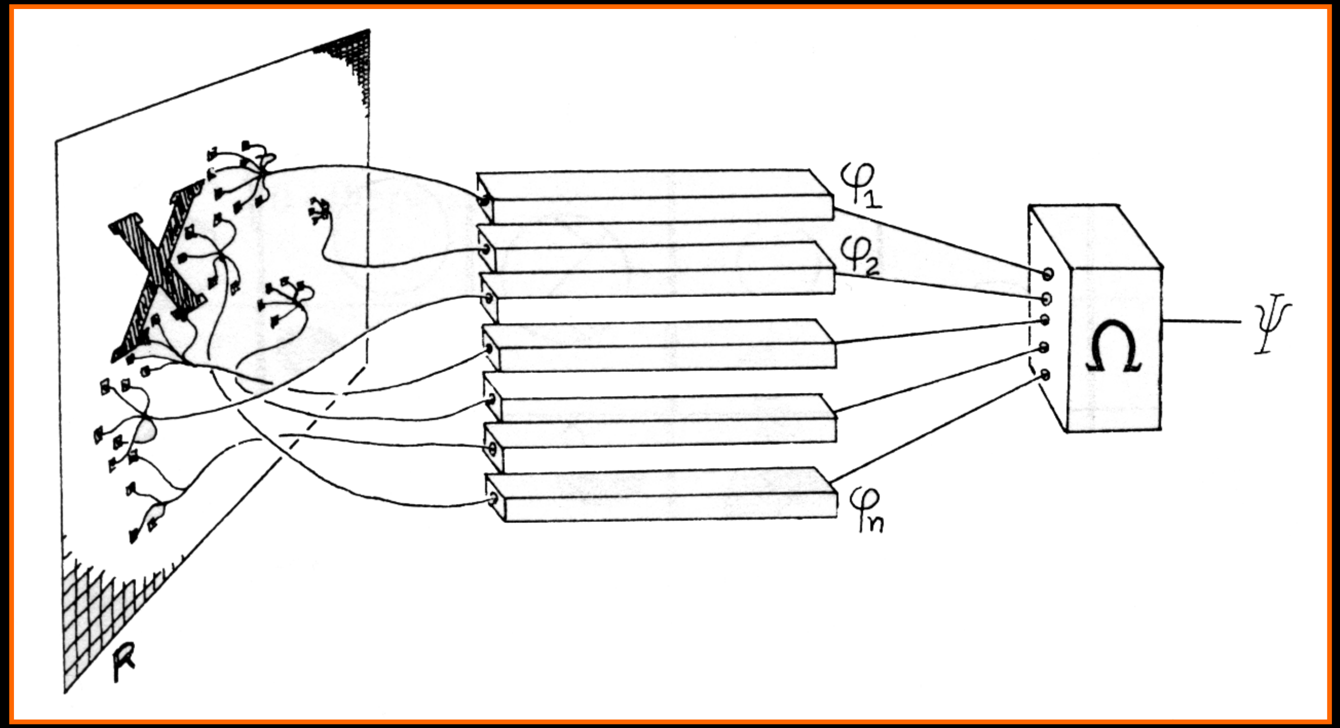
The Perceptron

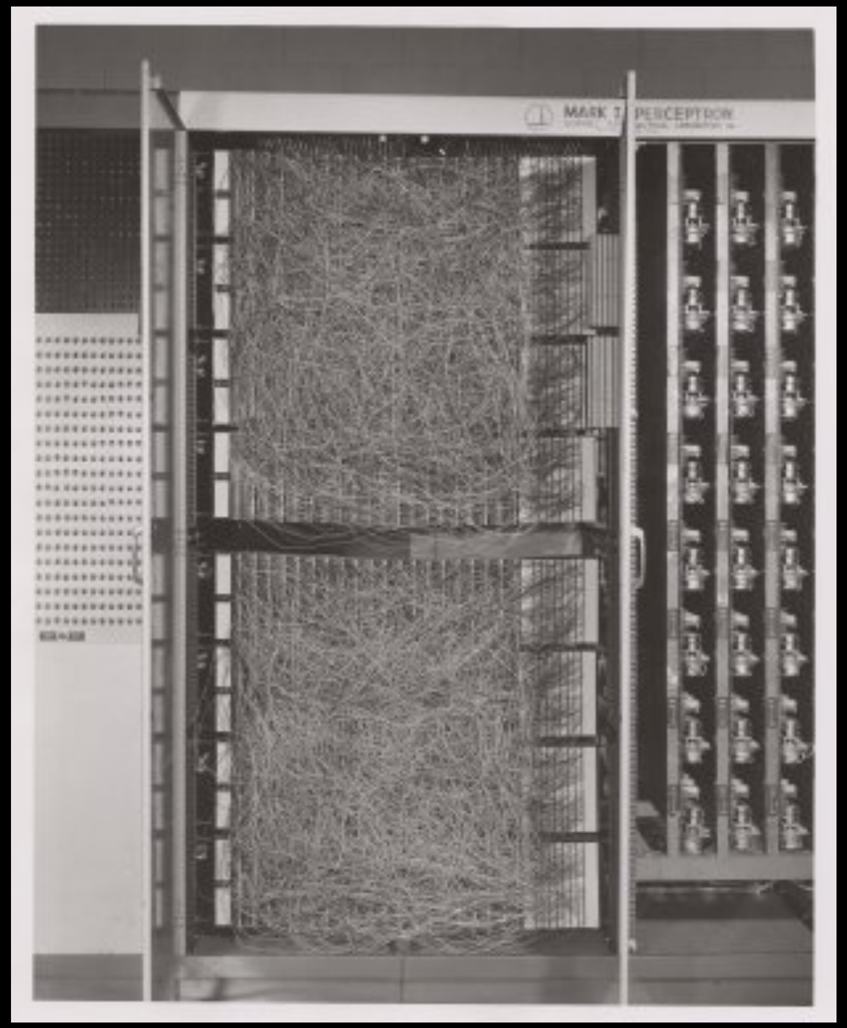
Frank Rosenblatt, 1957



The Perceptron

Frank Rosenblatt, 1957





Mark 1 Perceptron

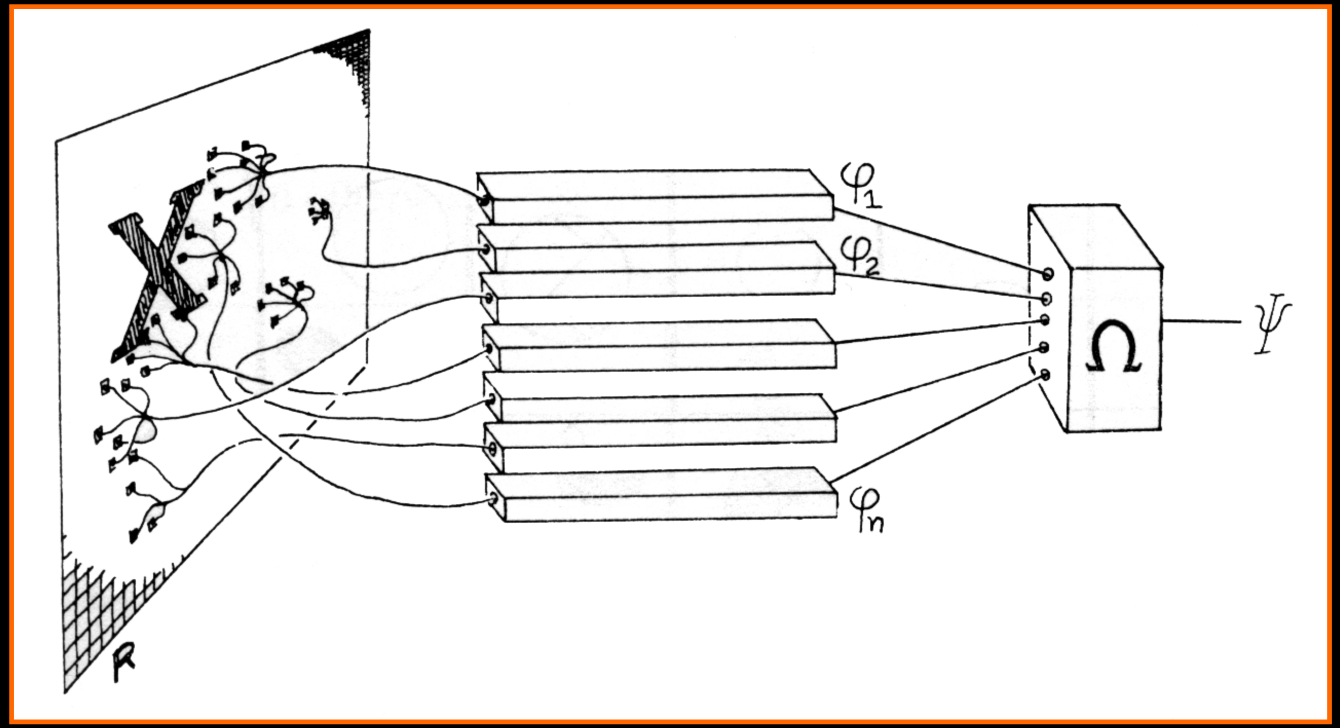
Input:
400 photocells

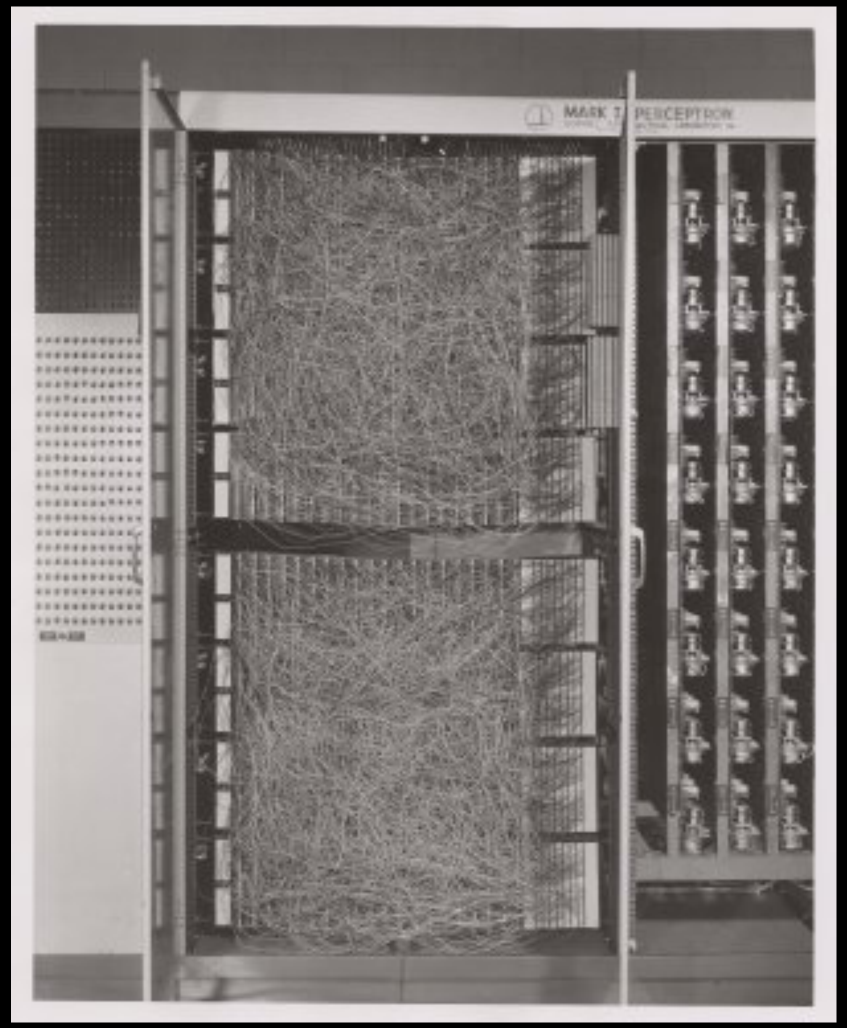
weights:
potentiometers

weight updates:
electric motors

The Perceptron

Frank Rosenblatt, 1957





Mark 1 Perceptron

Input:
400 photocells

weights:
potentiometers

weight updates:
electric motors

1958 New York Times...

NEW NAVY DEVICE LEARNS BY DOING

Psychologist Shows Embryo of Computer Designed to Read and Grow Wiser

WASHINGTON, July 7 (UPI)—The Navy revealed the embryo of an electronic computer today that it expects will be able to walk, talk, see, write, reproduce itself and be conscious of its existence.

The embryo—the Weather Bureau's $2,000,000 "704" computer—learned to differentiate between right and left after fifty attempts in the Navy's demonstration for newsmen.

The service said it would use this principle to build the first of its Perceptron thinking machines that will be able to read and write. It is expected to be finished in about a year at a cost of $100,000.

Dr. Frank Rosenblatt, designer of the Perceptron, conducted the demonstration. He said the machine would be the first device to think as the human brain. As do human beings, Perceptron will make mistakes at first, but will grow wiser as it gains experience, he said.

Dr. Rosenblatt, a research psychologist at the Cornell Aeronautical Laboratory, Buffalo, said Perceptrons might be fired to the planets as mechanical space explorers.

Without Human Controls

The Navy said the perceptron would be the first non-living mechanism "capable of receiving, recognizing and identifying its surroundings without any human training or control."

The "brain" is designed to remember images and information it has perceived itself. Ordinary computers remember only what is fed into them on punch cards or magnetic tape.

Later Perceptrons will be able to recognize people and call out their names and instantly translate speech in one language to speech or writing in another language, it was predicted.

Mr. Rosenblatt said in principle it would be possible to build brains that could reproduce themselves on an assembly line and which would be conscious of their existence.

In today's demonstration, the "704" was fed two cards, one with squares marked on the left side and the other with squares on the right side.

Learns by Doing

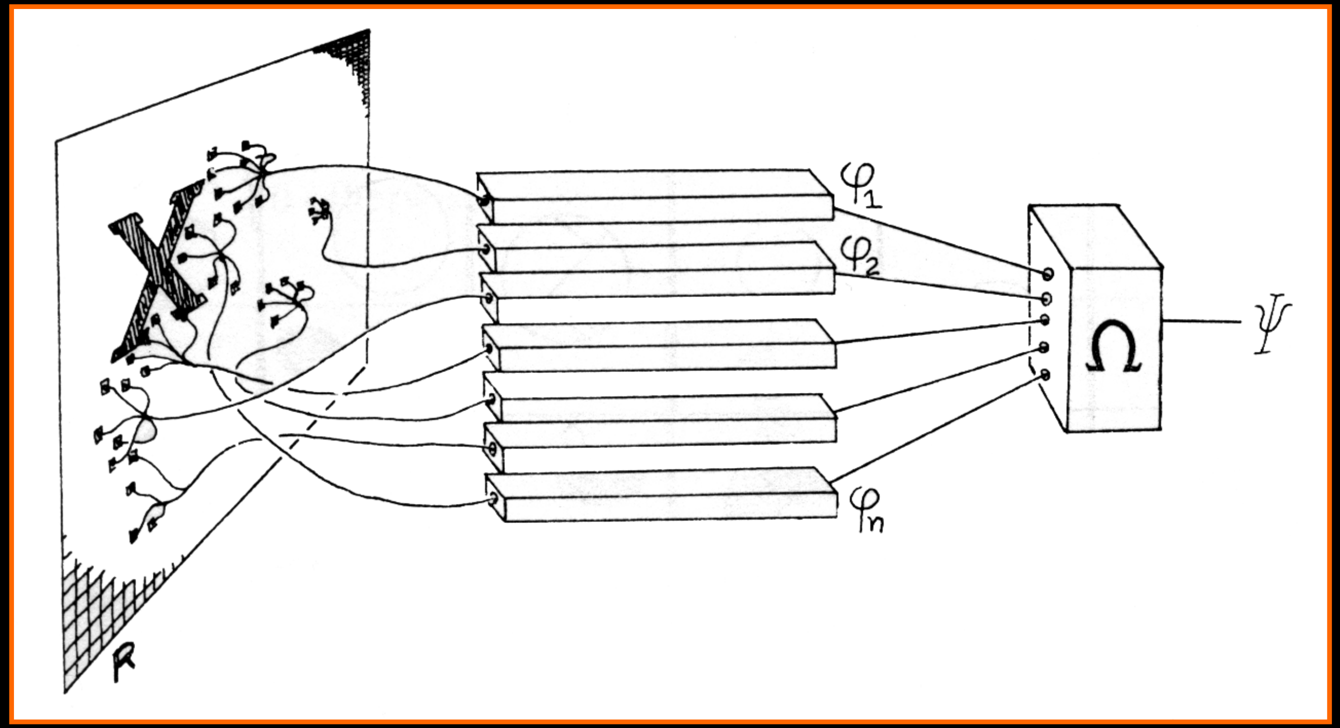
In the first fifty trials, the machine made no distinction between them. It then started registering a "Q" for the left squares and "O" for the right squares.

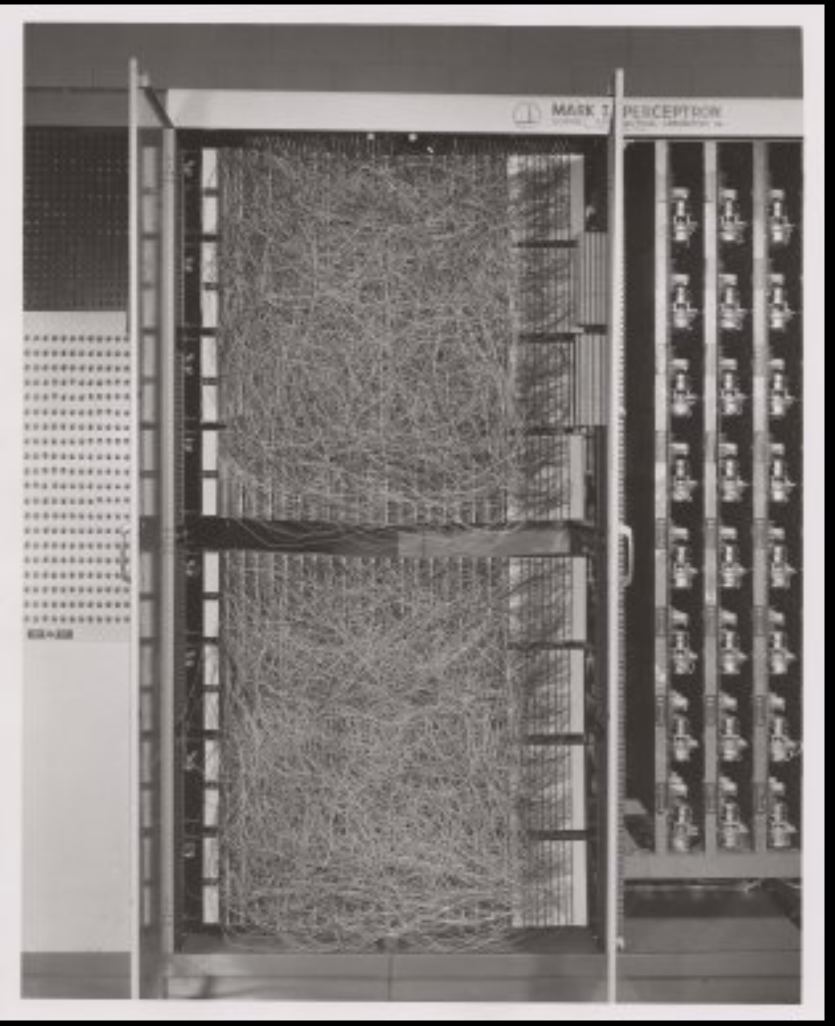
Dr. Rosenblatt said he could explain why the machine learned only in highly technical terms. But he said the computer had undergone a "self-induced change in the wiring diagram."

The first Perceptron will have about 1,000 electronic "association cells" receiving electrical impulses from an eye-like scanning device with 400 photo-cells. The human brain has 10,000,000,000 responsive cells, including 100,000,000 connections with the eyes.

The Perceptron

Frank Rosenblatt, 1957





Mark 1 Perceptron

Input:
400 photocells

weights:
potentiometers

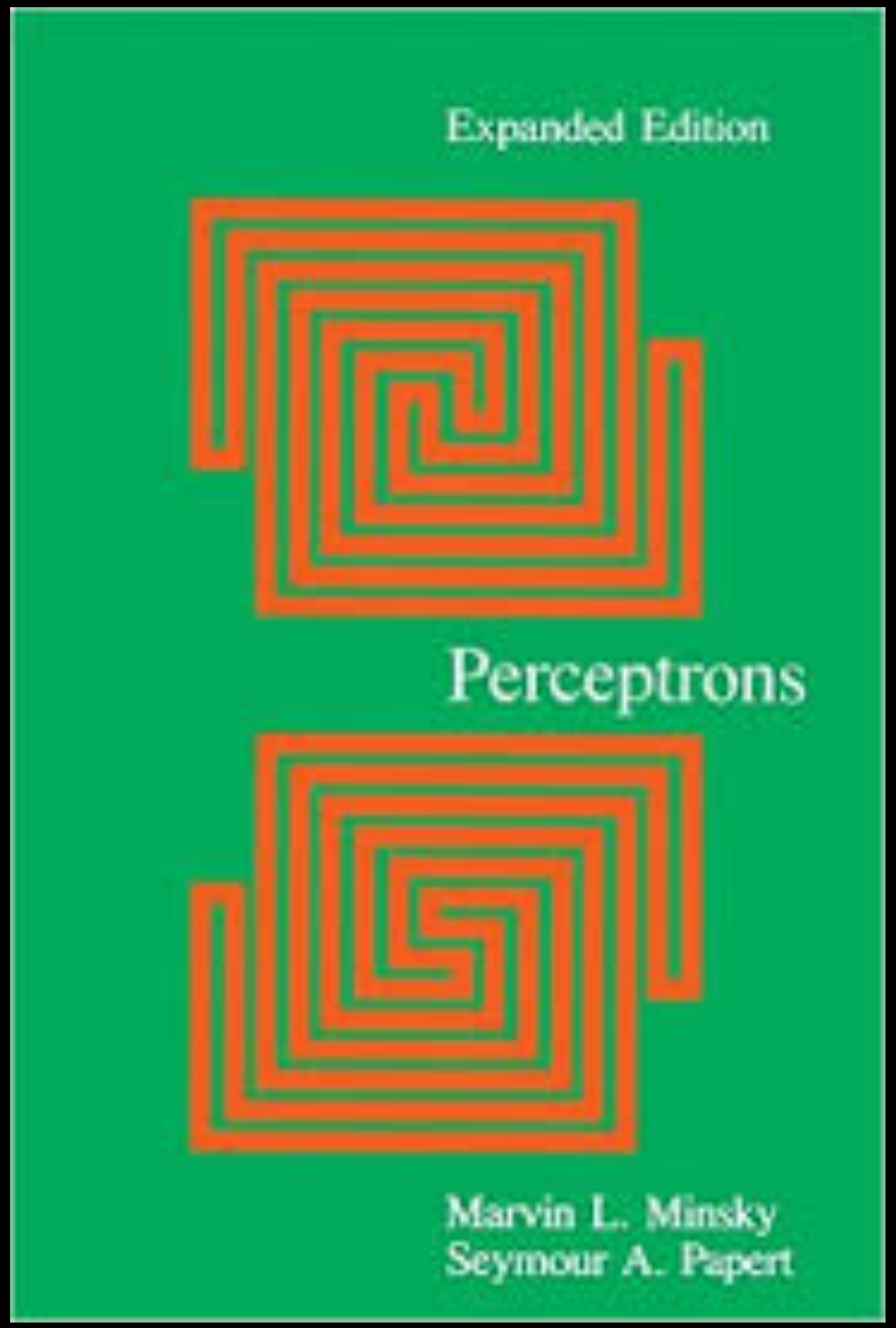
weight updates:
electric motors

- **Minsky & Papert** *(1969)*
  - **Perceptrons can't learn simple boolean functions (e.g., XOR)**

  ∴ **not computationally general**

The Perceptron

Frank Rosenblatt, 1957



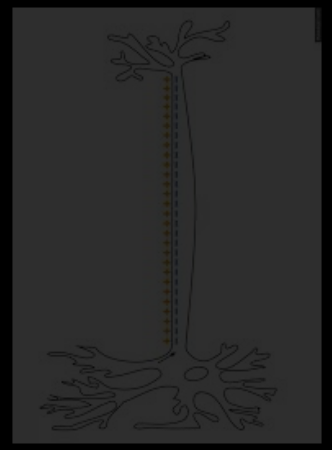
- **Minsky & Papert** *(1969)*
  - **Perceptrons can't learn simple boolean functions (e.g., XOR)**

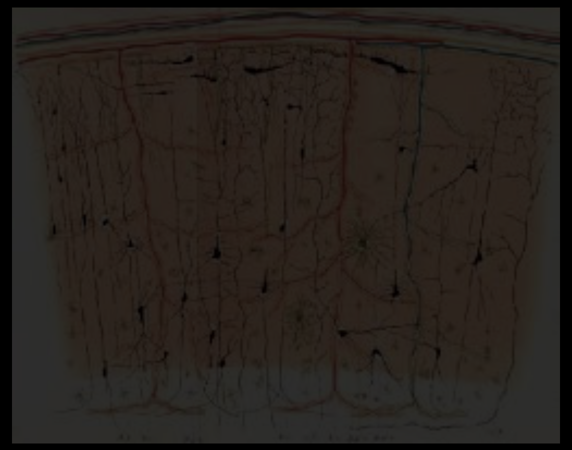
  ∴ **not computationally general**

Brief Historical Review

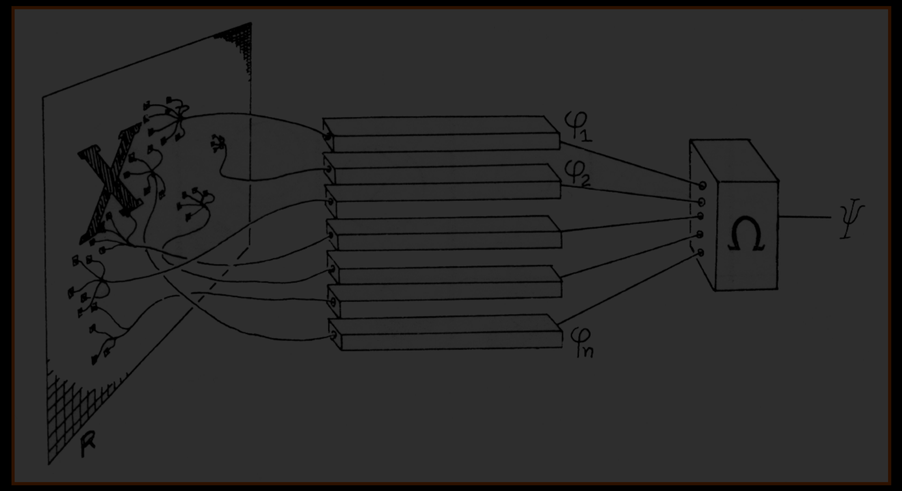
Brief Historical Review

- Spiking Neurons and Neural Assemblies
  - McCulloch & Pitts (neuroscience)
  - Hebb (psychology)
  - Norbert Weiner (mathematics)



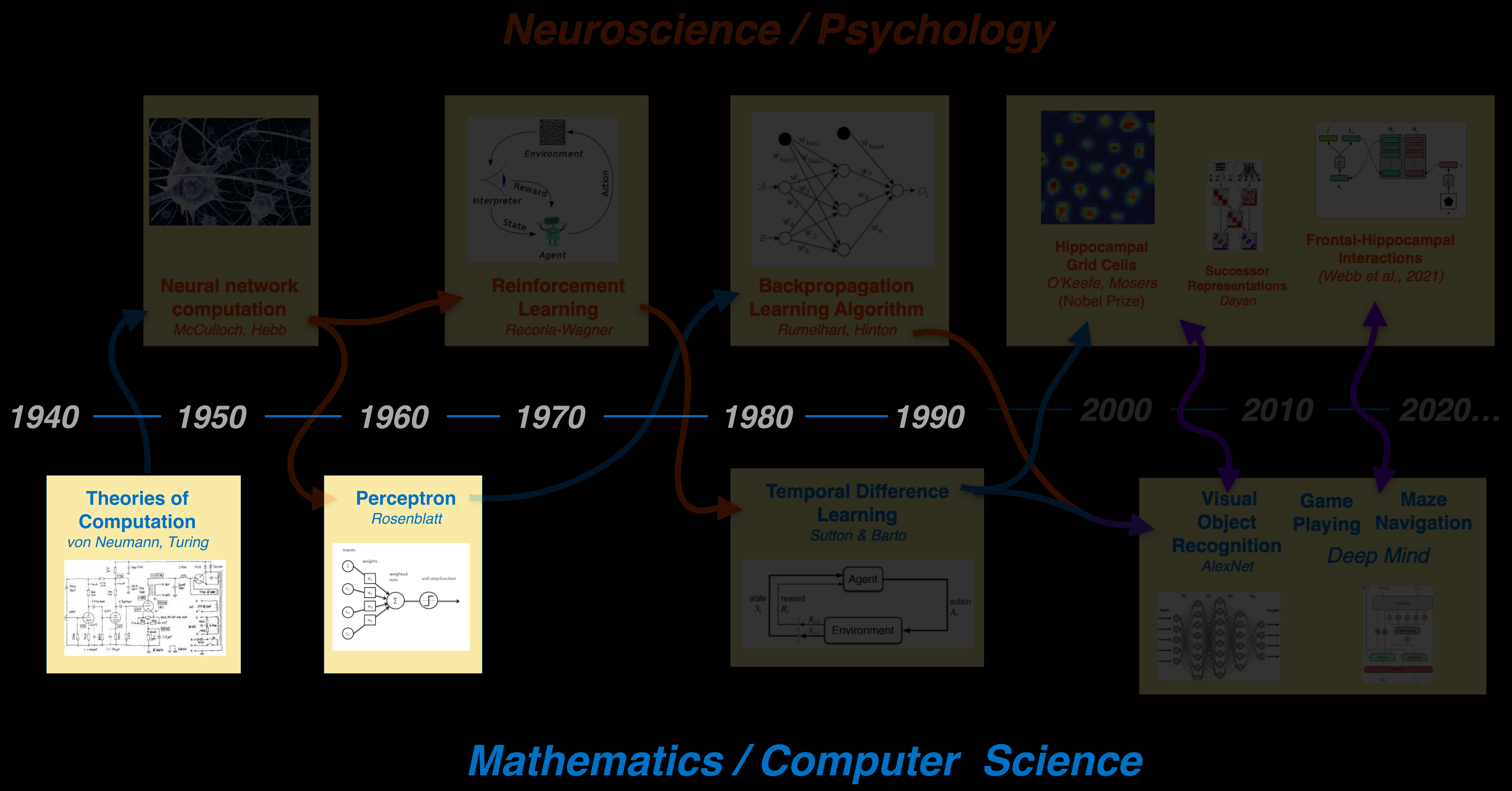


- Early neural network models
  - Rosenblatt's Perceptron
  - Minsky & Papert: demise of the Perceptron

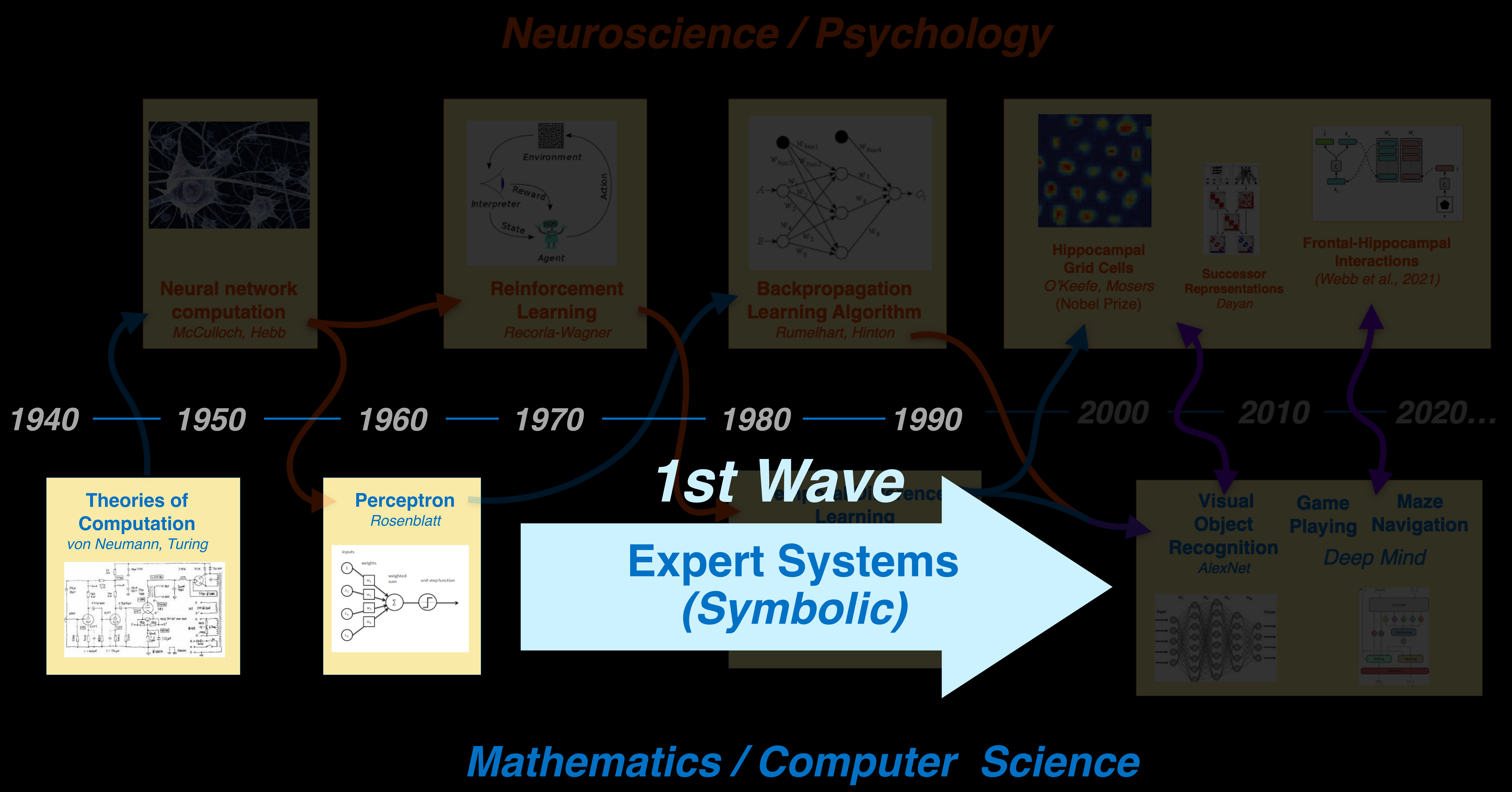


- AI, cognitive science and the symbolic approach:
  - The physical symbol system hypothesis (Newell & Simon)
  - von Neumann Architecture and the computer metaphor
  - The golden years of AI…

Brief Historical Review

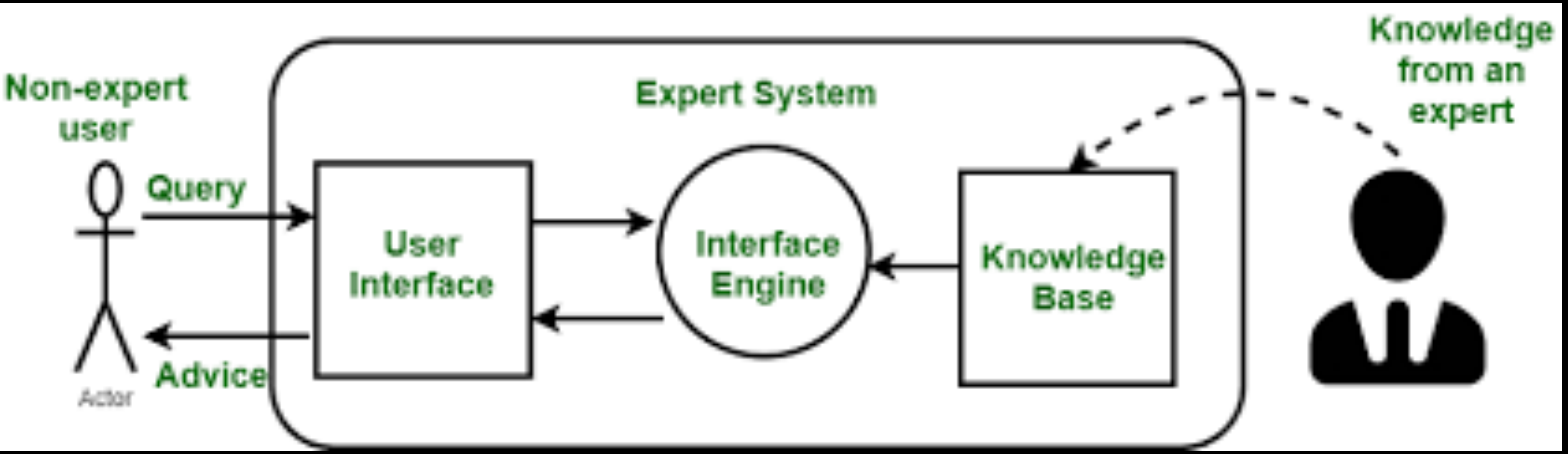


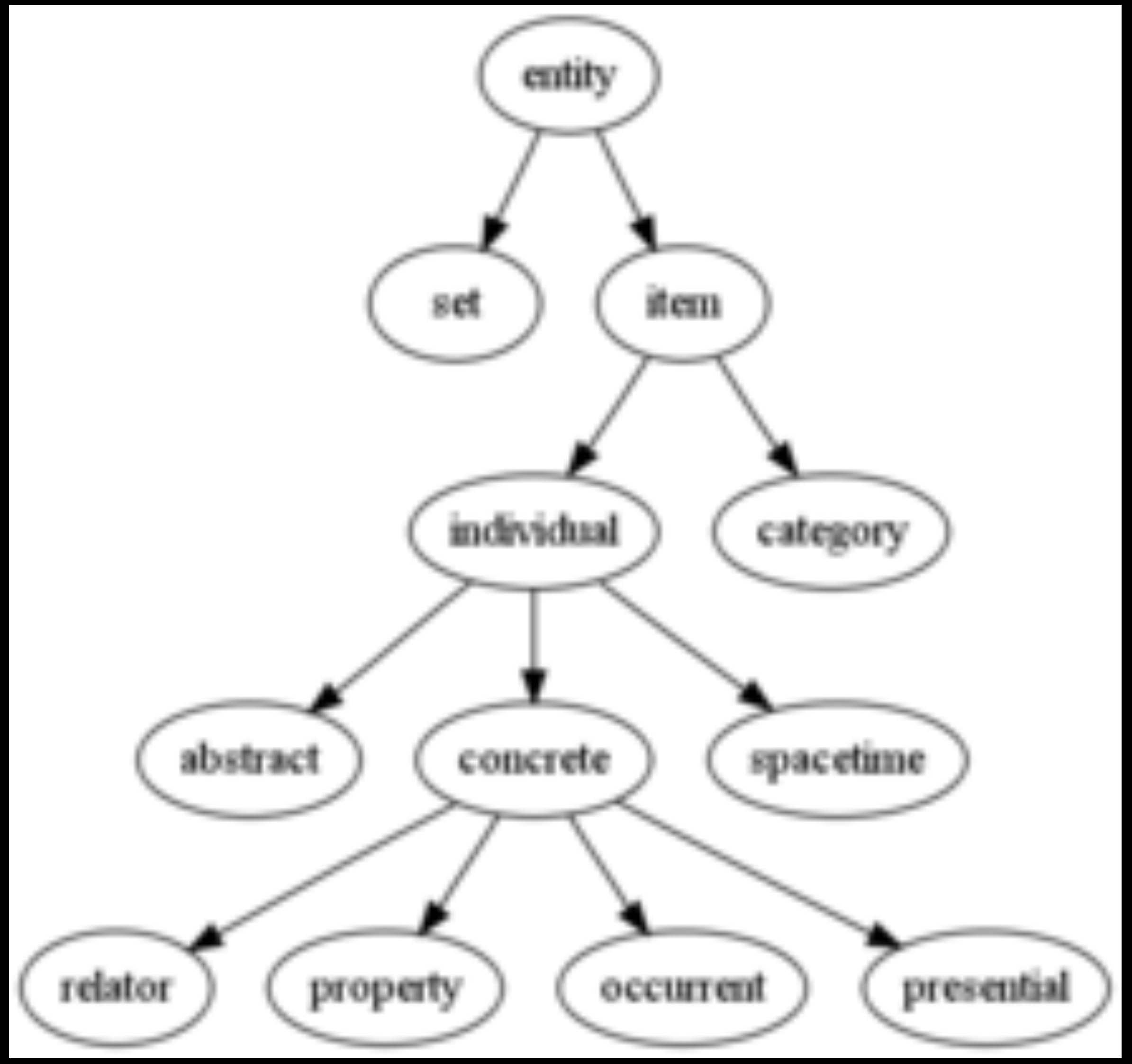
Brief Historical Review

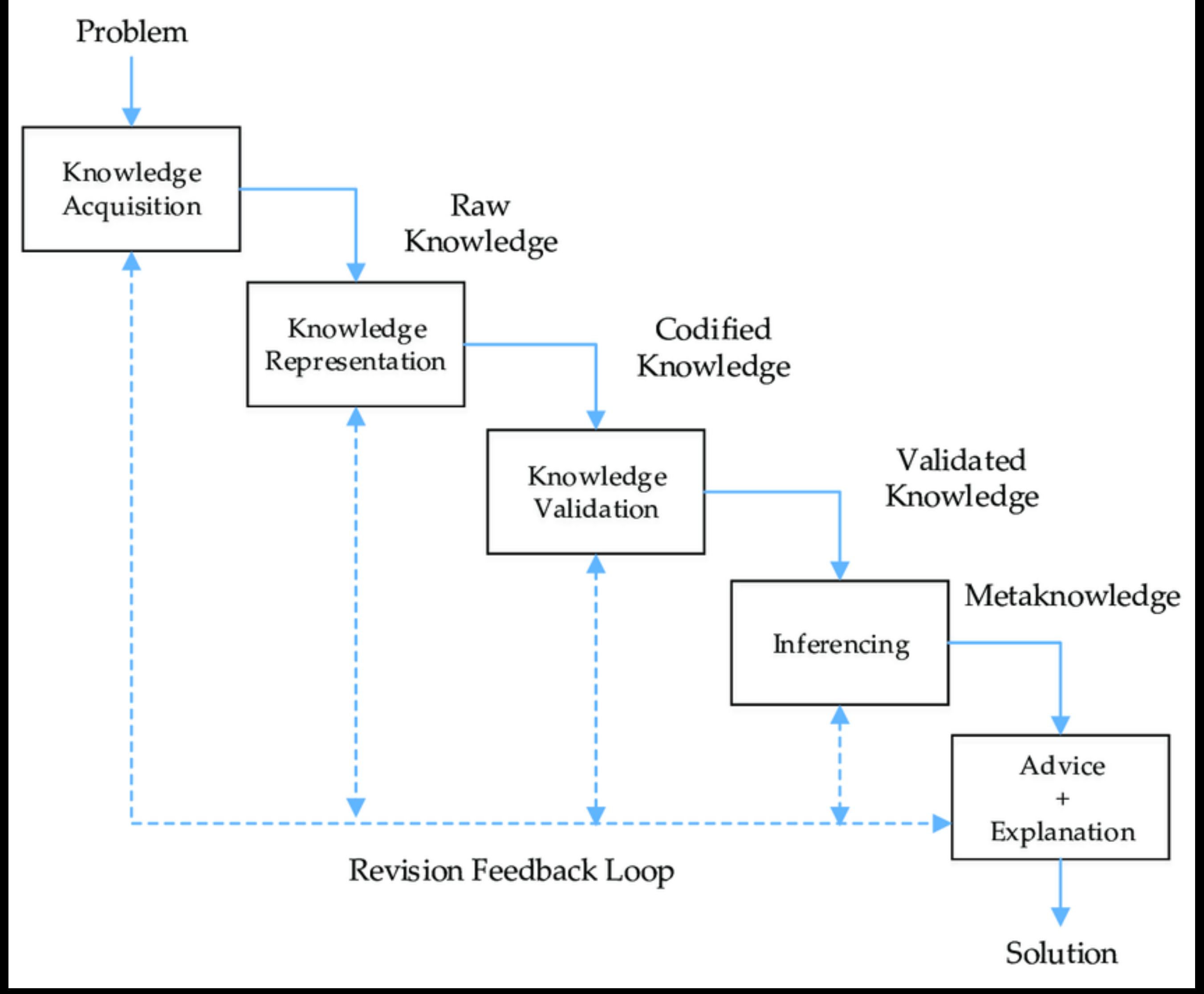


Classical AI

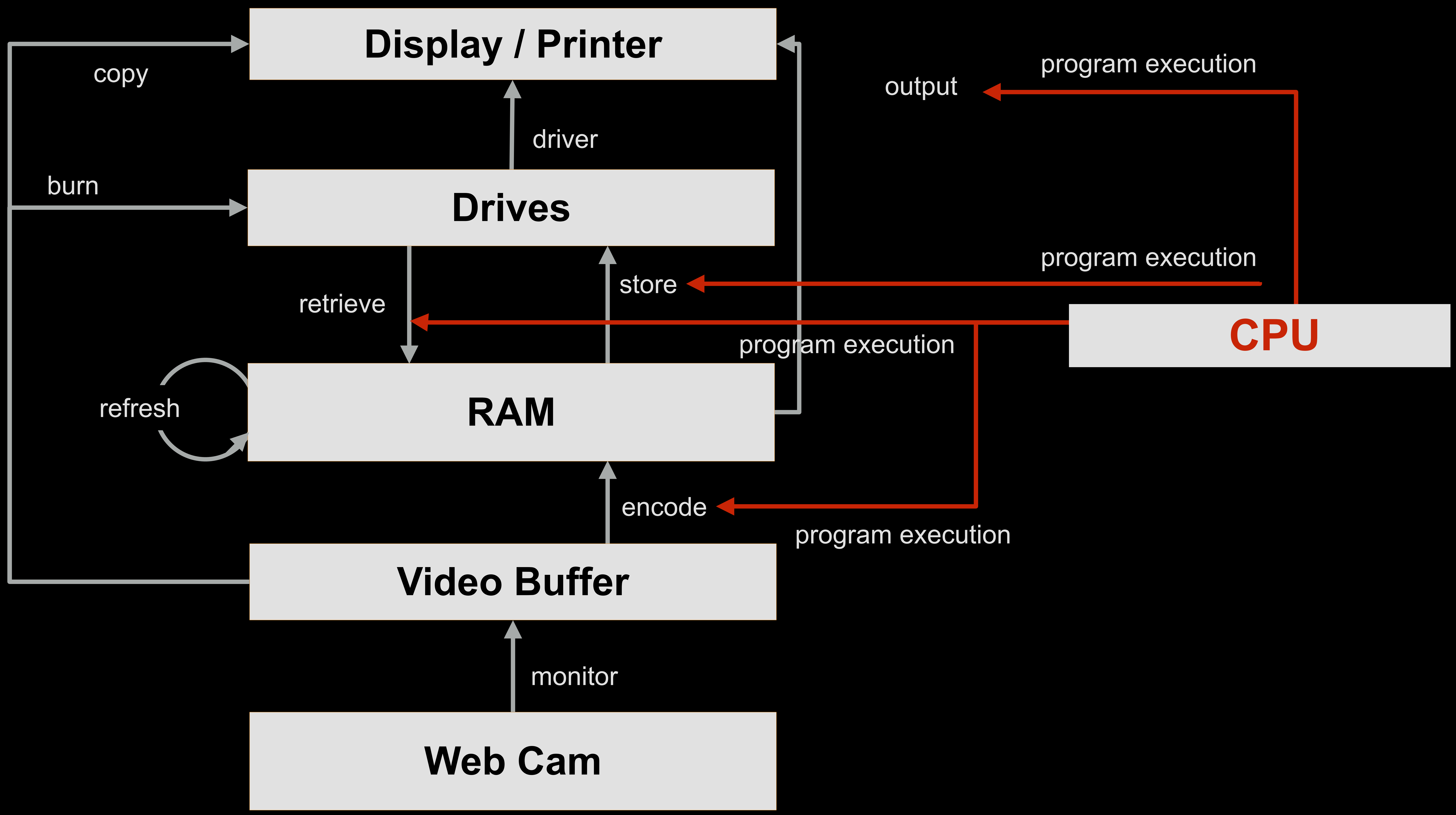
Expert Systems and Knowledge Engineering



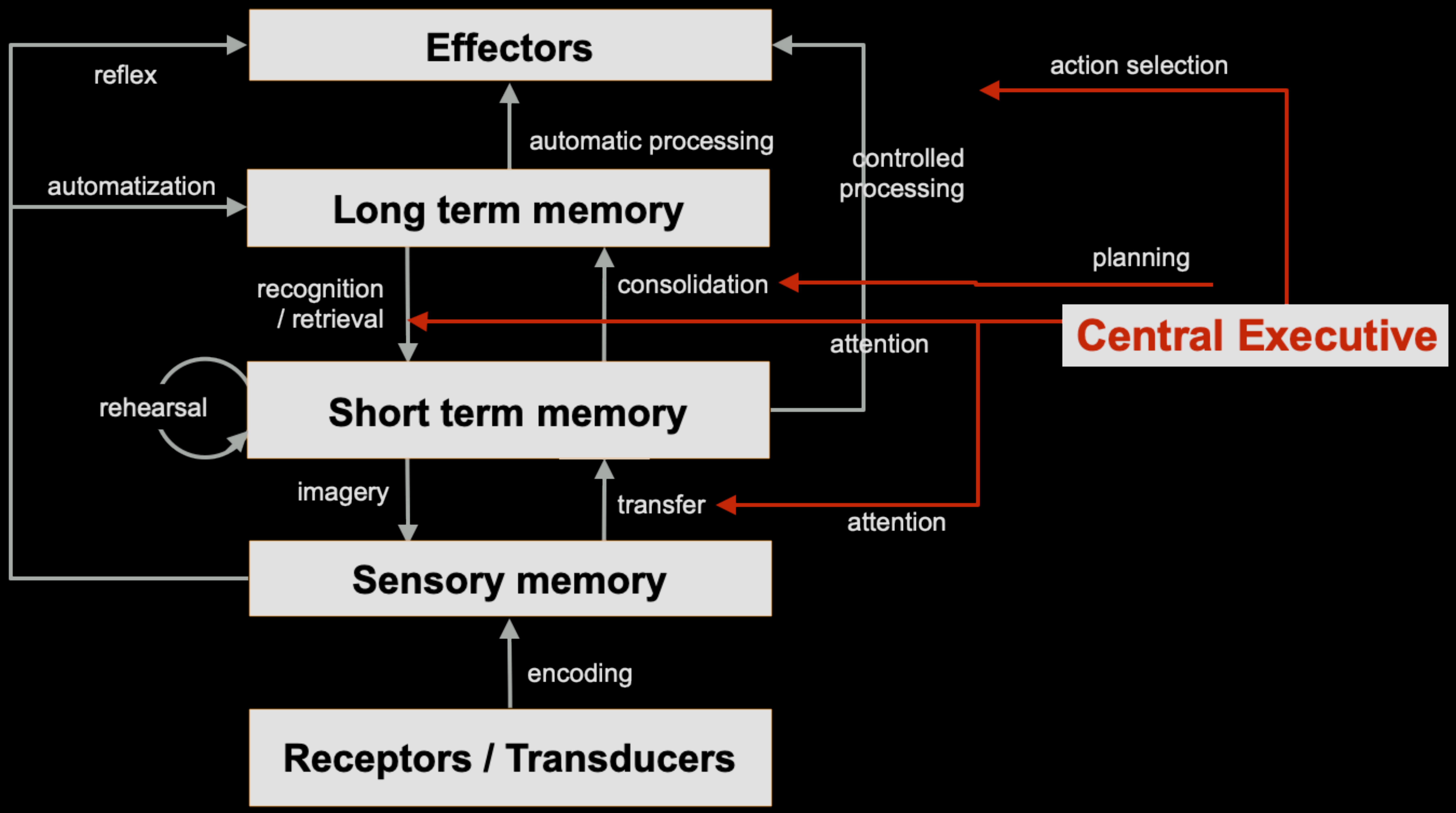




The Von Neumann Computer

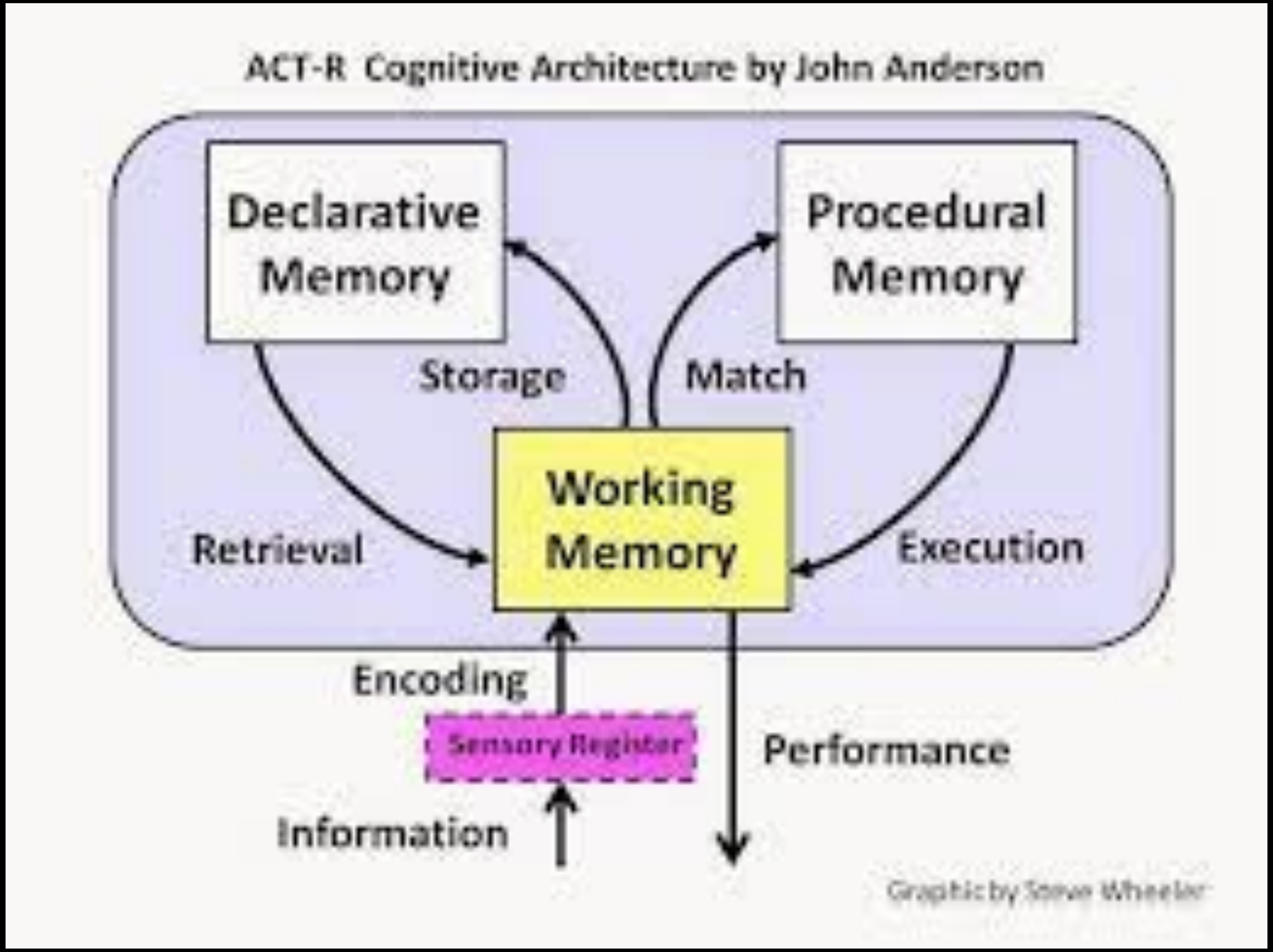


Metaphor of the Mind



Symbolic Models of Cognition

Production System Models - ACT-R



ACT-R

...the capacity to come up with abstract solutions to problems is one ability that is frequently cited with almost mystical awe. A good example of this is the ability to write recursive programs...

The simulation is capable of solving the same problems as [human] participants. It can actually interact with the same experimental software as the participants, execute the same scanning actions, read the same computer screen, and execute the same motor responses with very similar timing

We have studied extensively how people write recursive programs (e.g., Anderson, Farrell, & Sauers, 1984; Pirolli & Anderson, 1985). To test our understanding of the process, we have developed computer simulations that are themselves capable of writing recursive programs in the same way humans do. Underlying this skill are about 500 knowledge units called *production rules*. For instance, one of these production rules for programming recursion, which might apply in the midst of the problem solving, is

IF the goal is to identify the recursive relationship in a function with a number argument
THEN set as subgoals to
1. Find the value of the function for some N
2. Find the value of the function for $N-1$
3. Try to identify the relationship between the two answers.

Thus, in the case above, this might lead to finding that factorial(5) = 120 (Step 1), factorial(4) = 24 (Step 2), and that factorial (N) = factorial ($N-1$) $\times$ N (Step 3).

We (e.g., Anderson, Boyle, Corbett, & Lewis, 1990; Anderson, Corbett, Koedinger, & Pelletier, 1995; Anderson & Reiser, 1985) have created computer-based instructional systems, called *intelligent tutors*, for teaching cognitive skills based on this kind of production-rule analysis. By basing instruction on such rules, we have been able to increase students' rate of learning by a factor of 3.

ACT-R

...the capacity to come up with abstract solutions to problems is one ability that is frequently cited with almost mystical awe. A good example of this is the ability to write recursive programs...

All that there is to intelligence is the simple accrual and tuning of many small units of knowledge that in total produce complex cognition.
The whole is no more than the sum of its parts, but it has lots of parts

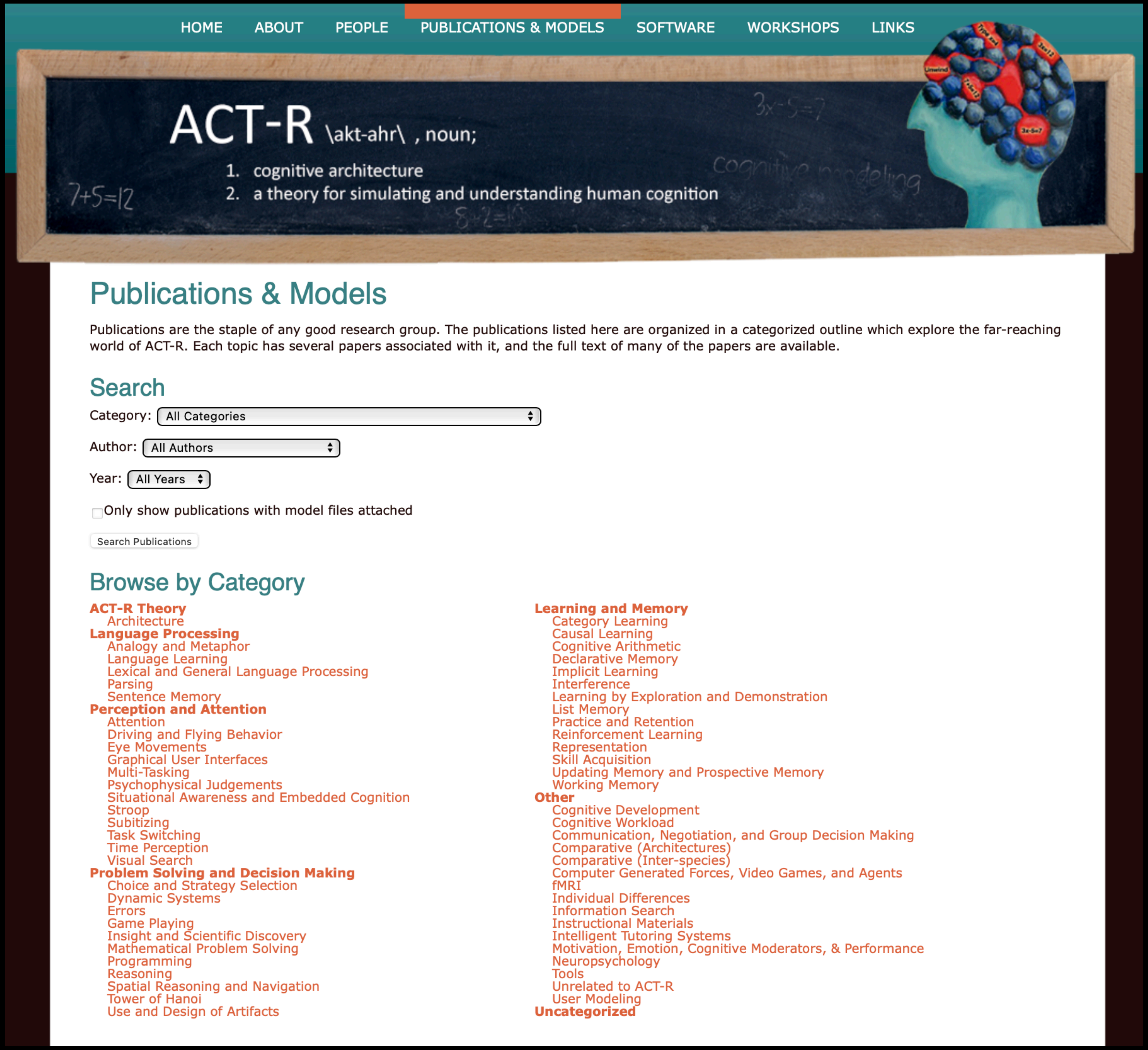
We have studied extensively how people write recursive programs (e.g., Anderson, Farrell, & Sauers, 1984; Pirolli & Anderson, 1985). To test our understanding of the process, we have developed computer simulations that are themselves capable of writing recursive programs in the same way humans do. Underlying this skill are about 500 knowledge units called *production rules*. For instance, one of these production rules for programming recursion, which might apply in the midst of the problem solving, is

IF the goal is to identify the recursive relationship in a function with a number argument
THEN set as subgoals to
1. Find the value of the function for some N
2. Find the value of the function for $N-1$
3. Try to identify the relationship between the two answers.

Thus, in the case above, this might lead to finding that factorial(5) = 120 (Step 1), factorial(4) = 24 (Step 2), and that factorial (N) = factorial ($N-1$) $\times$ N (Step 3).

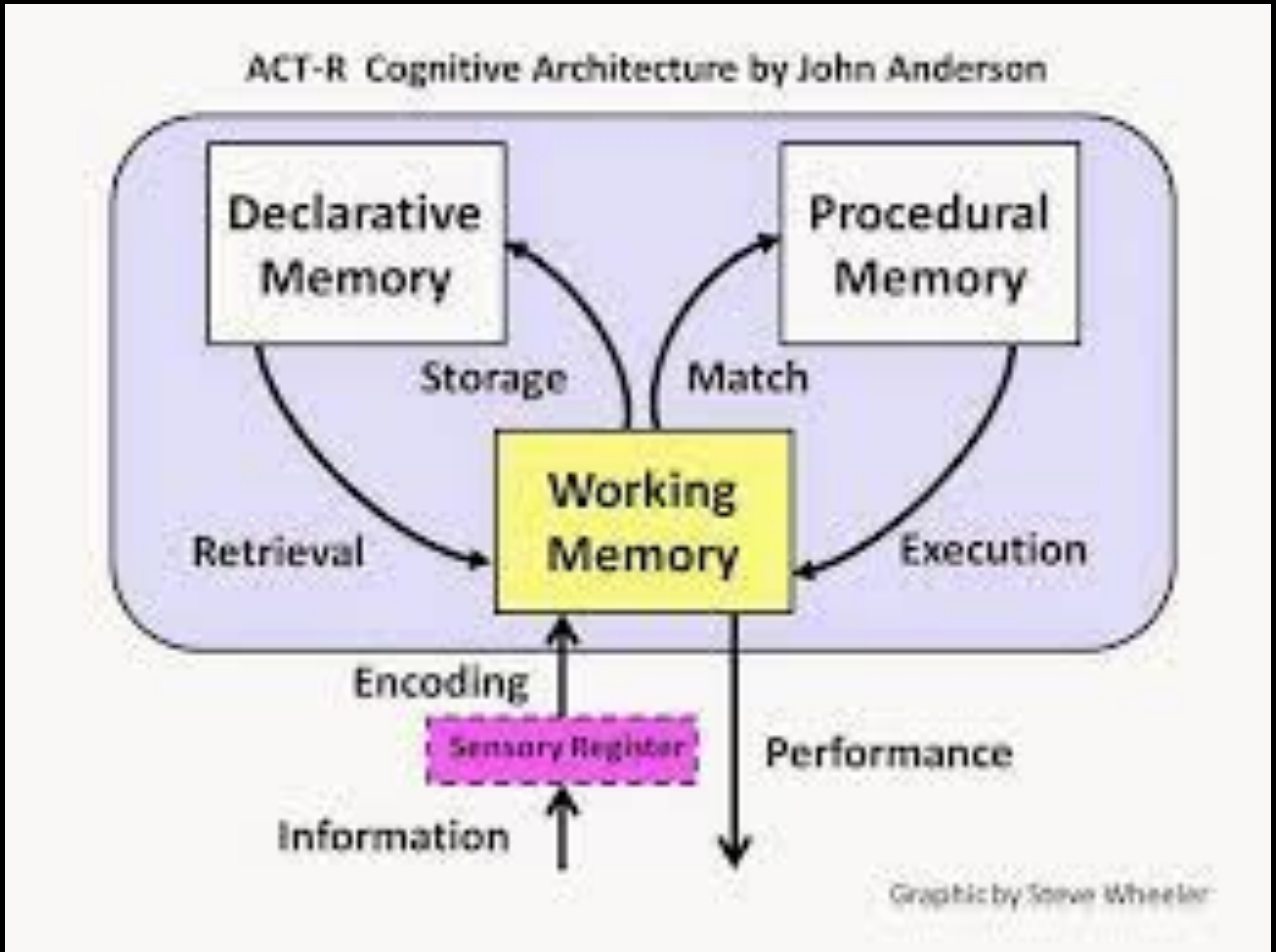
We (e.g., Anderson, Boyle, Corbett, & Lewis, 1990; Anderson, Corbett, Koedinger, & Pelletier, 1995; Anderson & Reiser, 1985) have created computer-based instructional systems, called *intelligent tutors,* for teaching cognitive skills based on this kind of production-rule analysis. By basing instruction on such rules, we have been able to increase students' rate of learning by a factor of 3.

ACT-R


HOME
ABOUT
PEOPLE
PUBLICATIONS & MODELS
SOFTWARE
WORKSHOPS
LINKS
ACT-R \akt-ahr\ , noun;
1. cognitive architecture
2. a theory for simulating and understanding human cognition
7+5=12
cognitive modeling
Publications & Models
Publications are the staple of any good research group. The publications listed here are organized in a categorized outline which explore the far-reaching world of ACT-R. Each topic has several papers associated with it, and the full text of many of the papers are available.
Search
Category: All Categories
Author: All Authors
Year: All Years
Only show publications with model files attached
Search Publications
Browse by Category
ACT-R Theory
Architecture
Language Processing
Analogy and Metaphor
Language Learning
Lexical and General Language Processing
Parsing
Sentence Memory
Perception and Attention
Attention
Driving and Flying Behavior
Eye Movements
Graphical User Interfaces
Multi-Tasking
Psychophysical Judgements
Situational Awareness and Embedded Cognition
Stroop
Subitizing
Task Switching
Time Perception
Visual Search
Problem Solving and Decision Making
Choice and Strategy Selection
Dynamic Systems
Errors
Game Playing
Insight and Scientific Discovery
Mathematical Problem Solving
Programming
Reasoning
Spatial Reasoning and Navigation
Tower of Hanoi
Use and Design of Artifacts
Learning and Memory
Category Learning
Causal Learning
Cognitive Arithmetic
Declarative Memory
Implicit Learning
Interference
Learning by Exploration and Demonstration
List Memory
Practice and Retention
Reinforcement Learning
Representation
Skill Acquisition
Updating Memory and Prospective Memory
Working Memory
Other
Cognitive Development
Cognitive Workload
Communication, Negotiation, and Group Decision Making
Comparative (Architectures)
Comparative (Inter-species)
Computer Generated Forces, Video Games, and Agents
fMRI
Individual Differences
Information Search
Instructional Materials
Intelligent Tutoring Systems
Motivation, Emotion, Cognitive Moderators, & Performance
Neuropsychology
Tools
Unrelated to ACT-R
User Modeling
Uncategorized

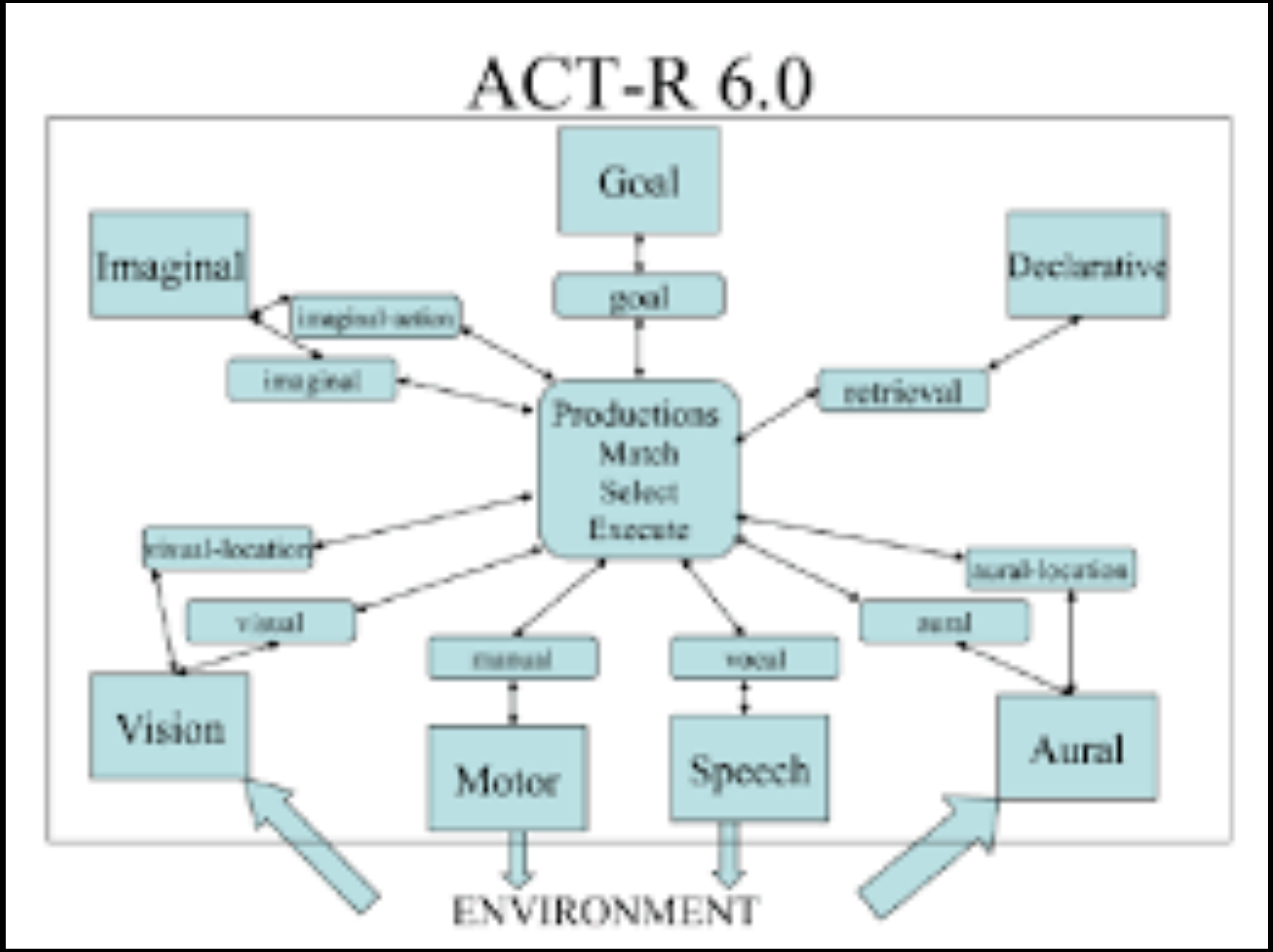
ACT-R

The whole is no more than the sum of its parts



ACT-R

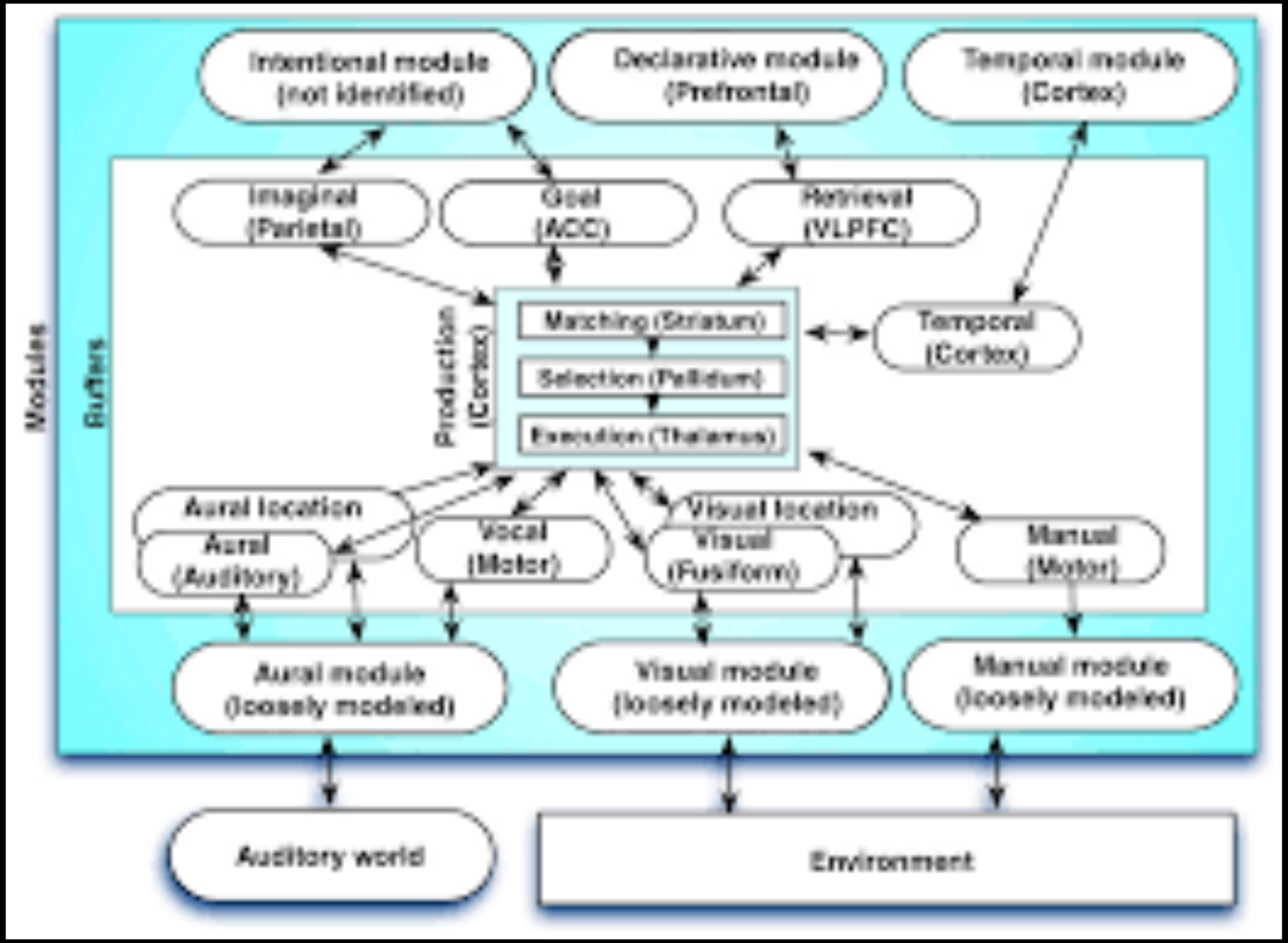
The whole is no more than the sum of its parts, but it has lots of parts



ACT-R

The whole is no more than the sum of its parts, but it has lots of parts

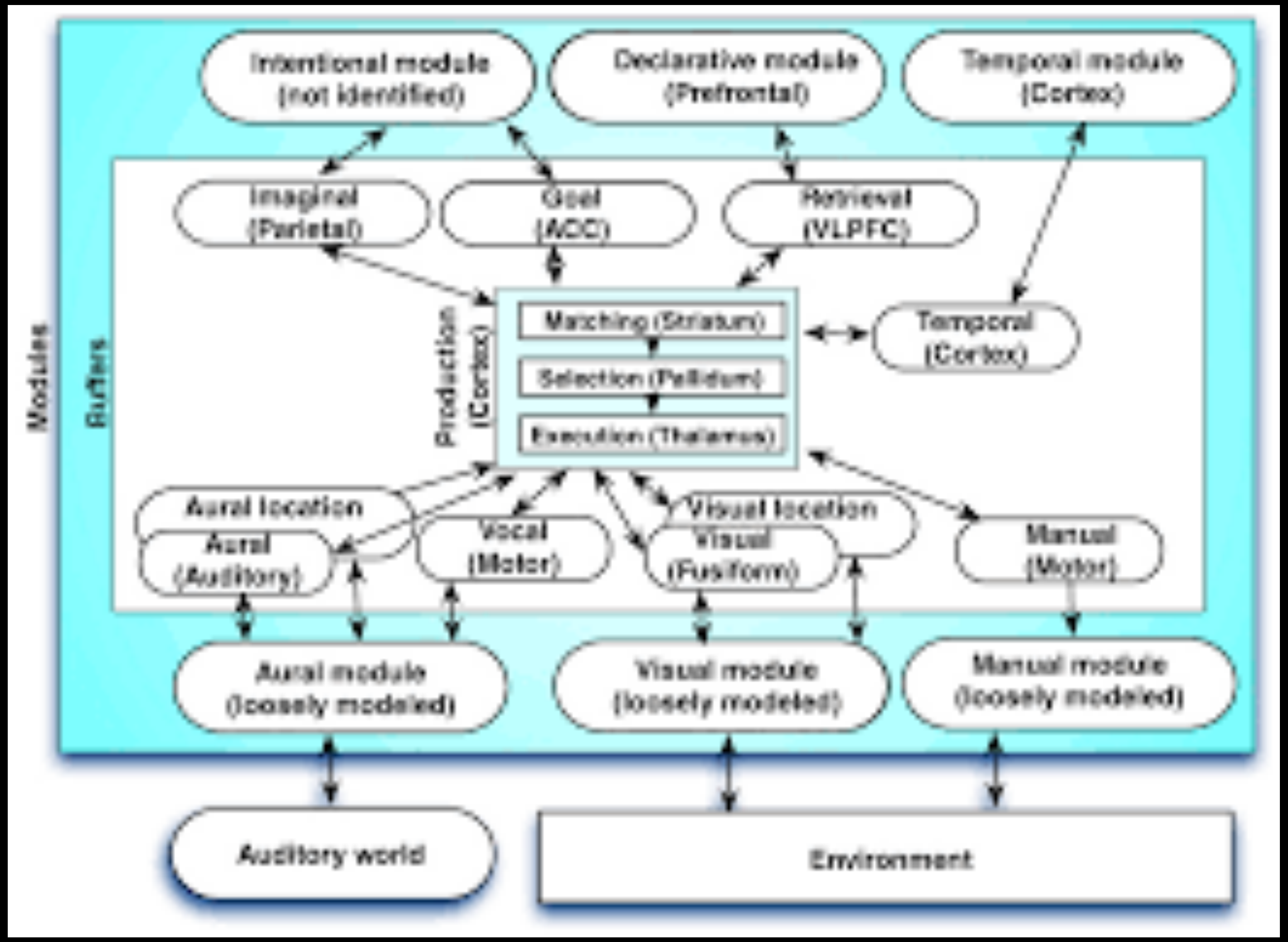
Really… _lots_ of parts



ACT-R

The whole is no more than the sum of its parts, but it has lots of parts

What about the brain?



Brief Historical Review

- Roots
  - McCulloch & Pitts (neuroscience)
  - Hebb (psychology)
- Early neural network models
  - Rosenblatt's Perceptron
  - Minsky & Papert: demise of the Perceptron
- AI and cognitive science
  - The physical symbol system hypothesis (Newell & Simon)
  - von Neumann Architecture and the computer metaphor
  - Knowledge engineering and the golden years of AI
  - Production system models of cognitive function
- Limits of the symbolic approach
  - Knowledge engineering (expert systems): programming vs. learning
  - Combinatorial explosion in highly contextual domains

    face recognition, natural language processing...

RED

THE
CAT

PFB

THE

CAT

Brief Historical Review

- Roots
  - McCulloch & Pitts (neuroscience)
  - Hebb (psychology)
- Early neural network models
  - Rosenblatt's Perceptron
  - Minsky & Papert: demise of the Perceptron
- AI and cognitive science
  - The physical symbol system hypothesis (Newell & Simon)
  - von Neumann Architecture and the computer metaphor
  - Knowledge engineering and the golden years of AI
  - Production system models of cognitive function
- Limits of the symbolic approach
  - Knowledge engineering: programming vs. learning
  - Combinatorial explosion in highly contextual domains

    face recognition, natural language processing...
  - **100 step challenge:**

Brief Historical Review

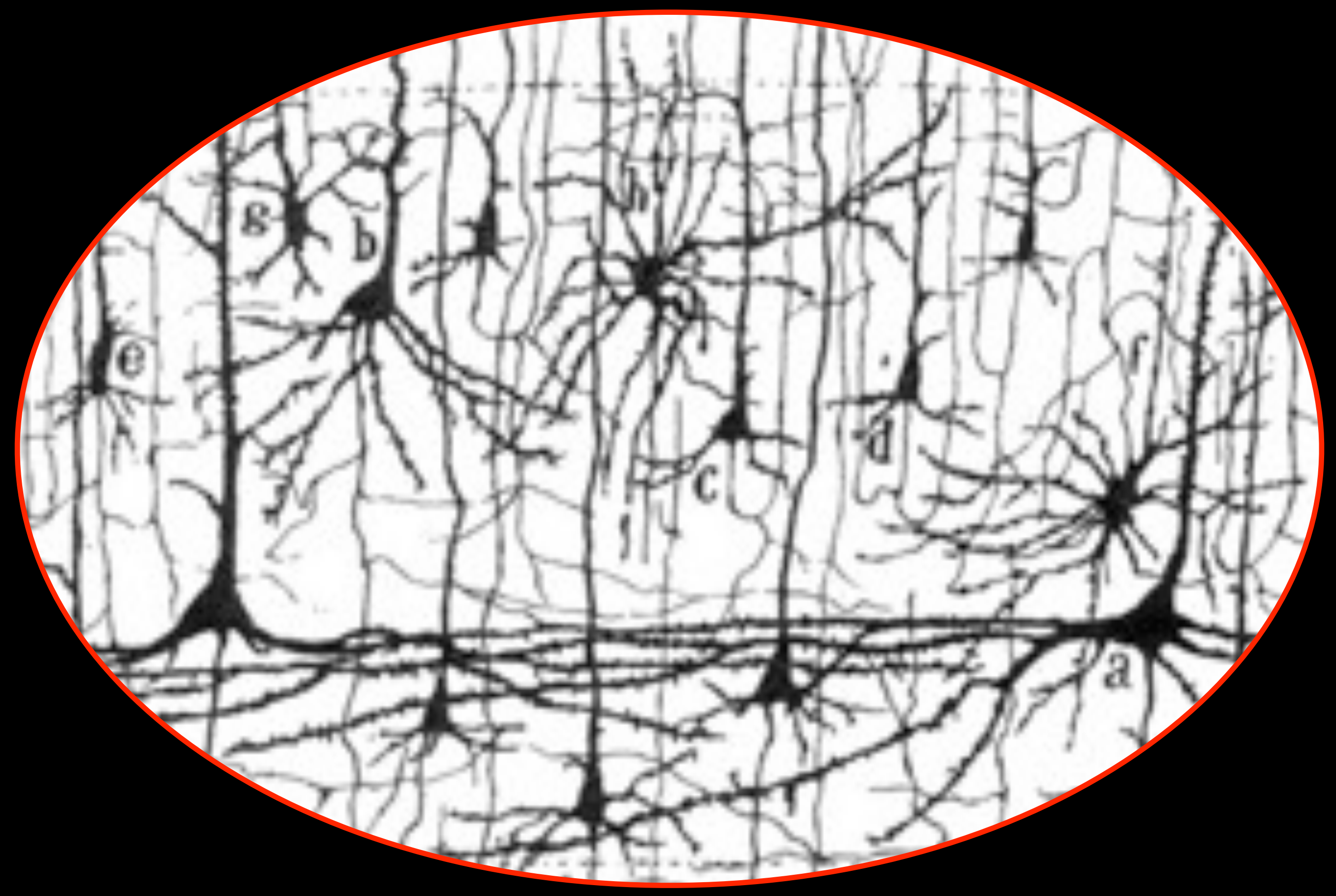
- Roots
  - McCulloch & Pitts (neuroscience)
  - Hebb (psychology)
- Early neural network models
  - Rosenblatt's Perceptron
  - Minsky & Papert: demise of the Perceptron
- AI and cognitive science
  - The physical symbol system hypothesis (Newell & Simon)
  - von Neumann Architecture and the computer metaphor
  - Knowledge engineering and the golden years of AI
  - Production system models of cognitive function
- Limits of the symbolic approach
  - Knowledge engineering: programming vs. learning
  - Combinatorial explosion in highly contextual domains

    face recognition, natural language processing...
  - **100 step challenge: the *brain* is instructive here, but…**

Scale of the Problem

Scale of the Problem

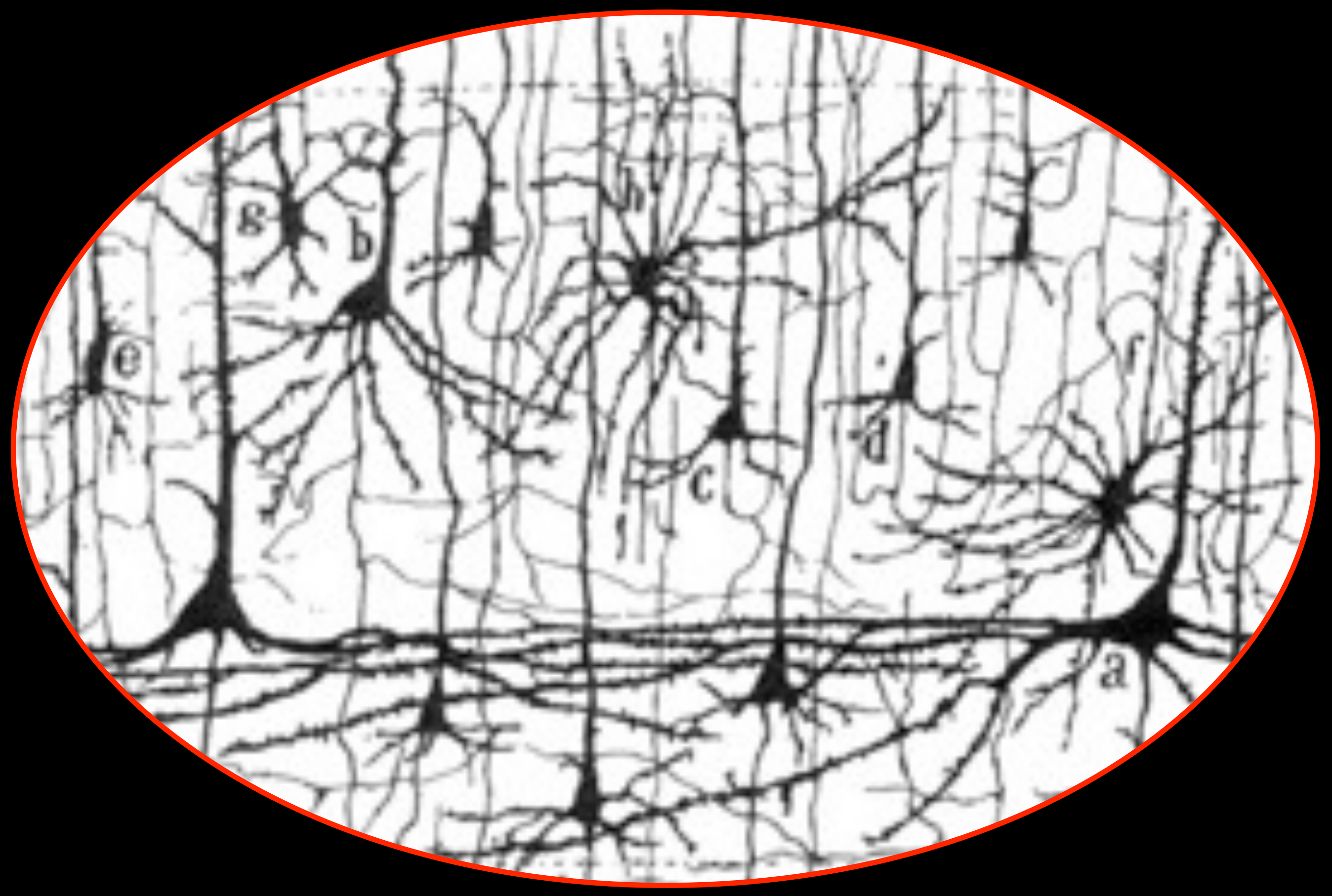
100 billion neurons



Scale of the Problem

100 billion neurons

100 thousand connections/neuron

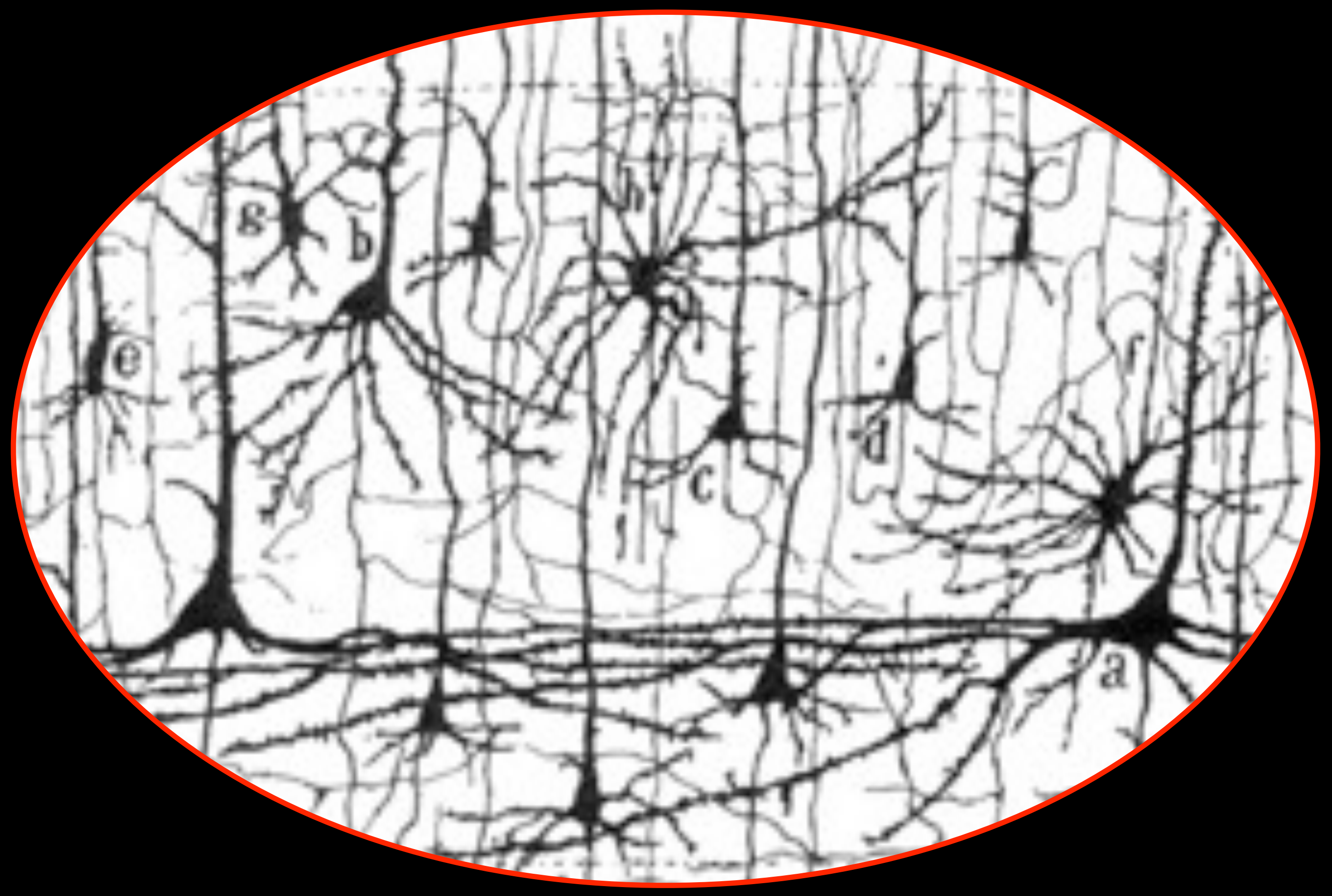


Scale of the Problem

100 billion neurons

100 thousand connections/neuron

= 100 trillion connections



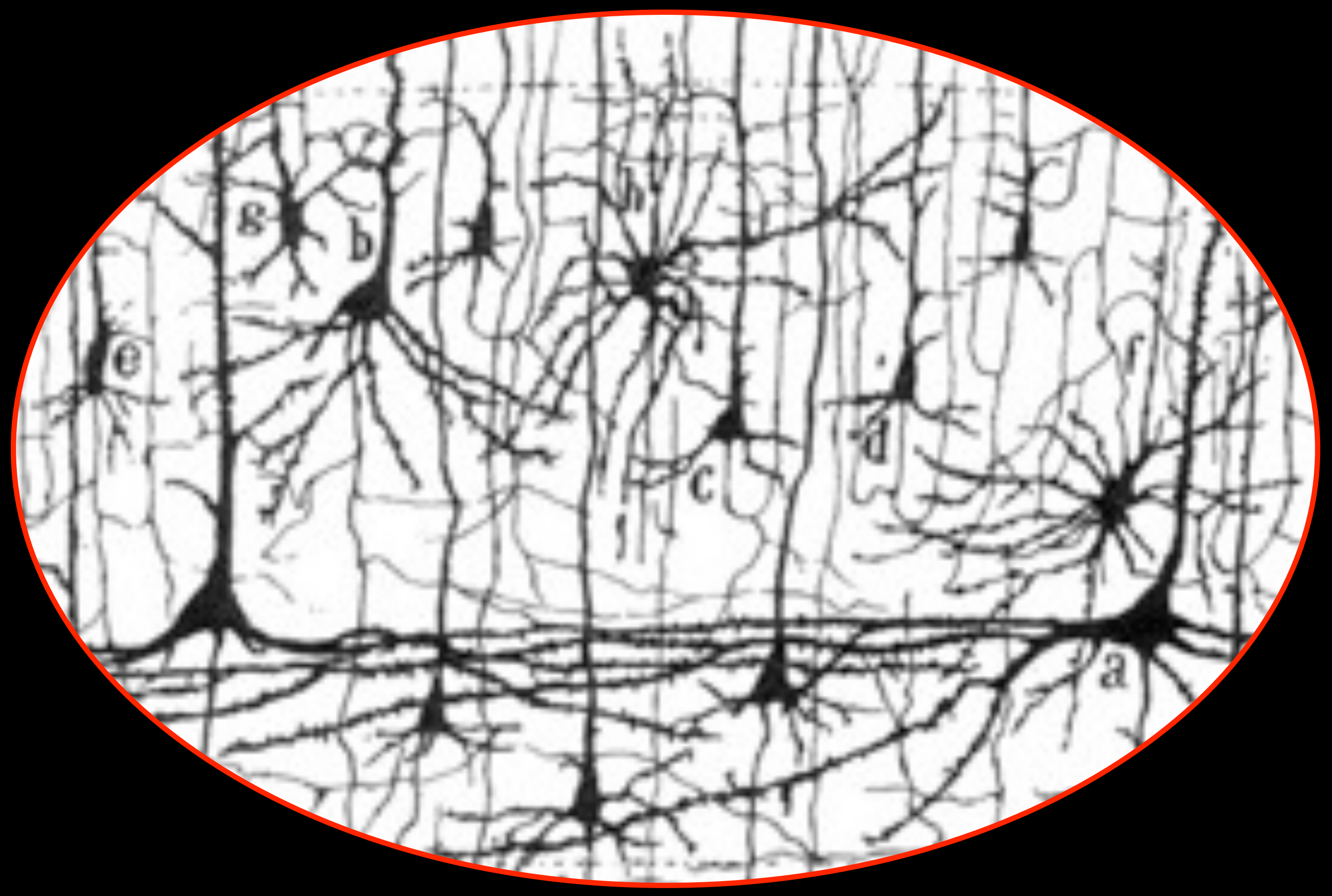
Scale of the Problem

100 billion neurons

100 thousand connections/neuron

= 100 trillion connections

More potential circuits than molecules in the universe



Scale of the Problem

100 billion neurons
100 thousand connections/neuron

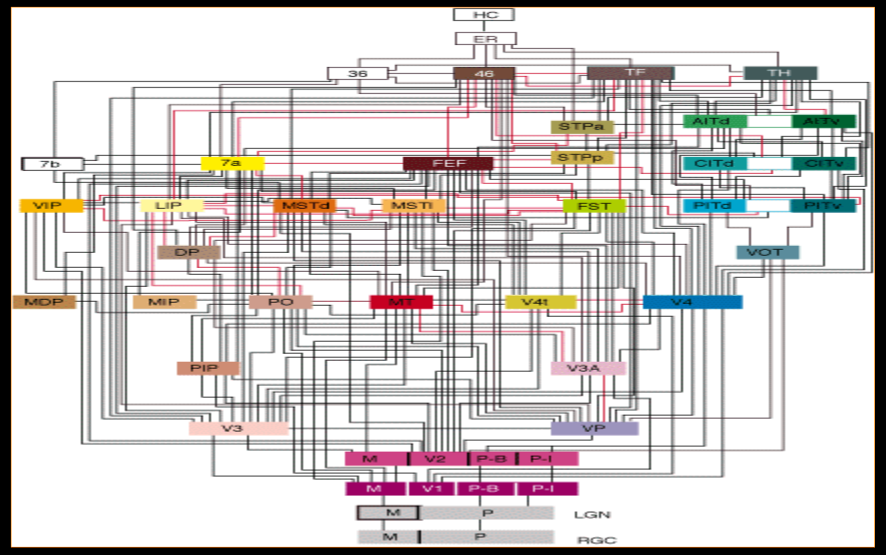
= 100 trillion connections

More potential circuits than molecules in the universe

"Bottom-Up" Approach

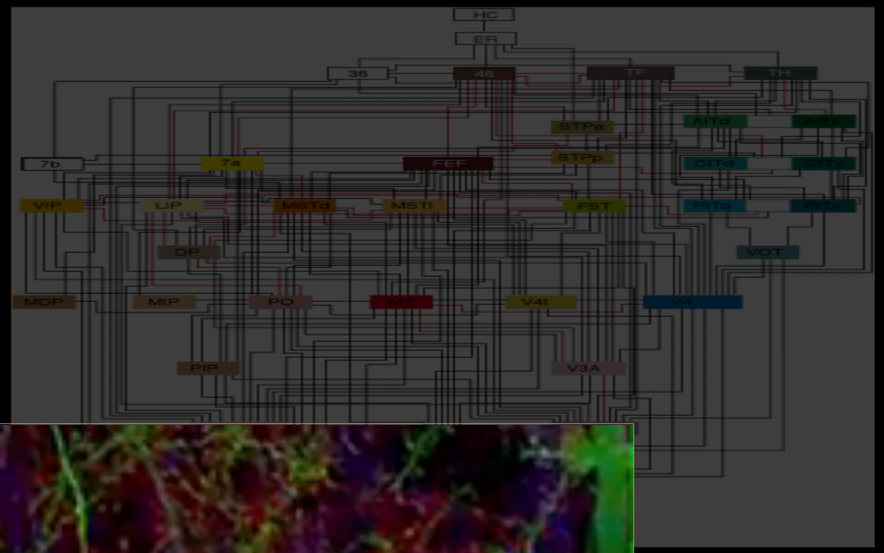
“Bottom-Up” Approach

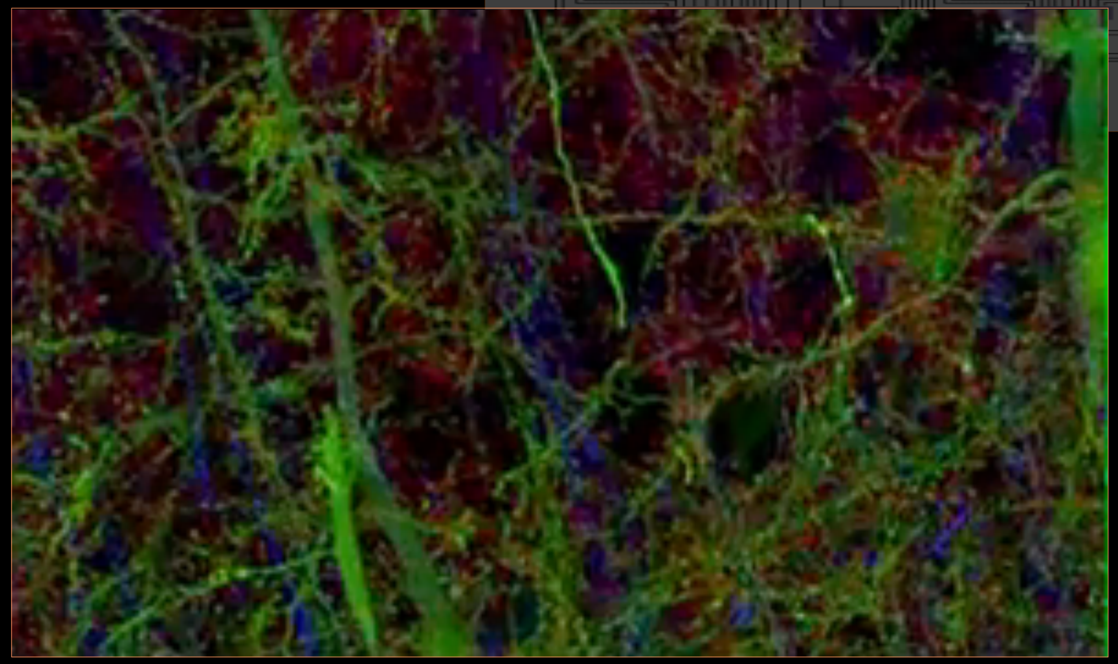
Go for entire wiring diagram…



"Bottom-Up" Approach

Go for entire wiring diagram…

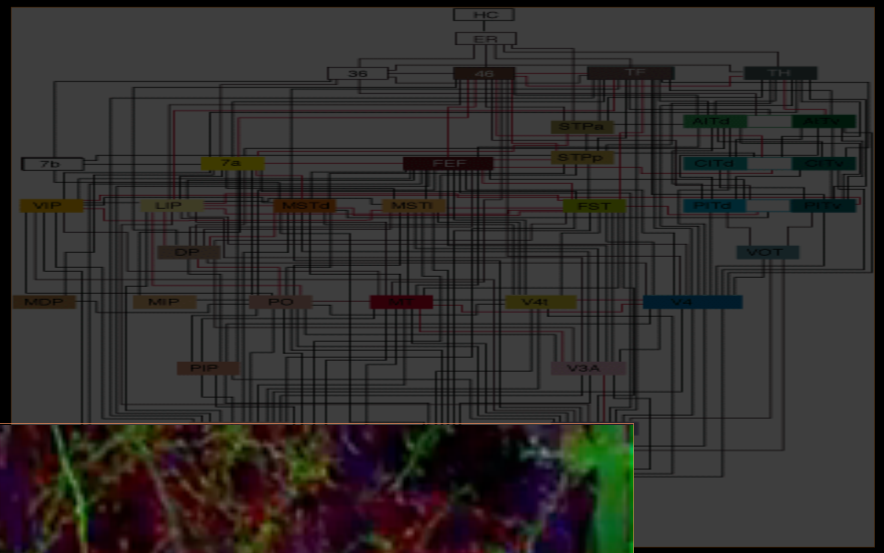


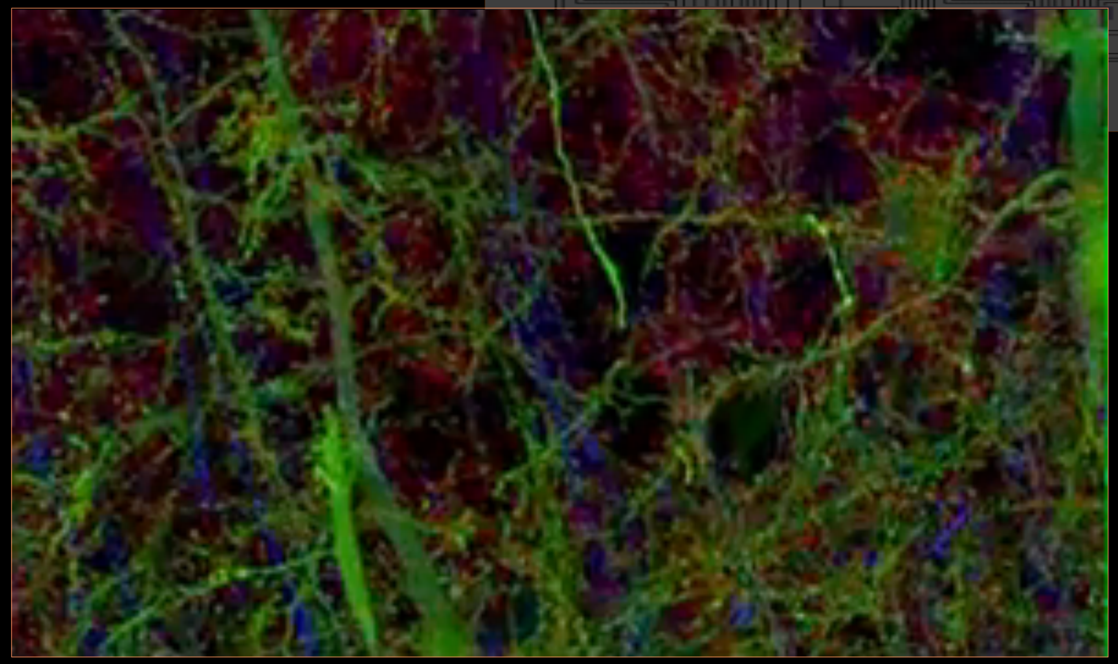


"Connectomics"
(map every connection)

“Bottom-Up” Approach

Go for entire wiring diagram…

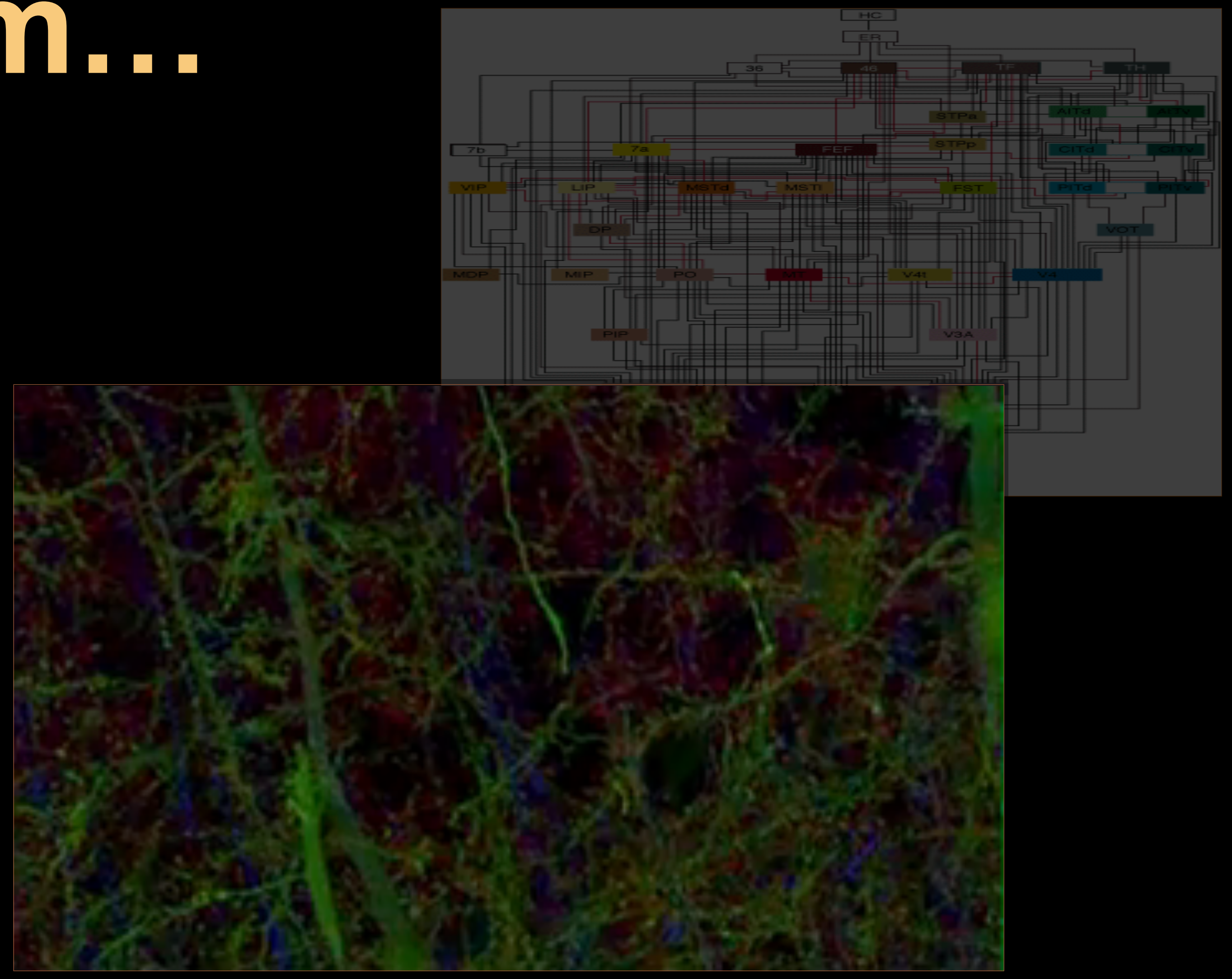




but...

“Bottom-Up” Approach

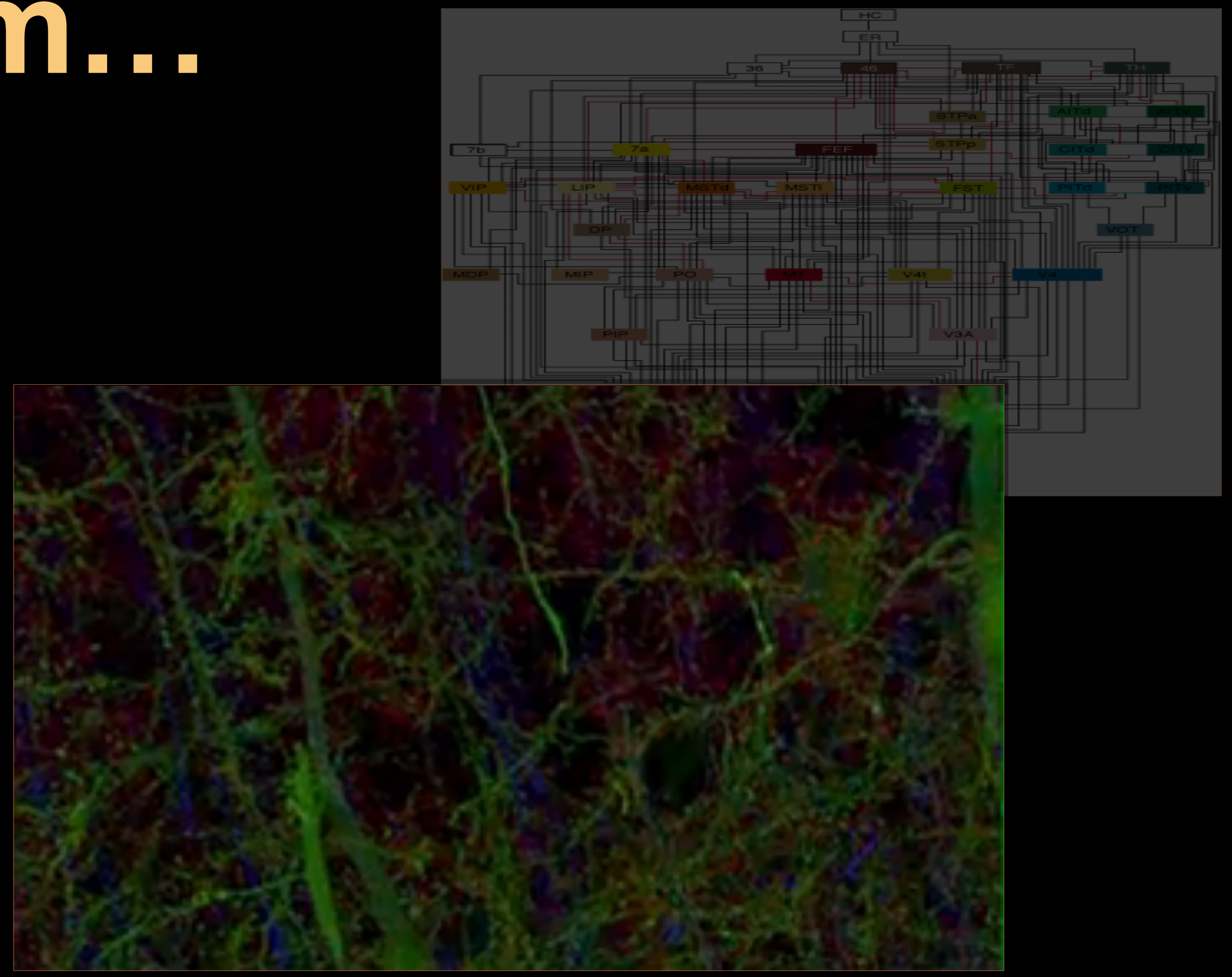
Go for entire wiring diagram…



The mouse brain alone
will take…

“Bottom-Up” Approach

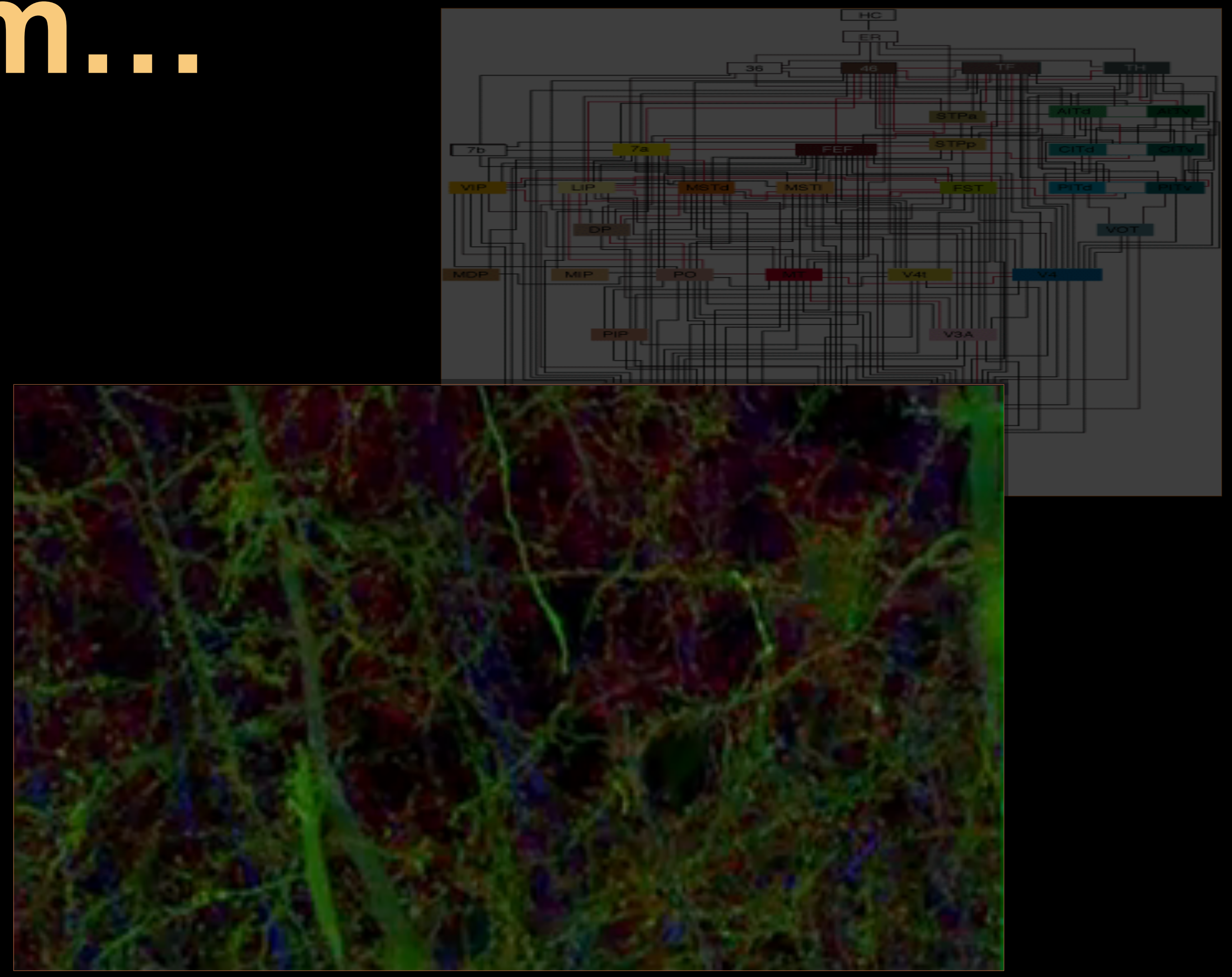
Go for entire wiring diagram…



???

“Bottom-Up” Approach

Go for entire wiring diagram…



And we’d only know structure, not *function*

The Connectionist (PDP) Approach

The Connectionist (PDP) Approach

- Brain-*like* computational *architecture*
  - biologically-inspired/plausible processing mechanisms

The Connectionist (PDP) Approach

- Brain-*like* computational *architecture*
  - biologically-inspired/plausible processing mechanisms
- Distributed representation
  - statistical structure / graded processing

The Connectionist (PDP) Approach

- Brain-*like* computational *architecture*
  - biologically-inspired/plausible processing mechanisms
- Distributed representation
  - statistical structure / graded processing
- Parallel processing
  - can meet 100 step challenge

The Connectionist (PDP) Approach

- Brain-*like* computational *architecture*
  – biologically-inspired/plausible processing mechanisms
- Distributed representation
  – statistical structure / graded processing
- Parallel processing
  – can meet 100 step challenge
- Self-organization through experience
  – general purpose statistical ***learning*** algorithms

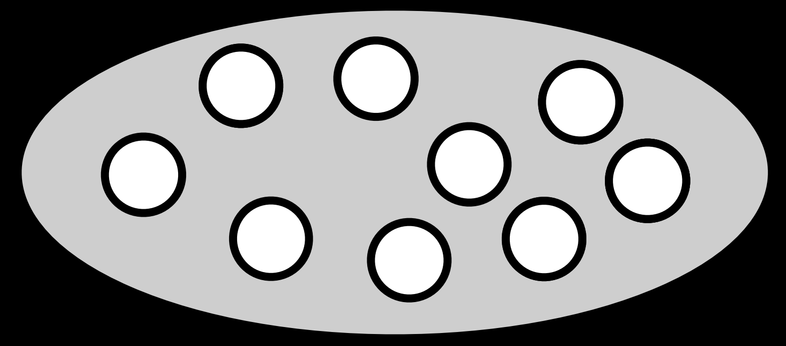
The Connectionist (PDP) Approach

- Brain-*like* computational *architecture*
  – biologically-inspired/plausible processing mechanisms
- Distributed representation
  – statistical structure / graded processing
- Parallel processing
  – can meet 100 step challenge
- Self-organization through experience
  – general purpose statistical *learning* algorithms

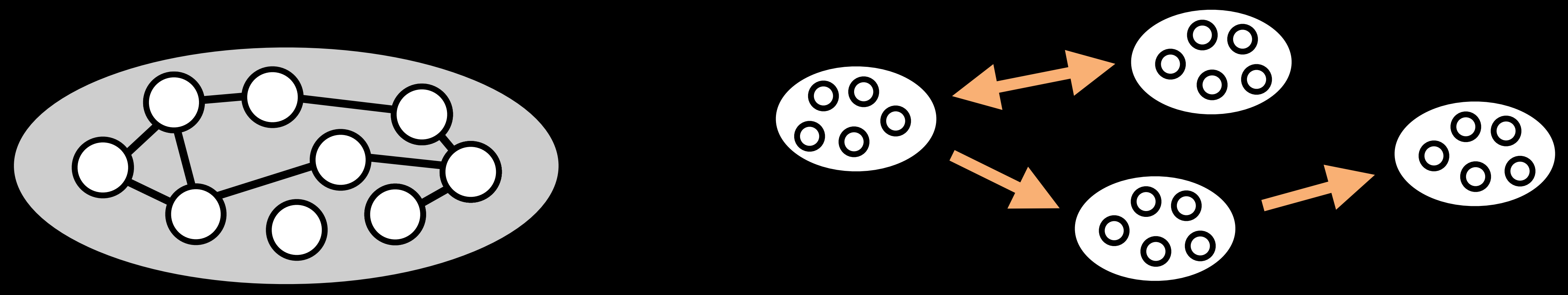
Skip PDP intro

Basic Elements

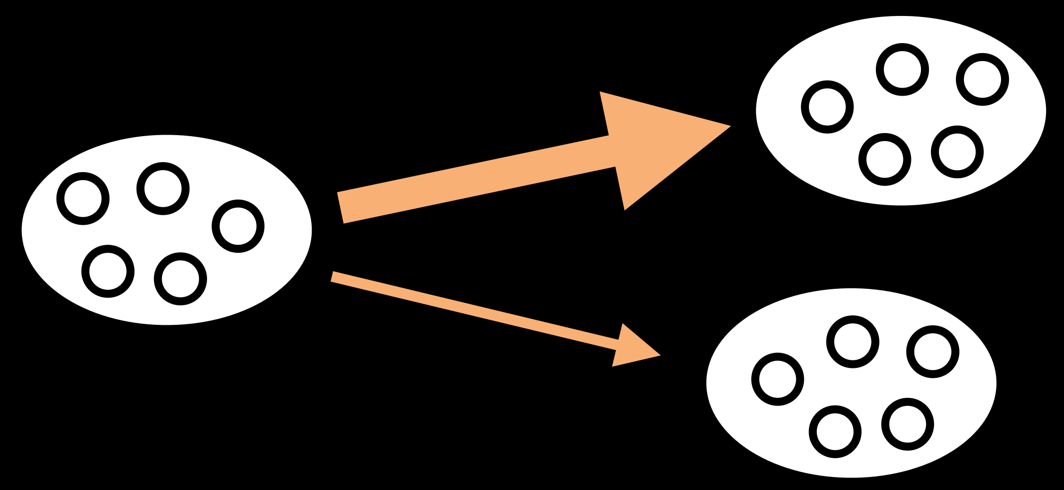
- **Units / Modules** (≈ neuron or population of neurons)



- **Connections / Pathways** (≈ synapses / projections / circuits)

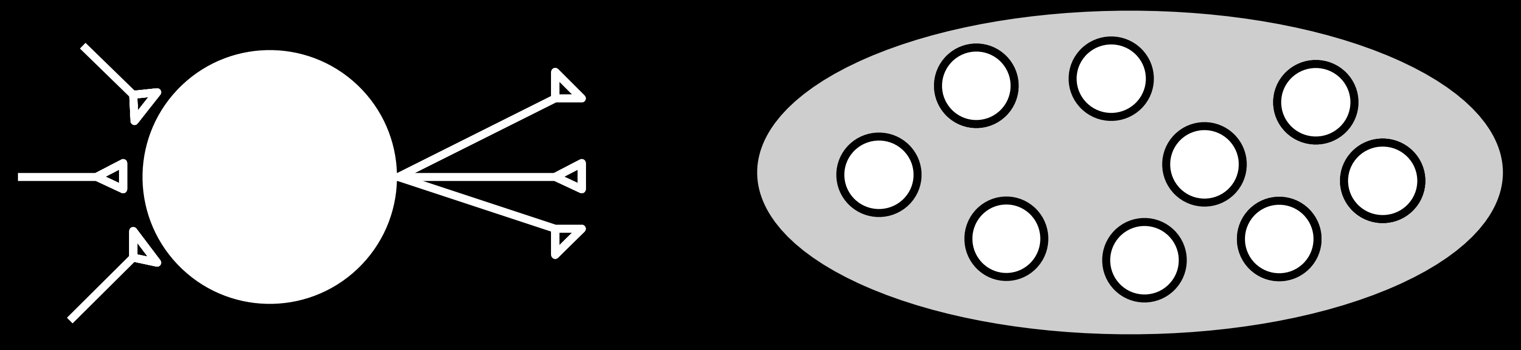


- **Learning Rules** (≈ synaptic plasticity)

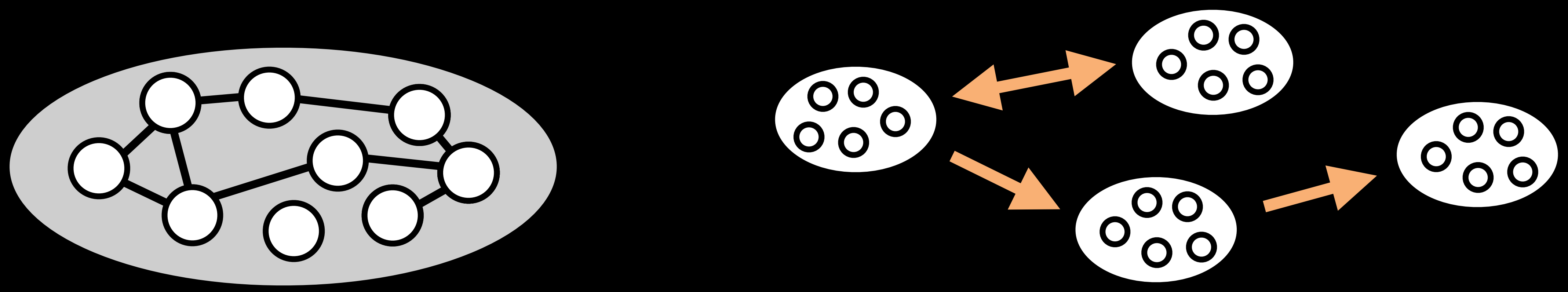


Basic Elements

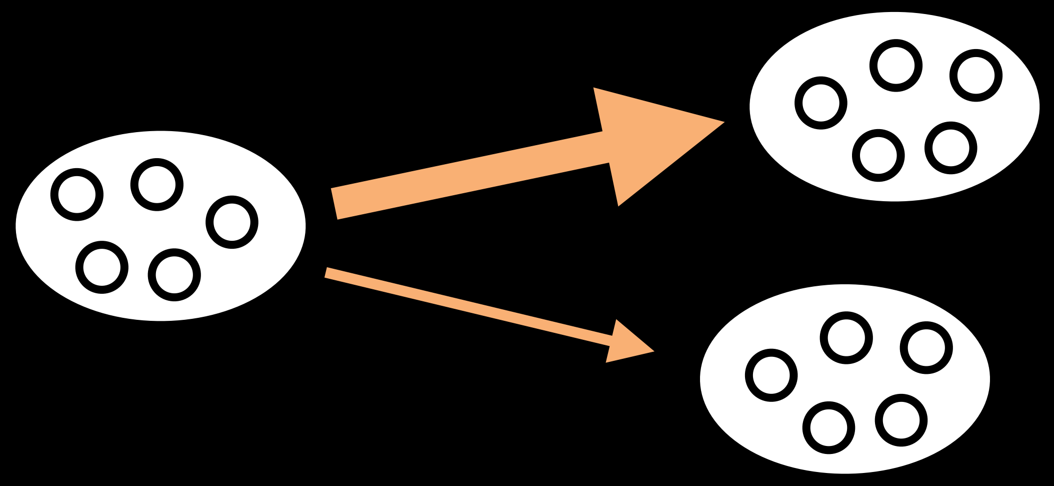
- **Units / Modules** (≈ neuron or population of neurons)



- **Connections / Pathways** (≈ synapses / projections / circuits)

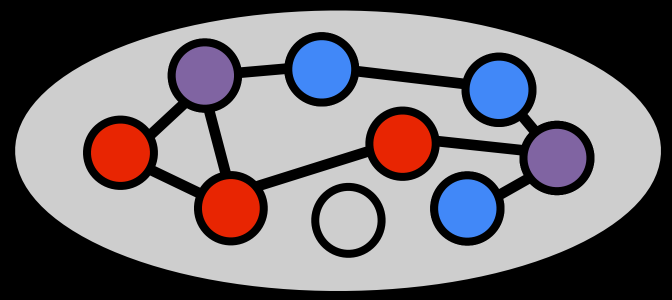


- **Learning Rules** (≈ synaptic plasticity)



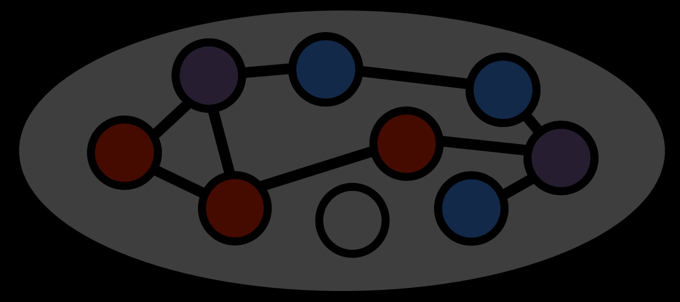
Psychological Constructs

Psychological Constructs

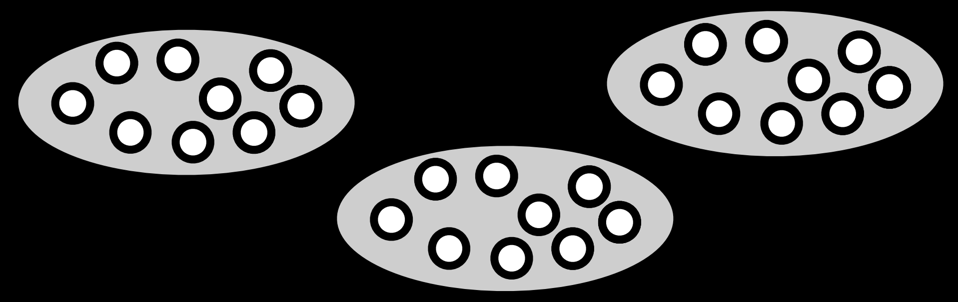
- **Representation** (units) 

Psychological Constructs

- **Representation** (units)

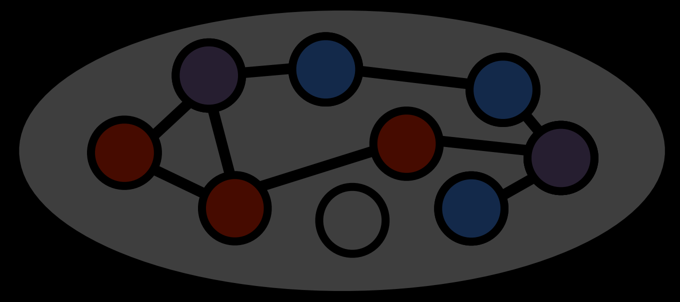


- **Functions** (modules)

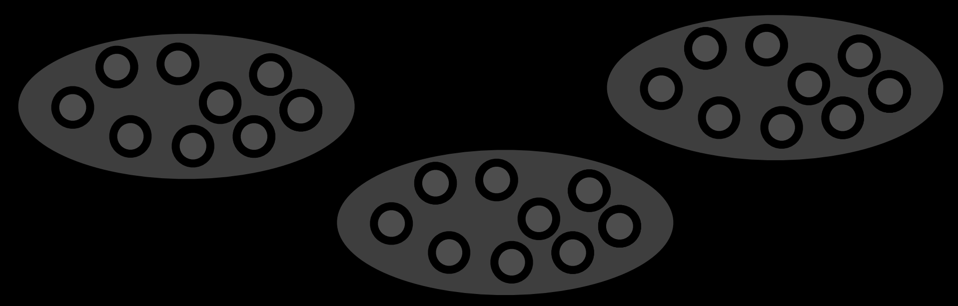


Psychological Constructs

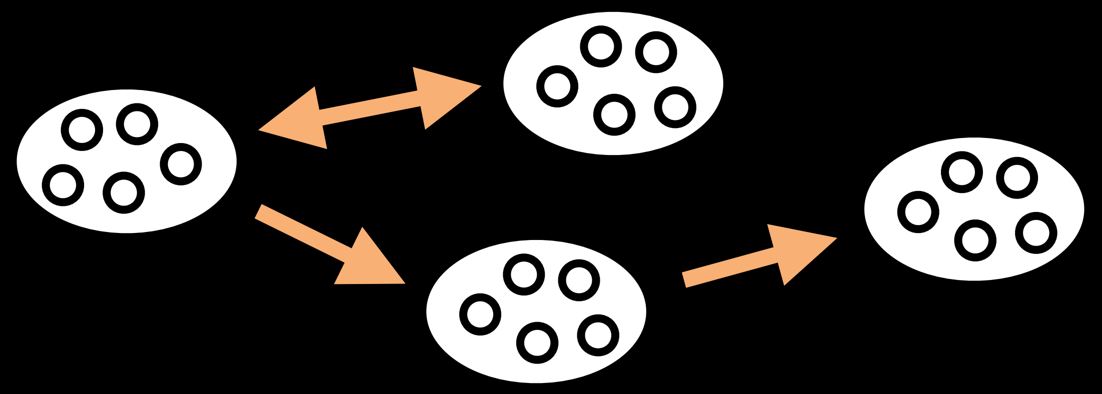
- **Representation** (units)



- **Functions** (modules)

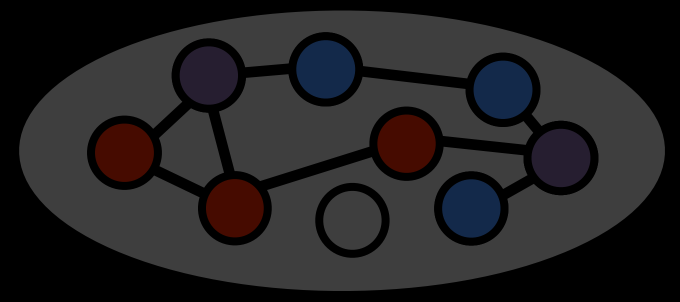


- **Processing** (flow of activity)

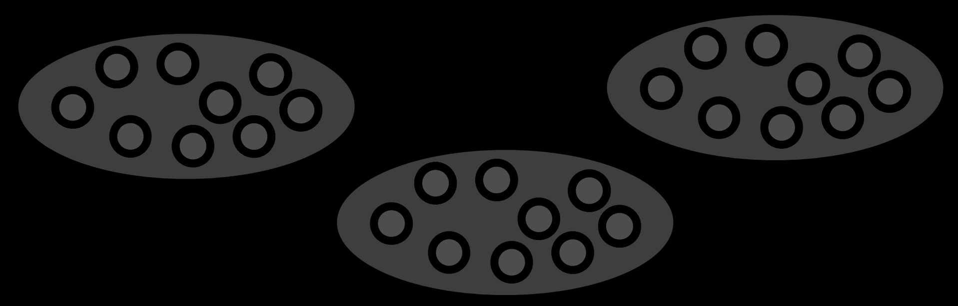


Psychological Constructs

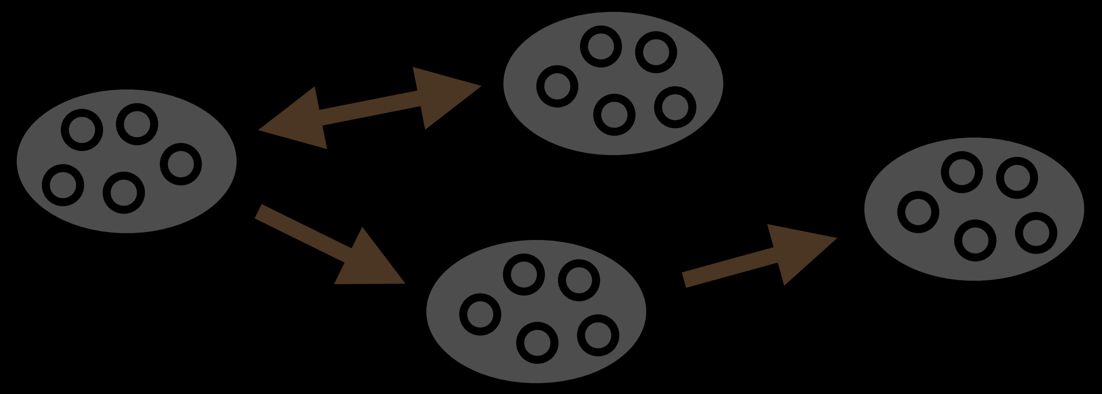
- **Representation** (units)



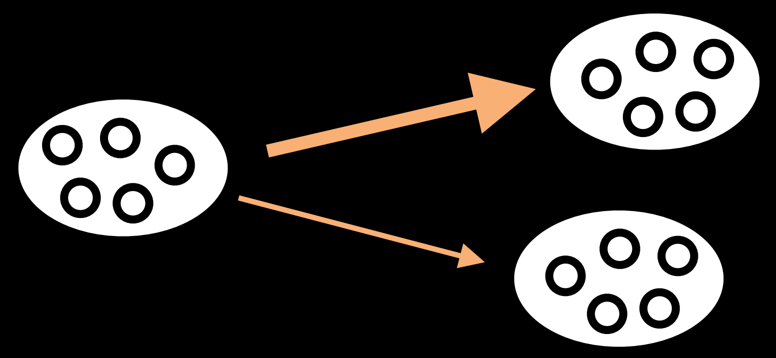
- **Functions** (modules)



- **Processing** (flow of activity)

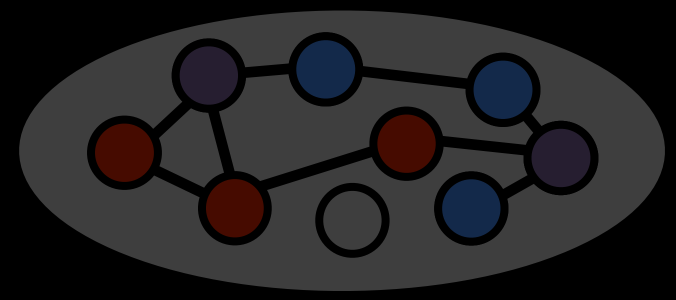


- **Learning** (weight modification)

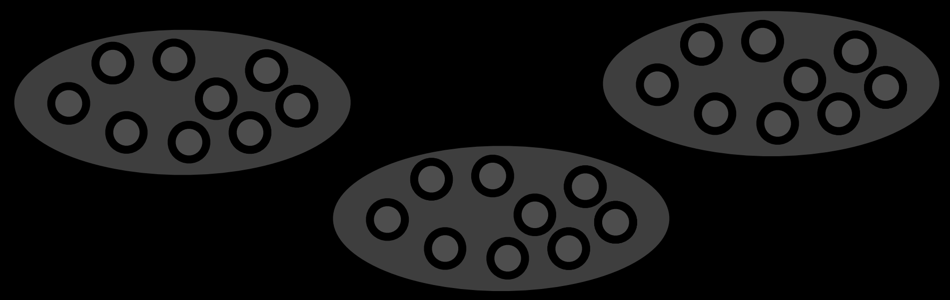


Psychological Constructs

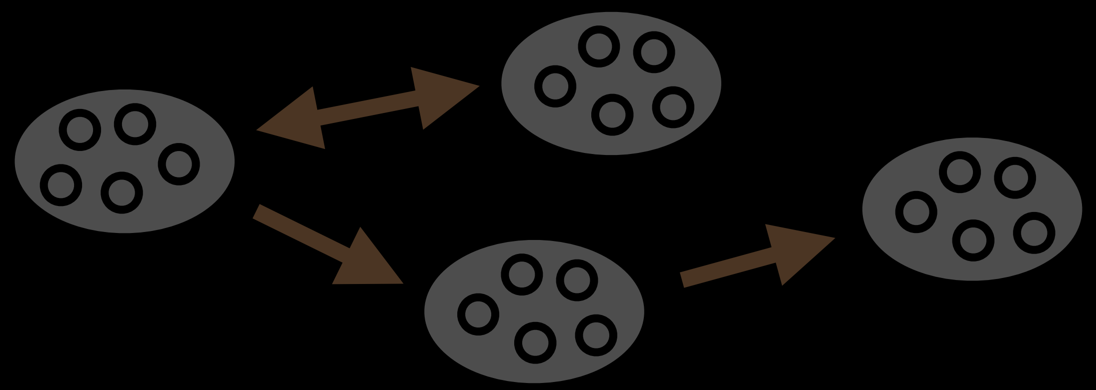
- **Representation** (units)



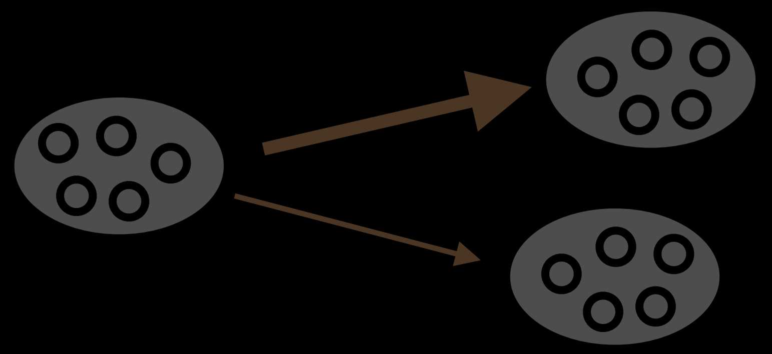
- **Functions** (modules)



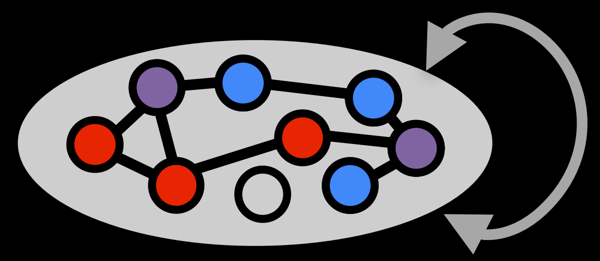
- **Processing** (flow of activity)



- **Learning** (weight modification)

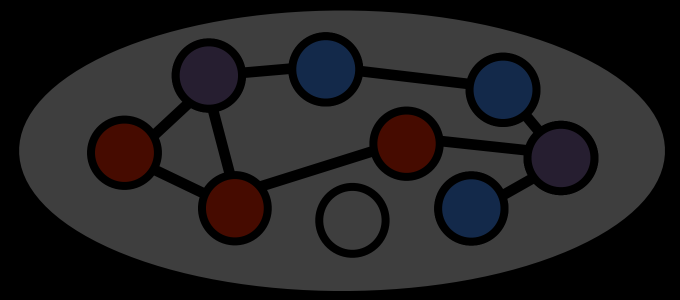


- **Memory** (active maintenance

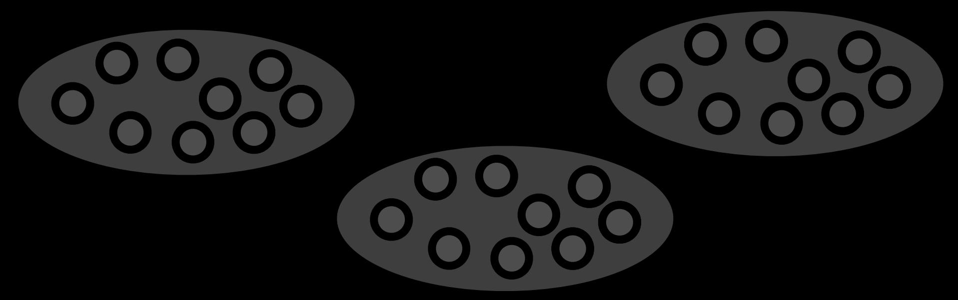


Psychological Constructs

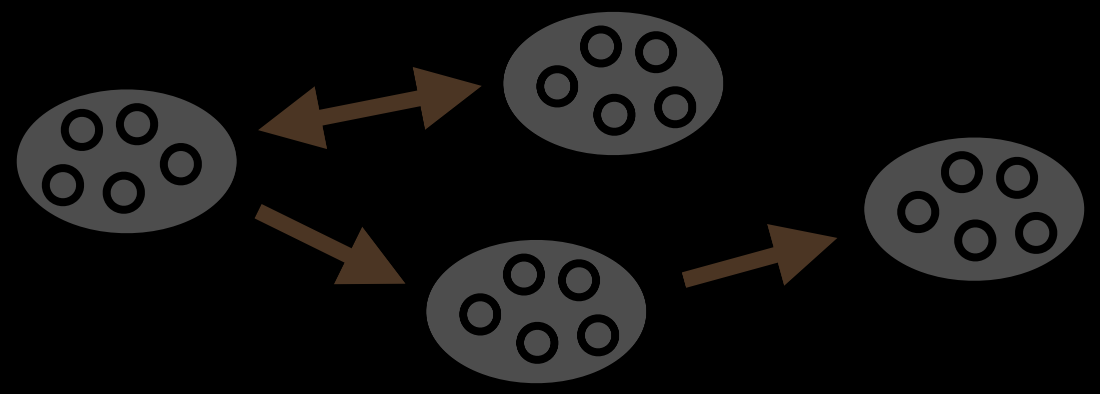
- **Representation** (units)



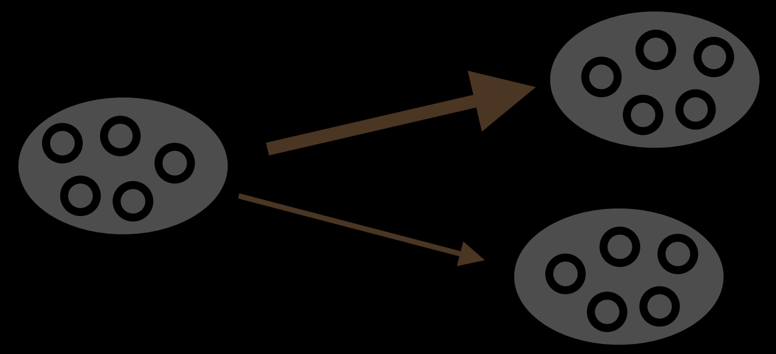
- **Functions** (modules)



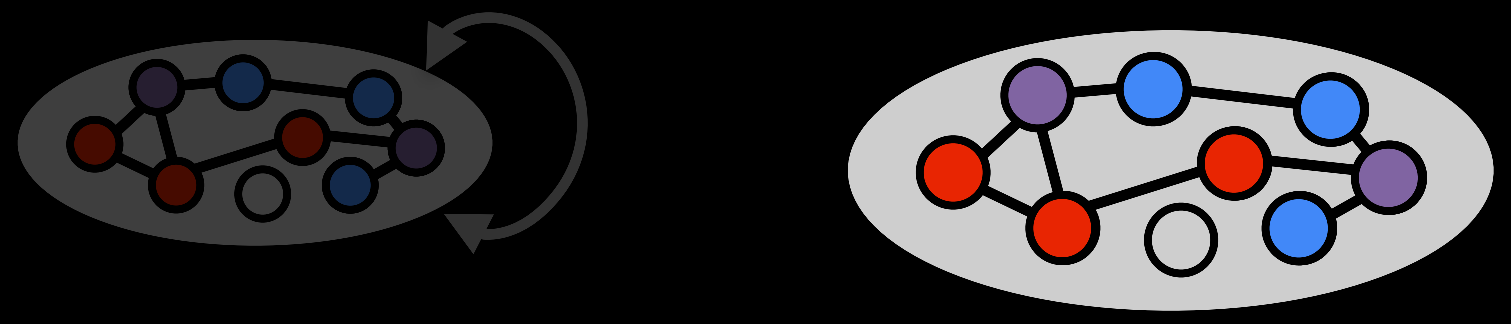
- **Processing** (flow of activity)



- **Learning** (weight modification)

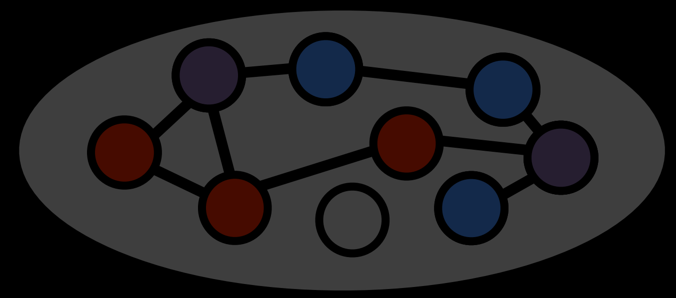


- **Memory** (active maintenance, pattern completion

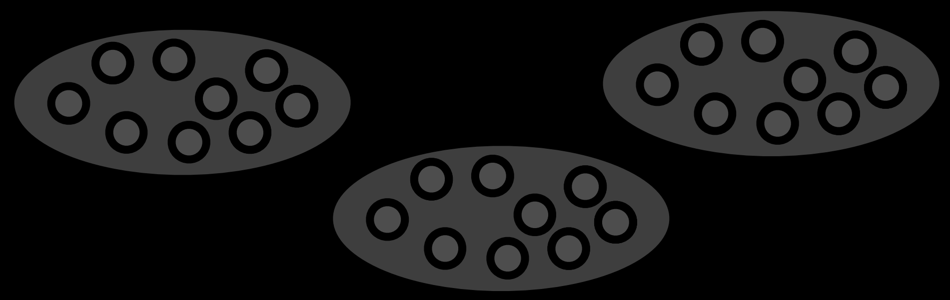


Psychological Constructs

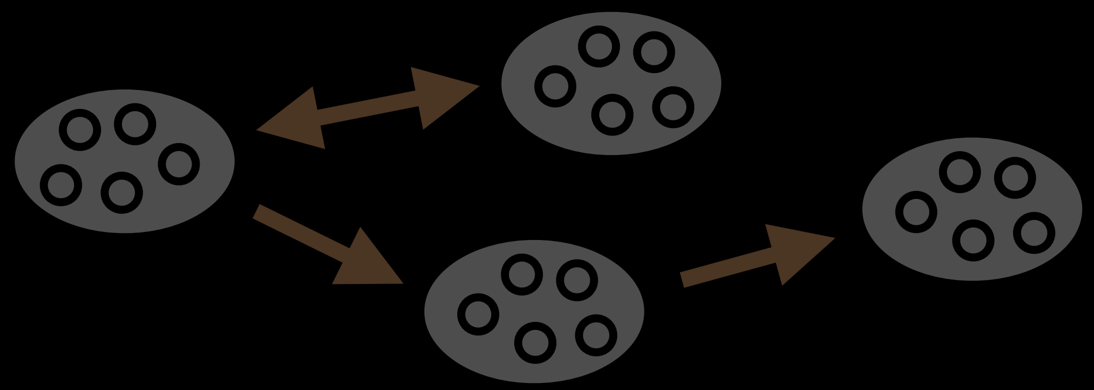
- **Representation** (units)



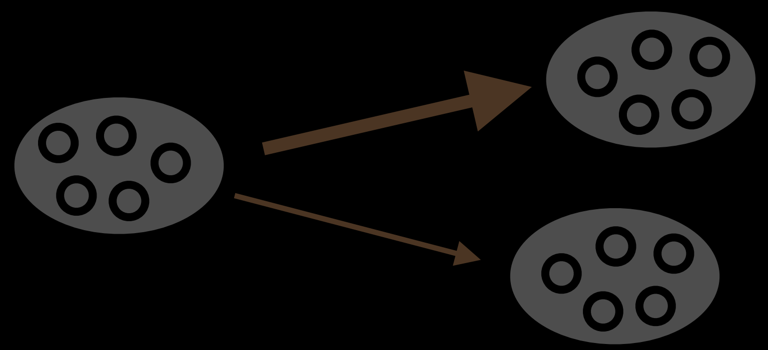
- **Functions** (modules)



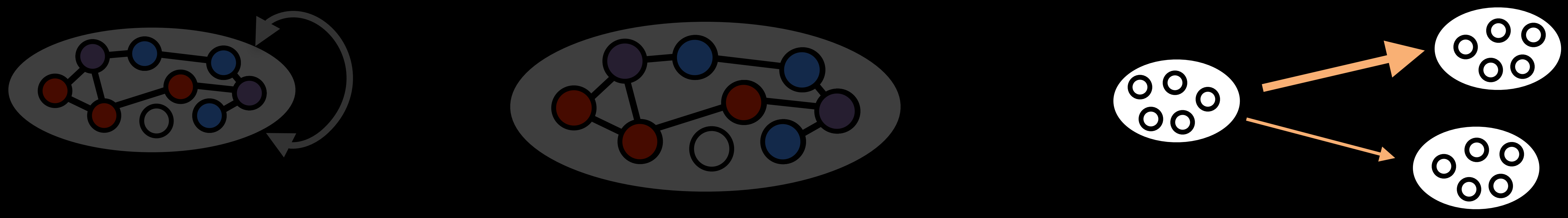
- **Processing** (flow of activity)



- **Learning** (weight modification)



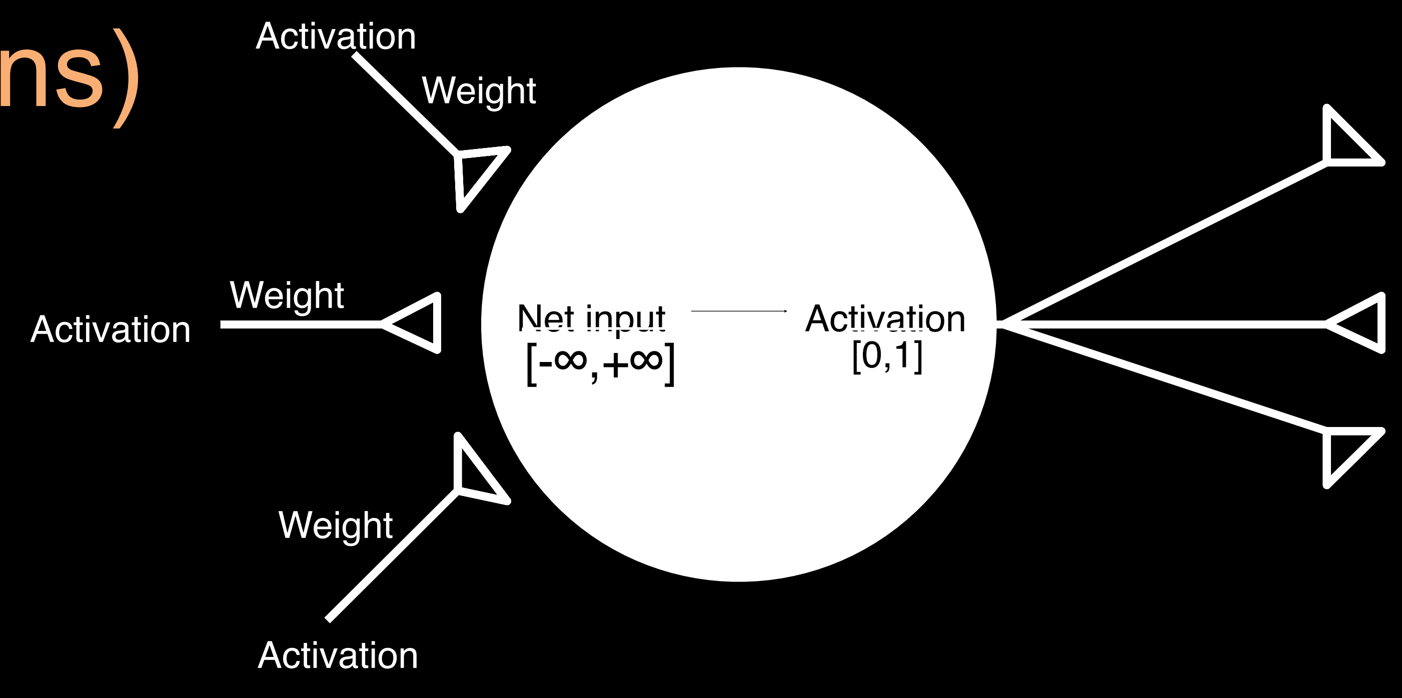
- **Memory** (active maintenance, pattern completion, or weight modification)



Representation

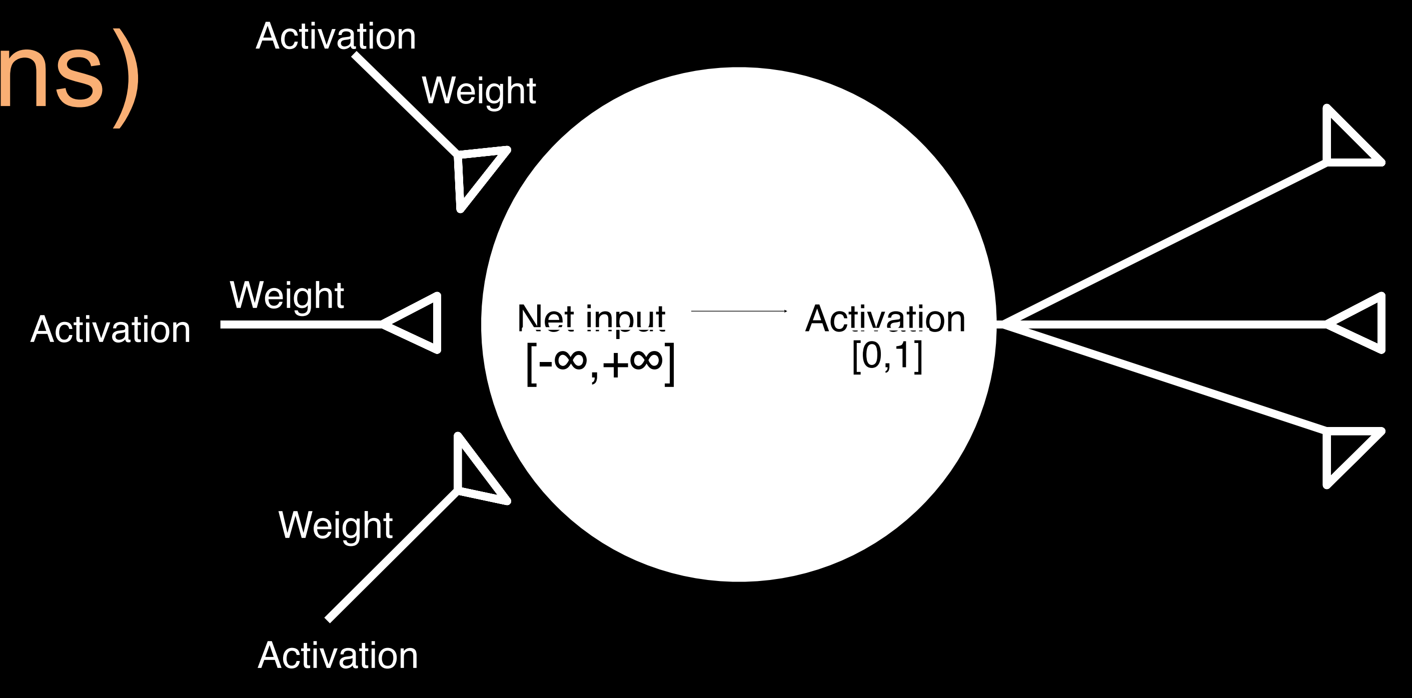
Representation

- **Units** (≈ neurons or population of neurons)



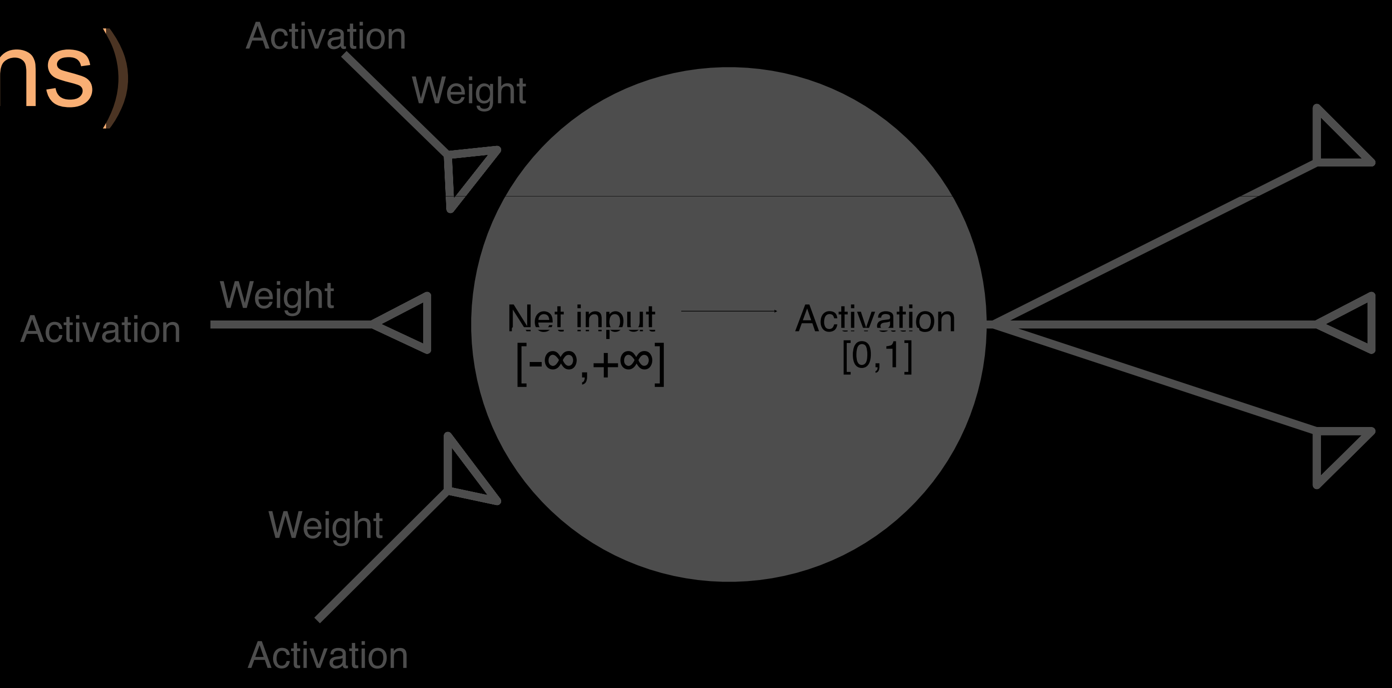
Representation

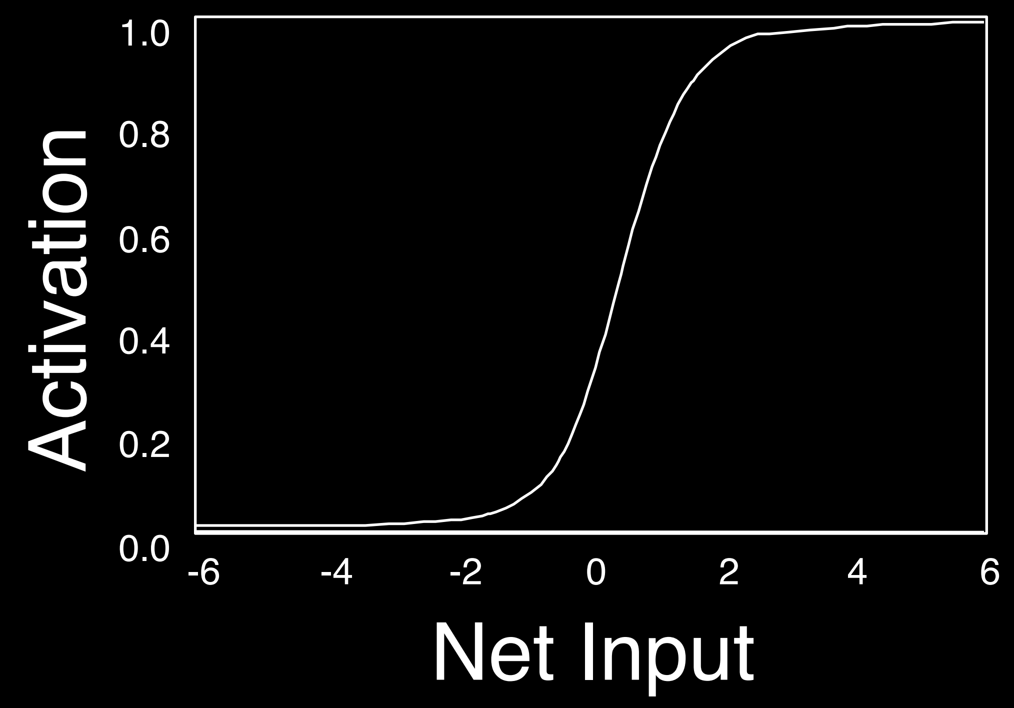
- **Units** (≈ neurons or population of neurons)
  - **Activity level**
    (≈ firing frequency or probability of firing)



Representation

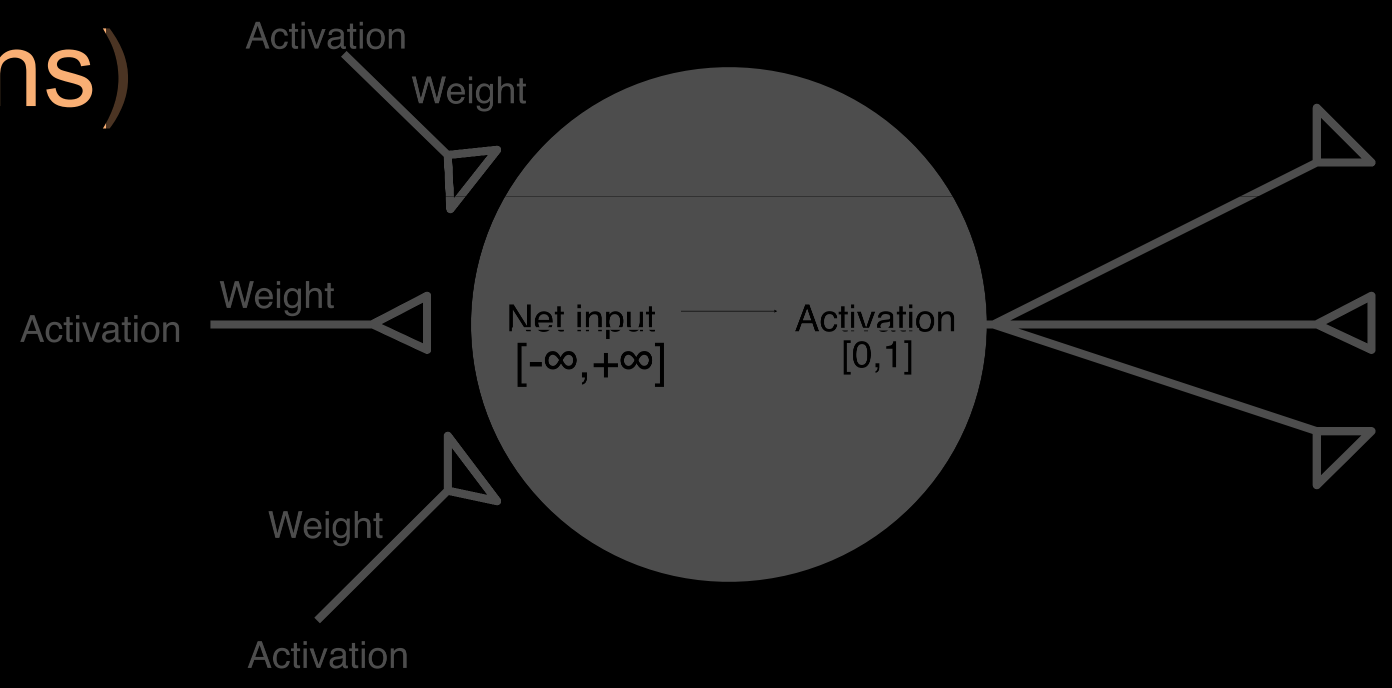
- **Units** (≈ neurons or population of neurons)
  - **Activity level**
    (≈ firing frequency or probability of firing)
  - **Activation (transfer) function**
    - **Integrate & fire**
    - **Thresholded (piecewise) linear**
    - **Continuous valued (sigmoid function, e.g. logistic)**

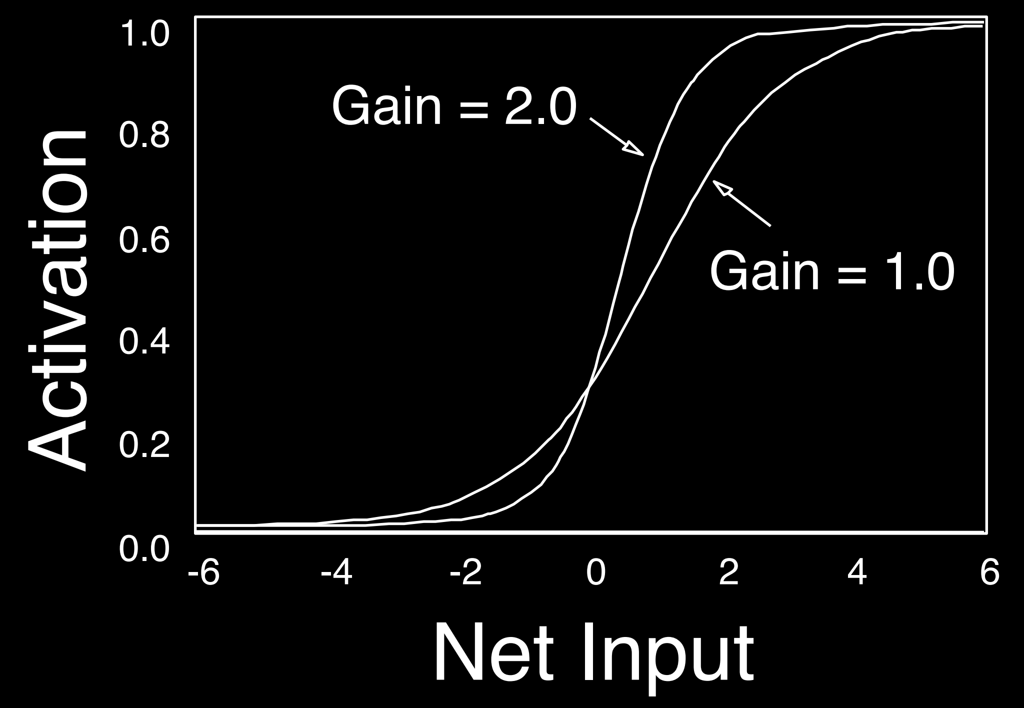




Representation

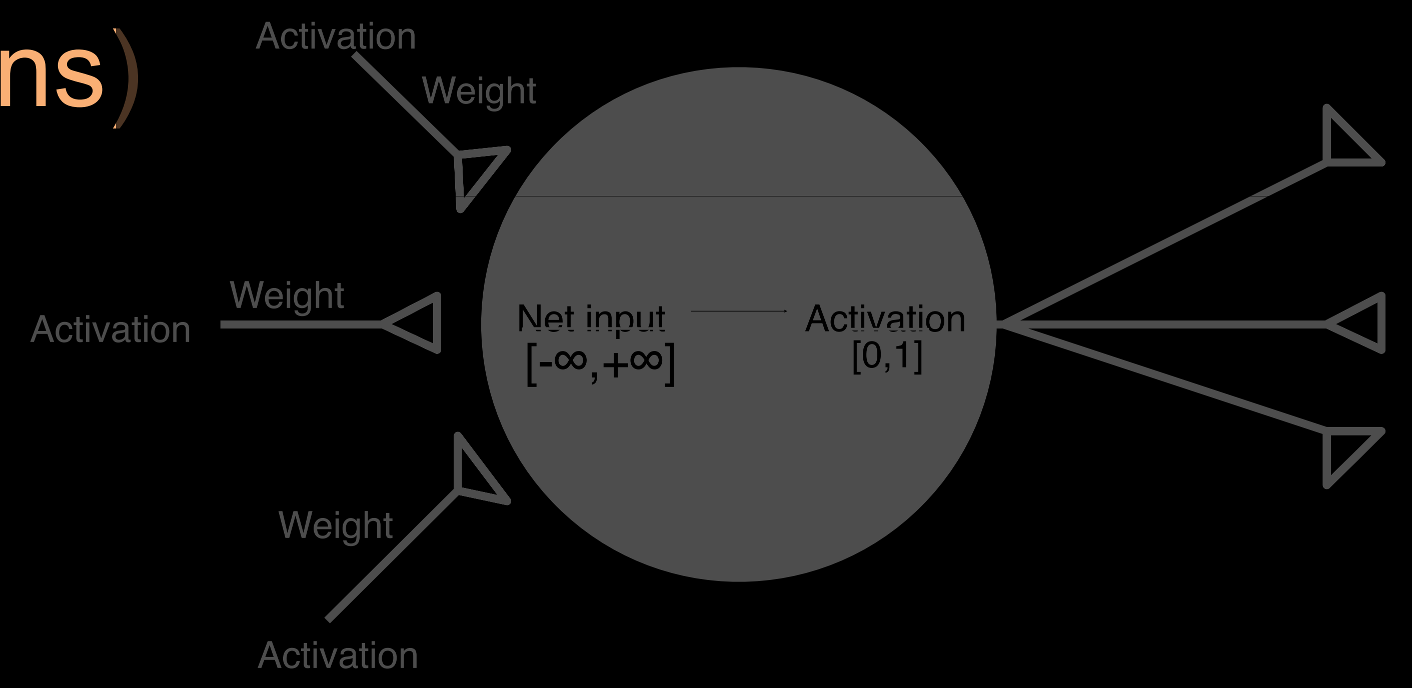
- **Units** (≈ neurons or population of neurons)
  - **Activity level**
    (≈ firing frequency or probability of firing)
  - **Activation (transfer) function**
    - **Integrate & fire**
    - **Thresholded (piecewise) linear**
    - **Continuous valued (sigmoid function, e.g. logistic)**
  - **Noise**
  - **Modulation**

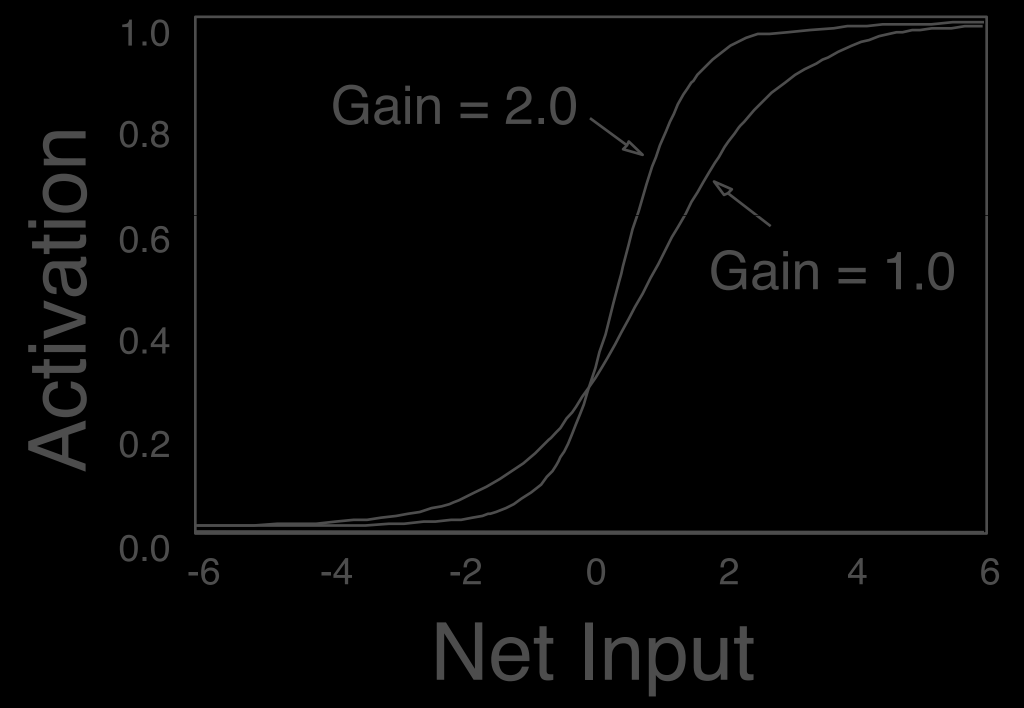




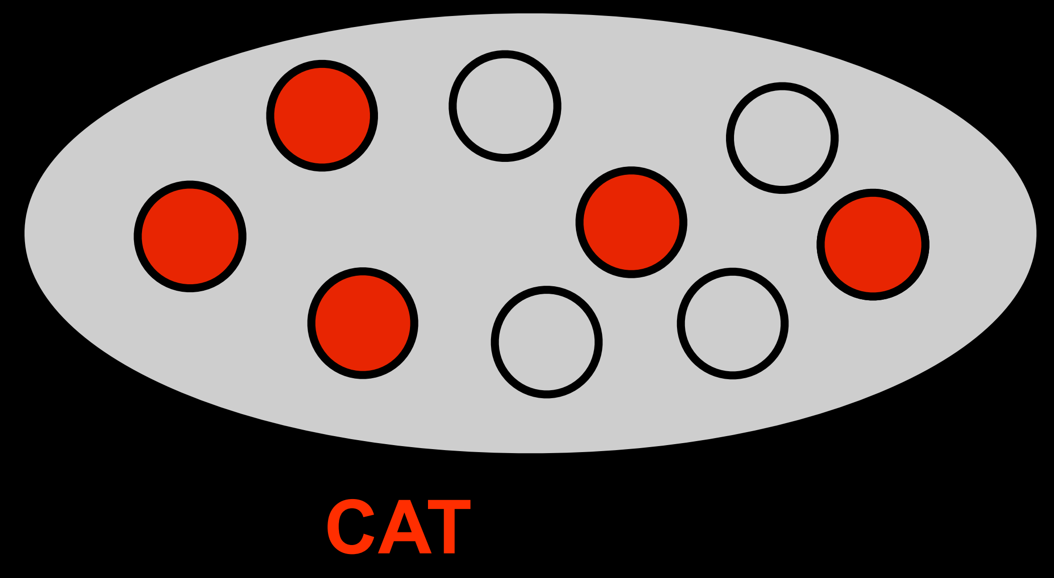
Representation

- **Units** (≈ neurons or population of neurons)
  - **Activity level**
    (≈ firing frequency or probability of firing)
  - **Activation (transfer) function**
    - **Integrate & fire**
    - **Thresholded (piecewise) linear**
    - **Continuous valued (sigmoid function, e.g. logistic)**
  - **Noise**
  - **Modulation**



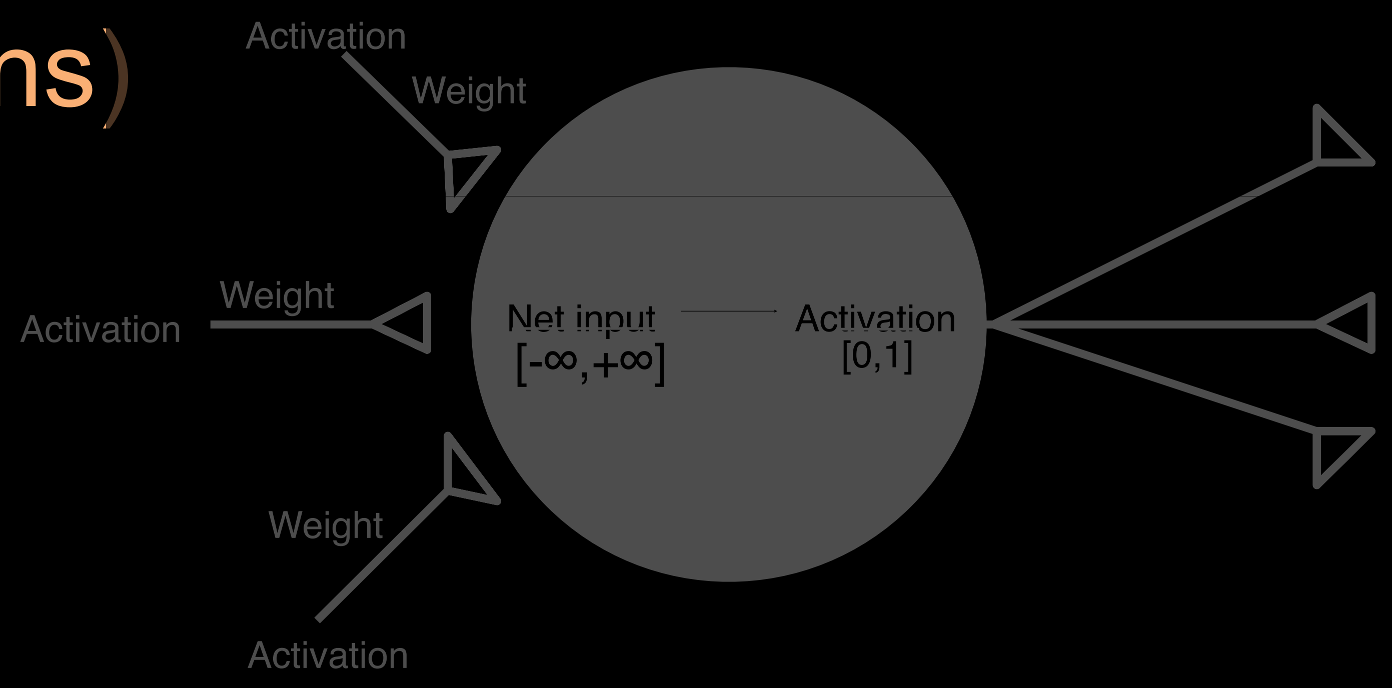


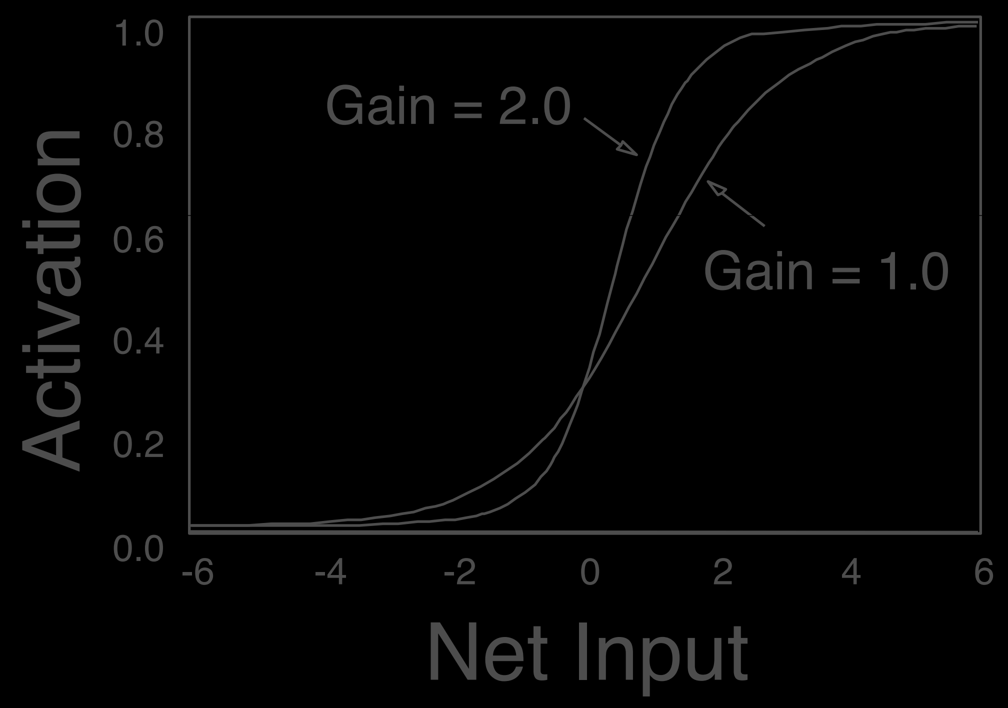
- **Patterns of activity** (≈ population code)
  - **Distributed representation**



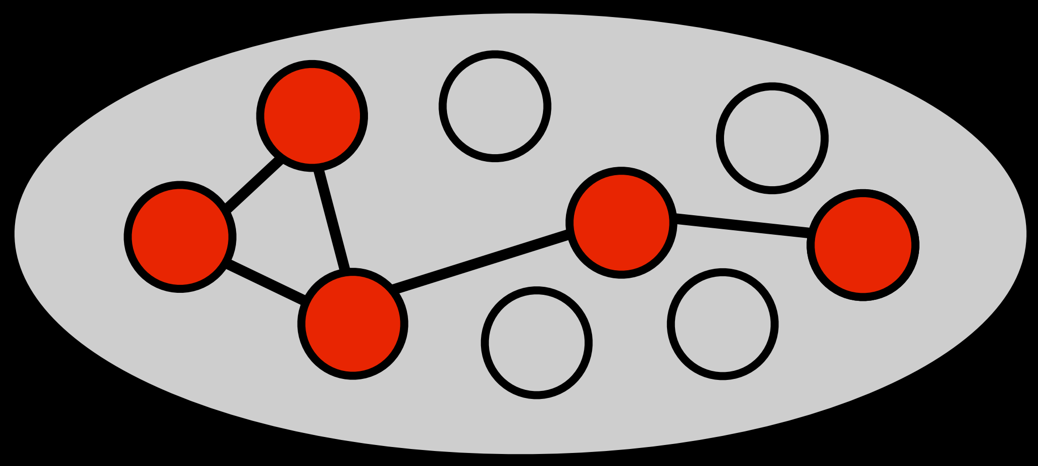
Representation

- **Units** (≈ neurons or population of neurons)
  - **Activity level**
    (≈ firing frequency or probability of firing)
  - **Activation (transfer) function**
    - **Integrate & fire**
    - **Thresholded (piecewise) linear**
    - **Continuous valued (sigmoid function, e.g. logistic)**
  - **Noise**
  - **Modulation**





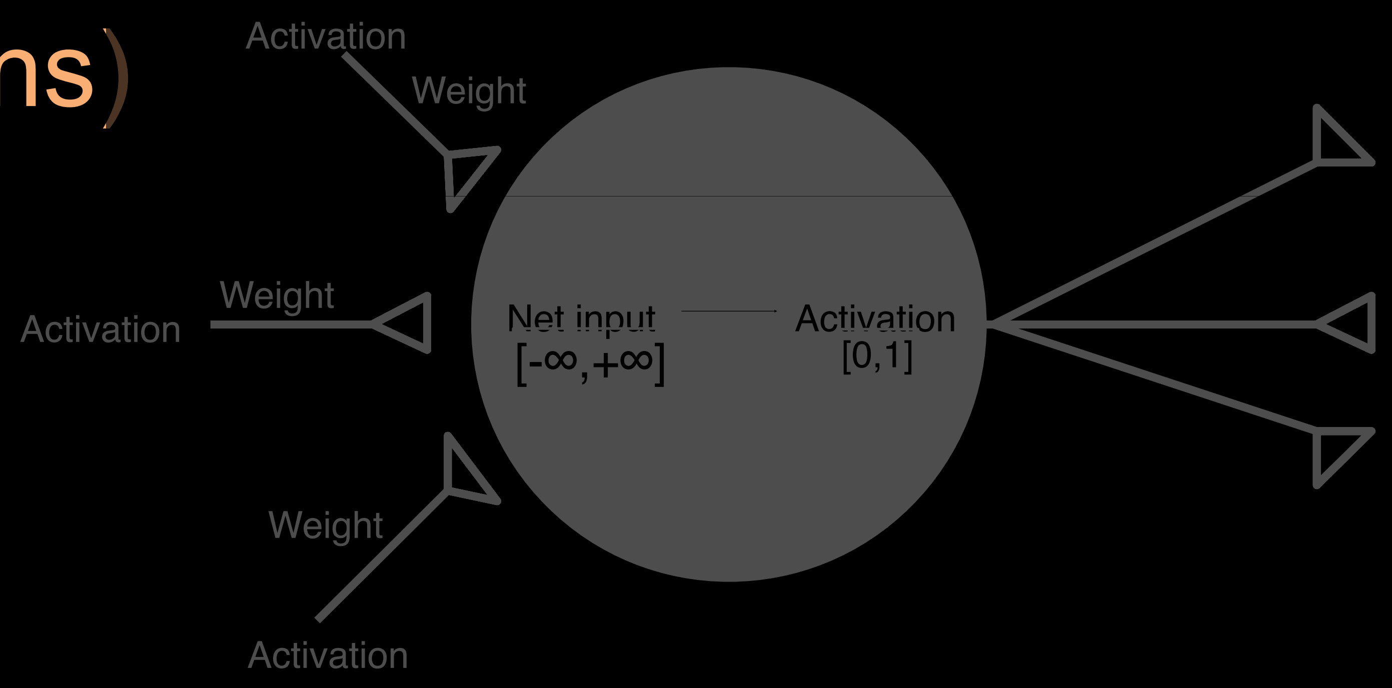
- **Patterns of activity** (≈ population code)
  - **Distributed representation**
  - **Relationships:**
    - associations between units/patterns

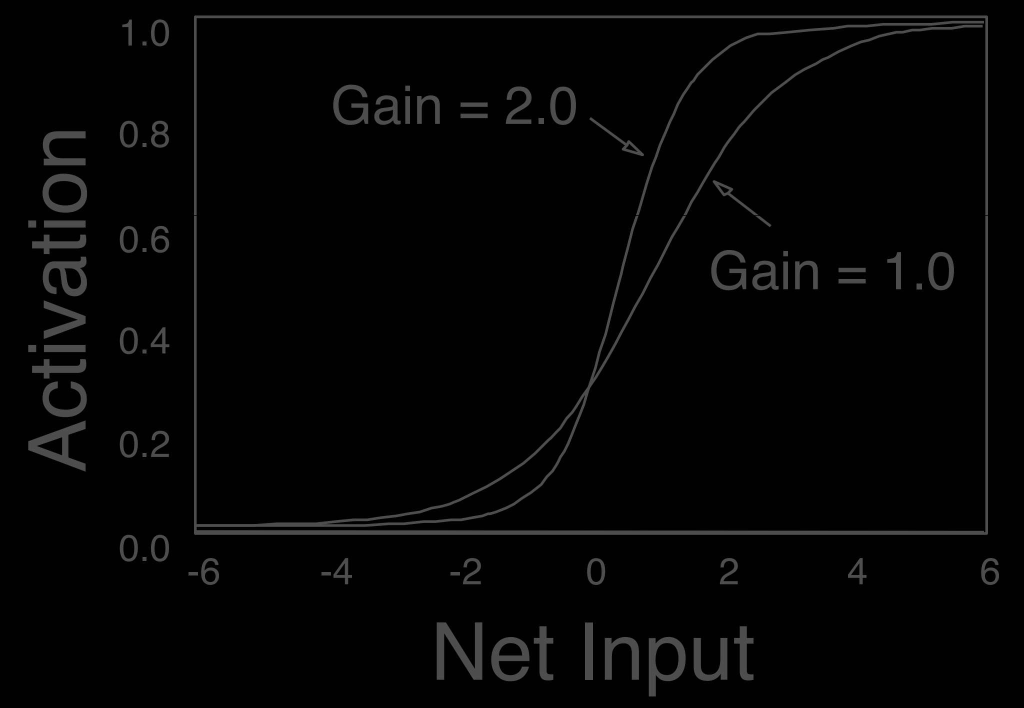


CAT

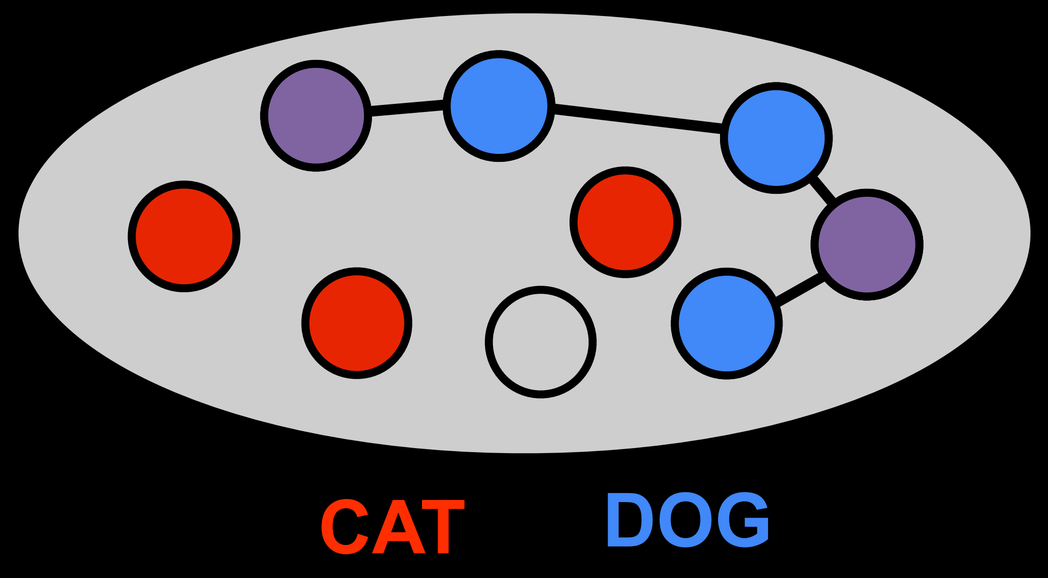
Representation

- **Units** (≈ neurons or population of neurons)
  - **Activity level**
    (≈ firing frequency or probability of firing)
  - **Activation (transfer) function**
    - **Integrate & fire**
    - **Thresholded (piecewise) linear**
    - **Continuous valued (sigmoid function, e.g. logistic)**
  - **Noise**
  - **Modulation**





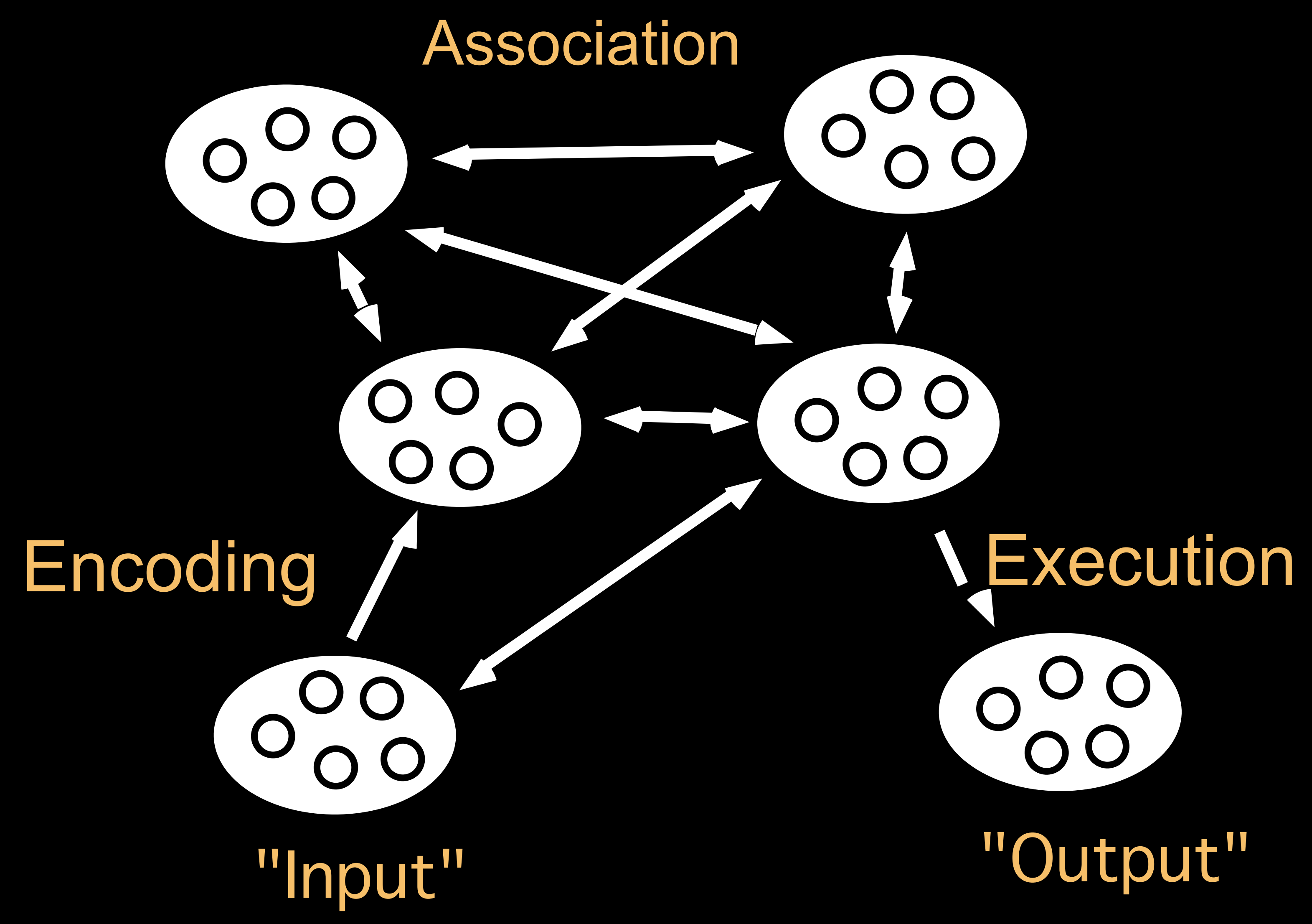
- **Patterns of activity** (≈ population code)
  - **Distributed representation**
  - **Relationships:**
    - associations between units/patterns
    - overlap of patterns



Specialization

- **Modules** (≈ brain areas)

sets of units responsible for:

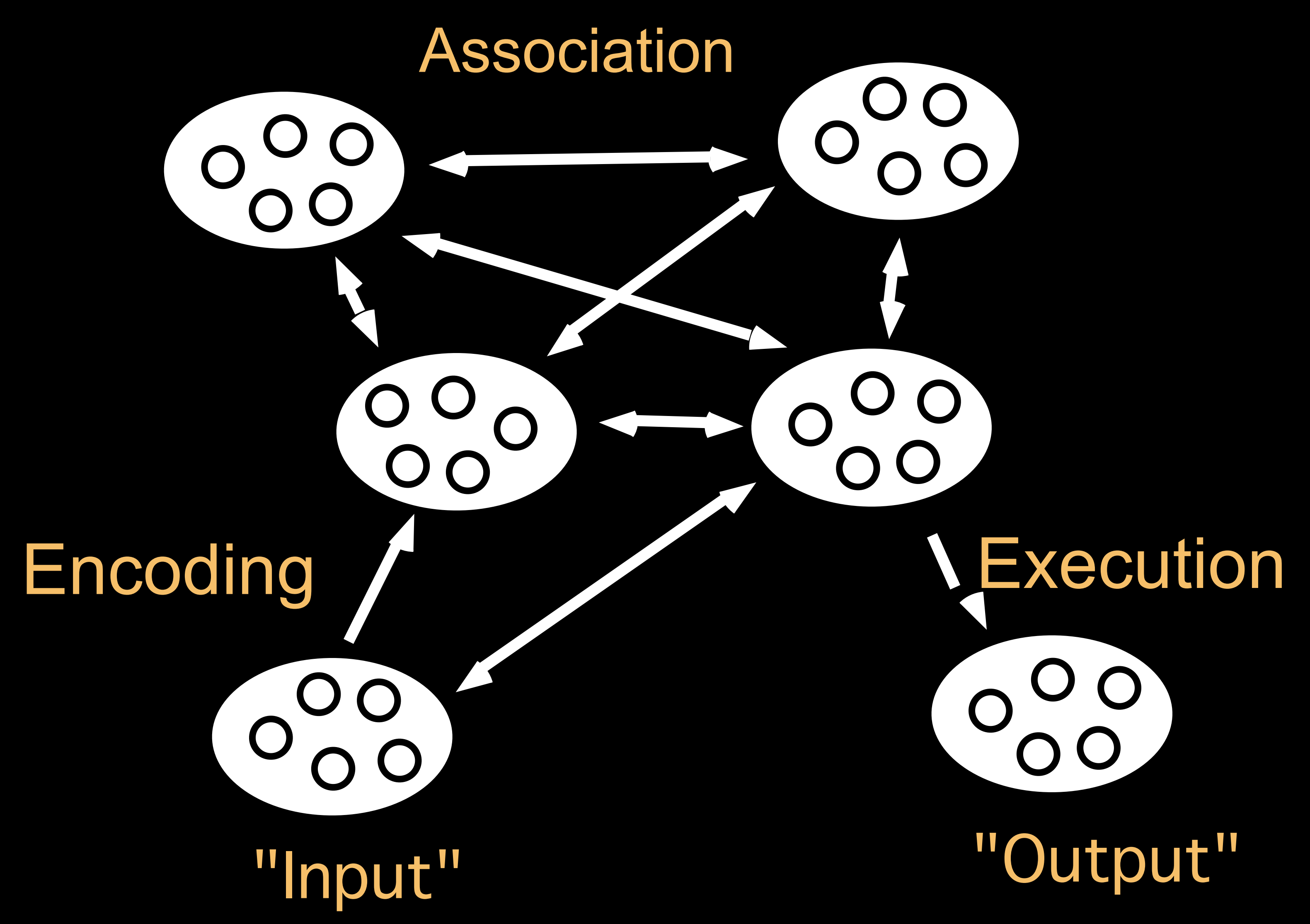


Specialization

- **Modules** (≈ brain areas)

sets of units responsible for:

representing a particular type of information
stimulus (input), semantic (hidden), motor (output) etc.



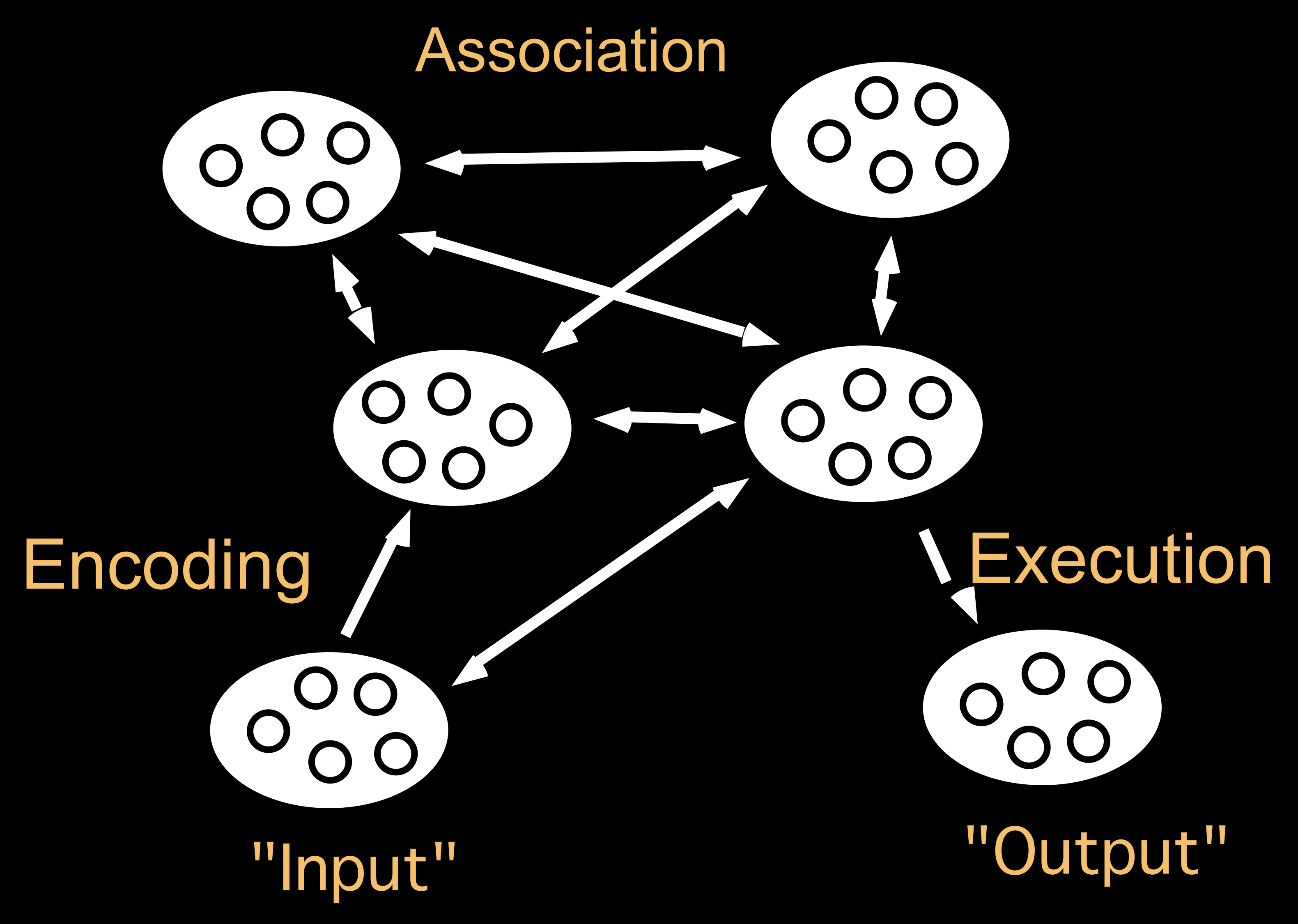
Specialization

- **Modules** (≈ brain areas)

sets of units responsible for:

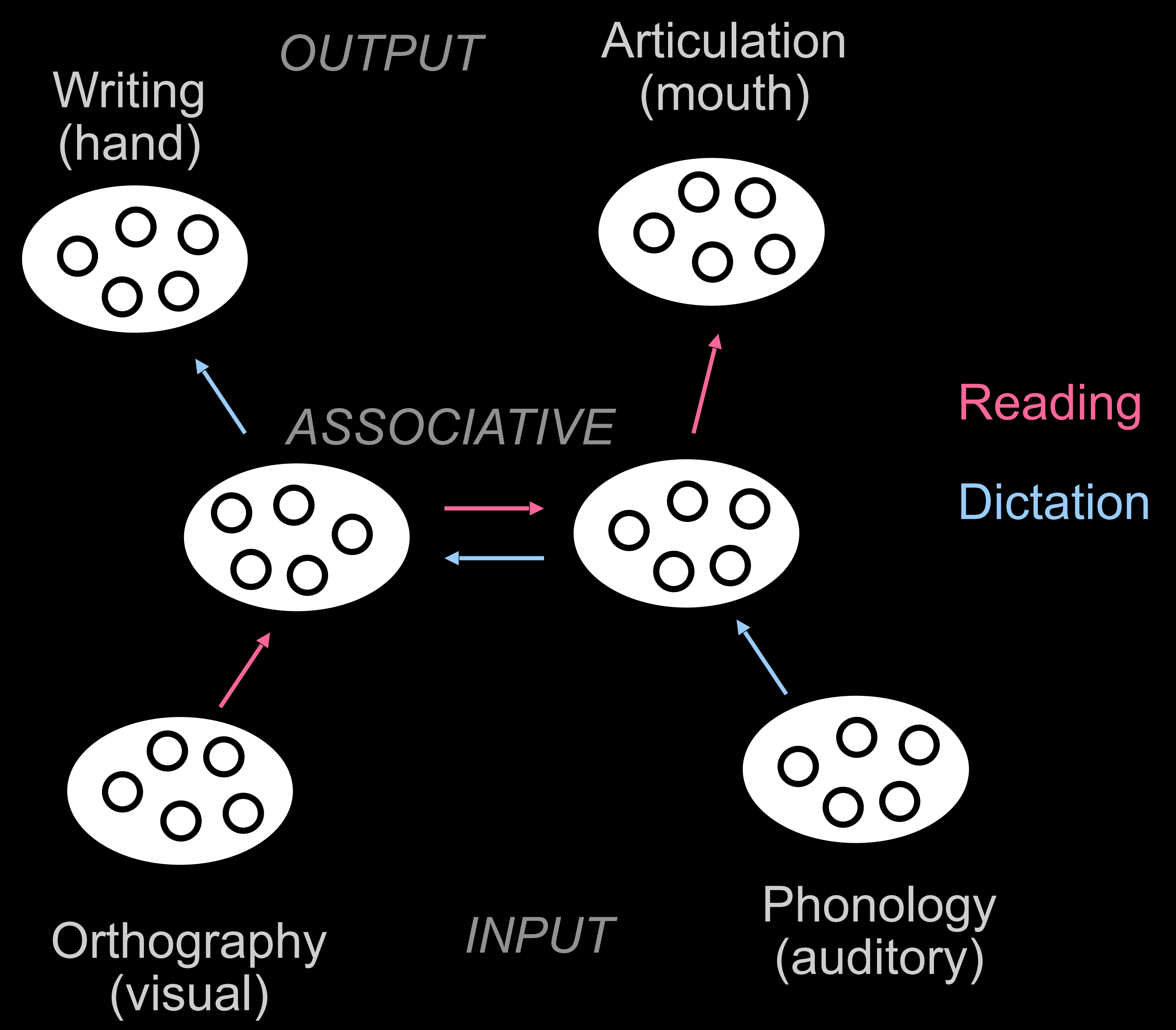
carrying out a particular function

sensory encoding (input), associative (hidden), motor control (output), etc.



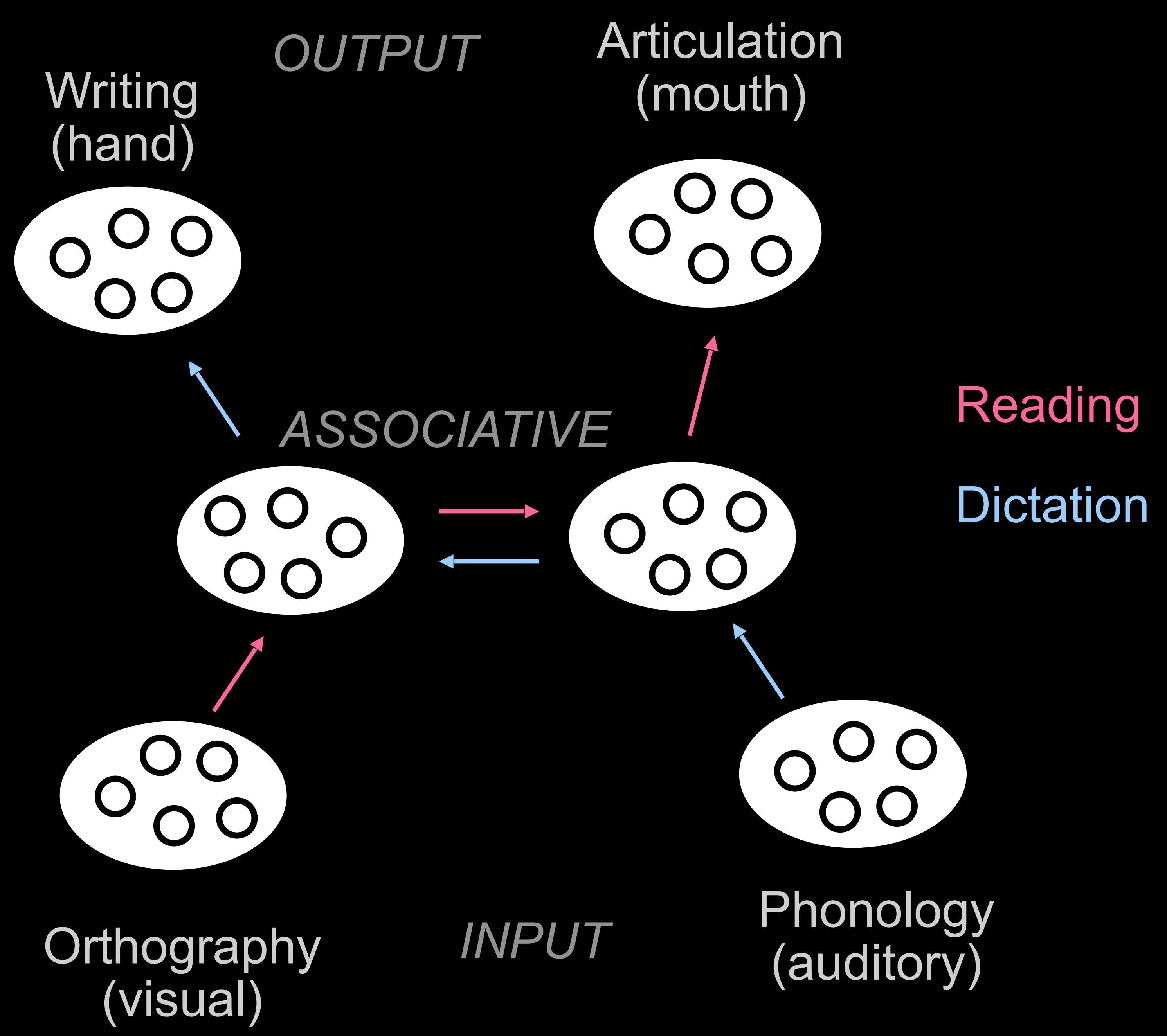
Processing

- Flow of activity among units / between modules



Processing

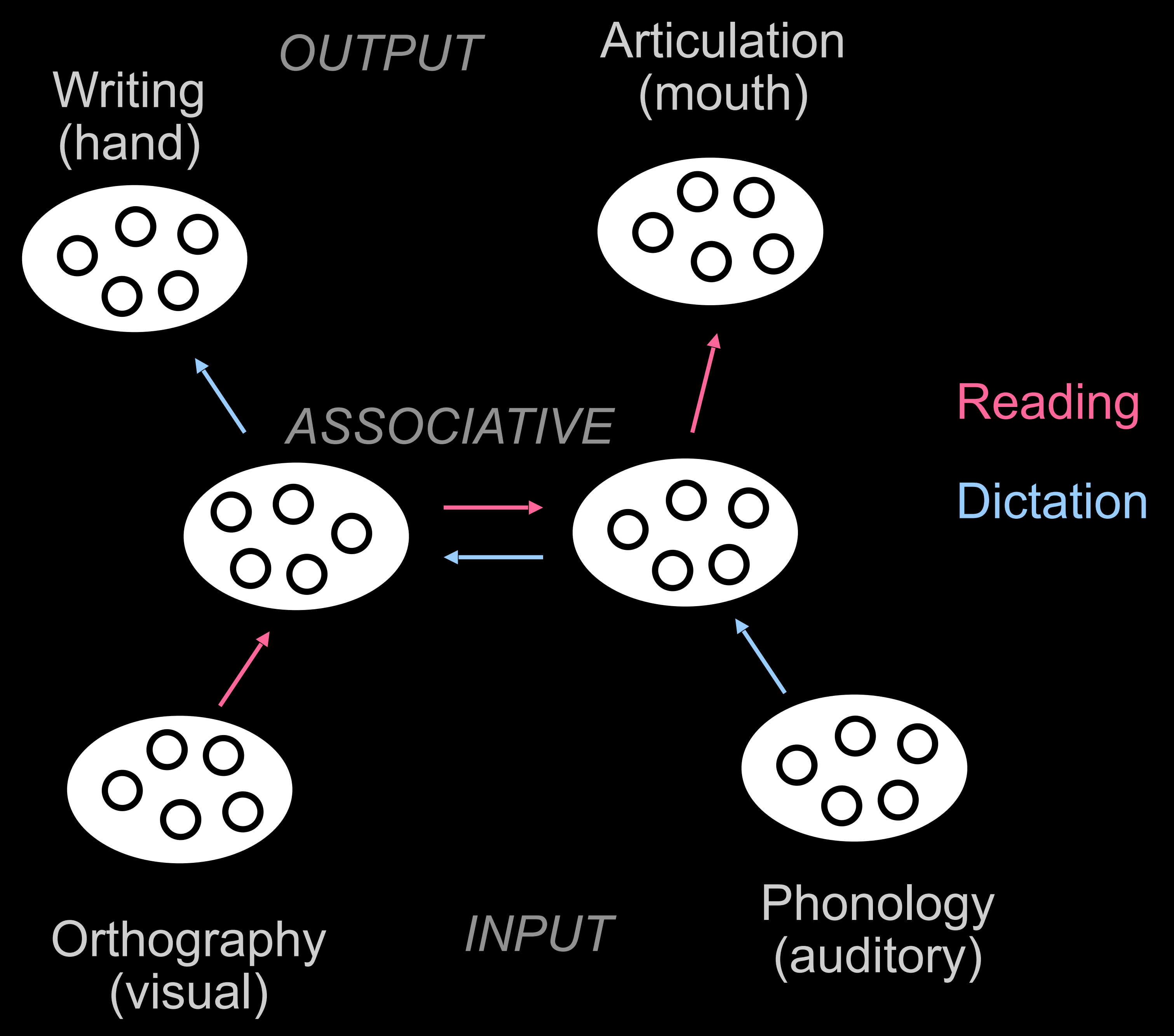
- Flow of activity among units / between modules
  - Input output mappings (pathways)



Processing

- Flow of activity among units / between modules

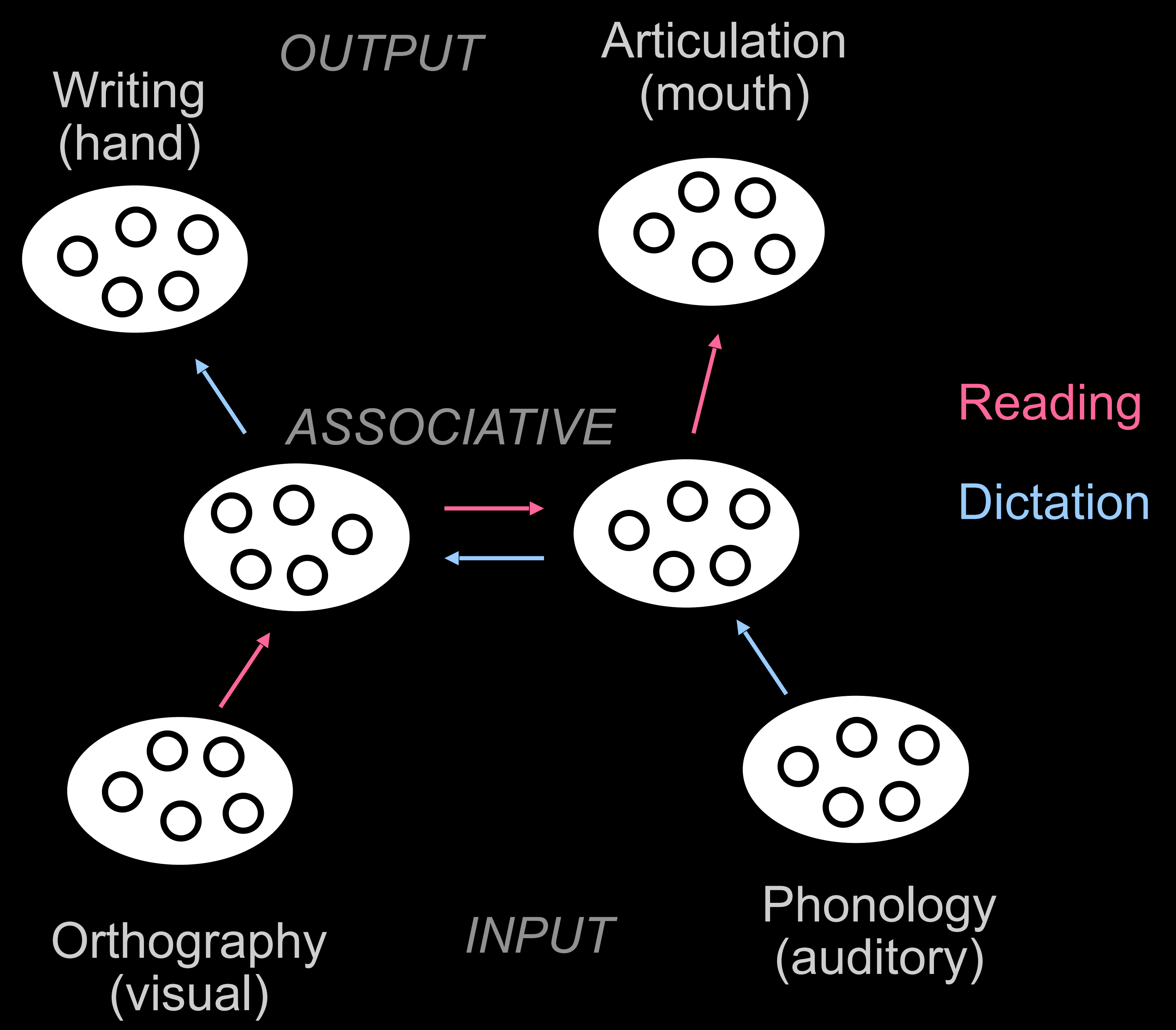
  - Interference



Processing

- Flow of activity among units / between modules

  - Control = modulation



Learning

- **Weight modification** (≈ synaptic plasticity)

Learning

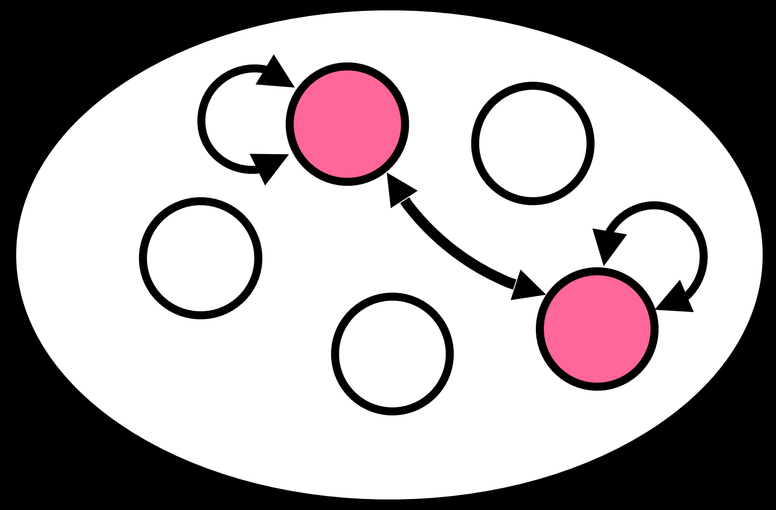
- **Weight modification** (≈ synaptic plasticity)
  - **Unsupervised** (self-organizing)
    - **Simple associative** (Hebbian)
    - **Competitive** (K-winner take all)

Learning

- **Weight modification** (≈ synaptic plasticity)
  - **Unsupervised** (self-organizing)
    - **Simple associative** (Hebbian)
    - **Competitive** (K-winner take all)
  - **Supervised (trained)**
    - **Reinforcement** (temporal differences)
    - **Structured** (backpropagation)

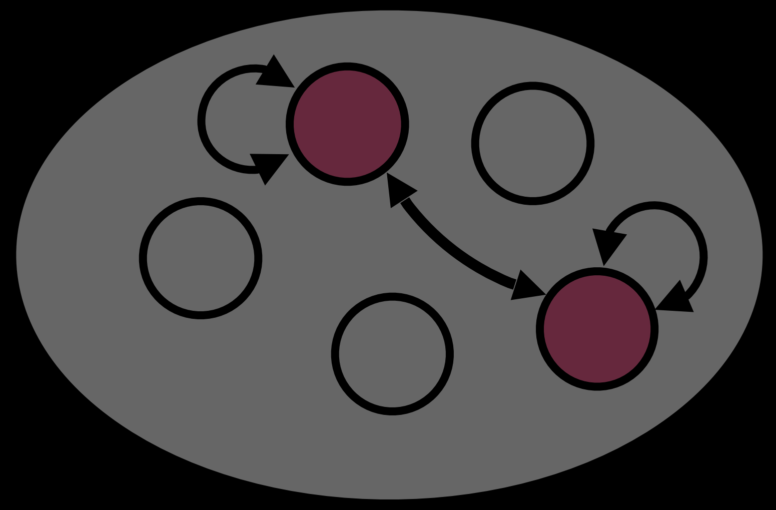
Memory

- Short Term
  - sustained pattern of activity

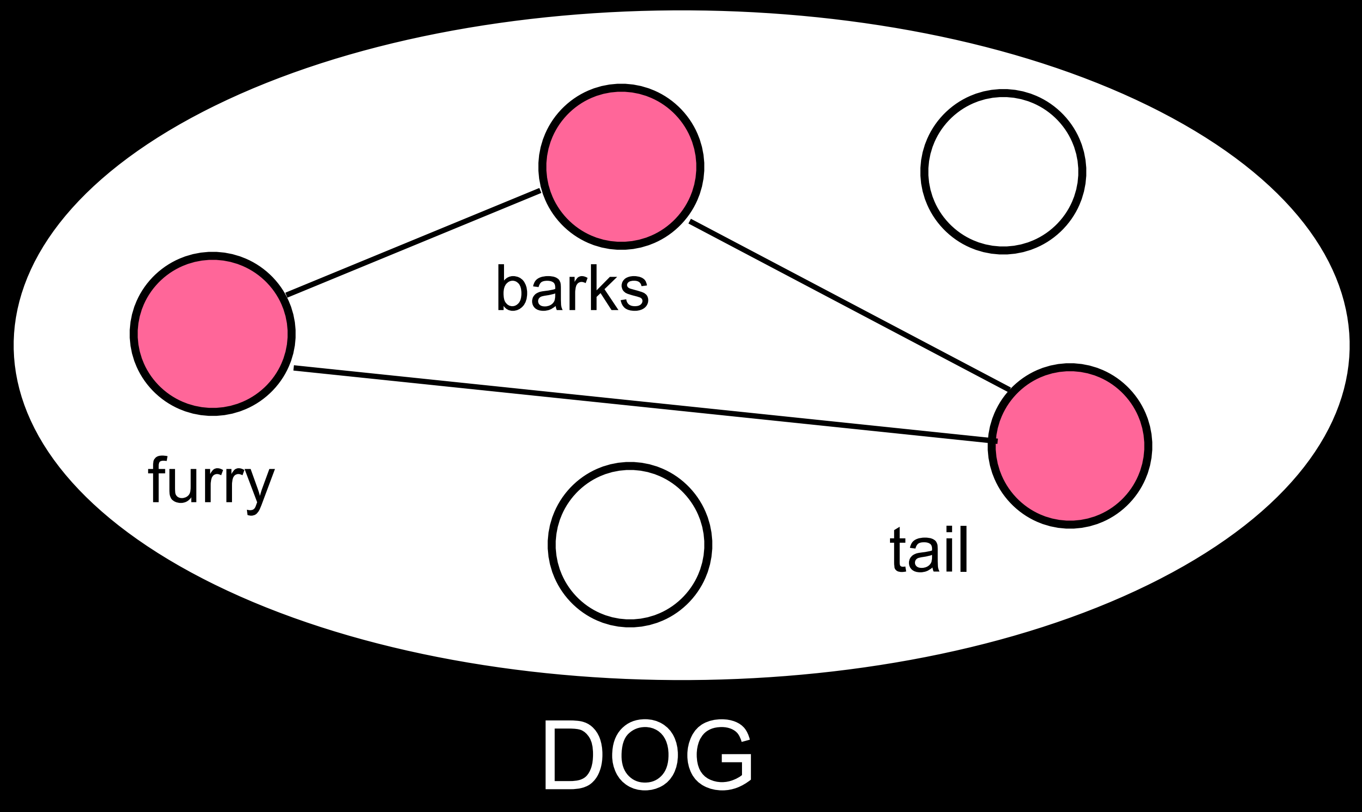


Memory

- Short Term
  - sustained pattern of activity

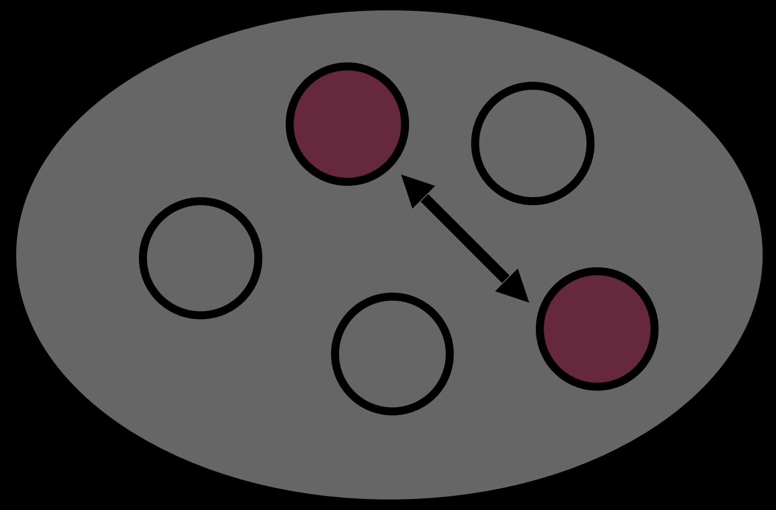


- Long Term
  - connections among associated features

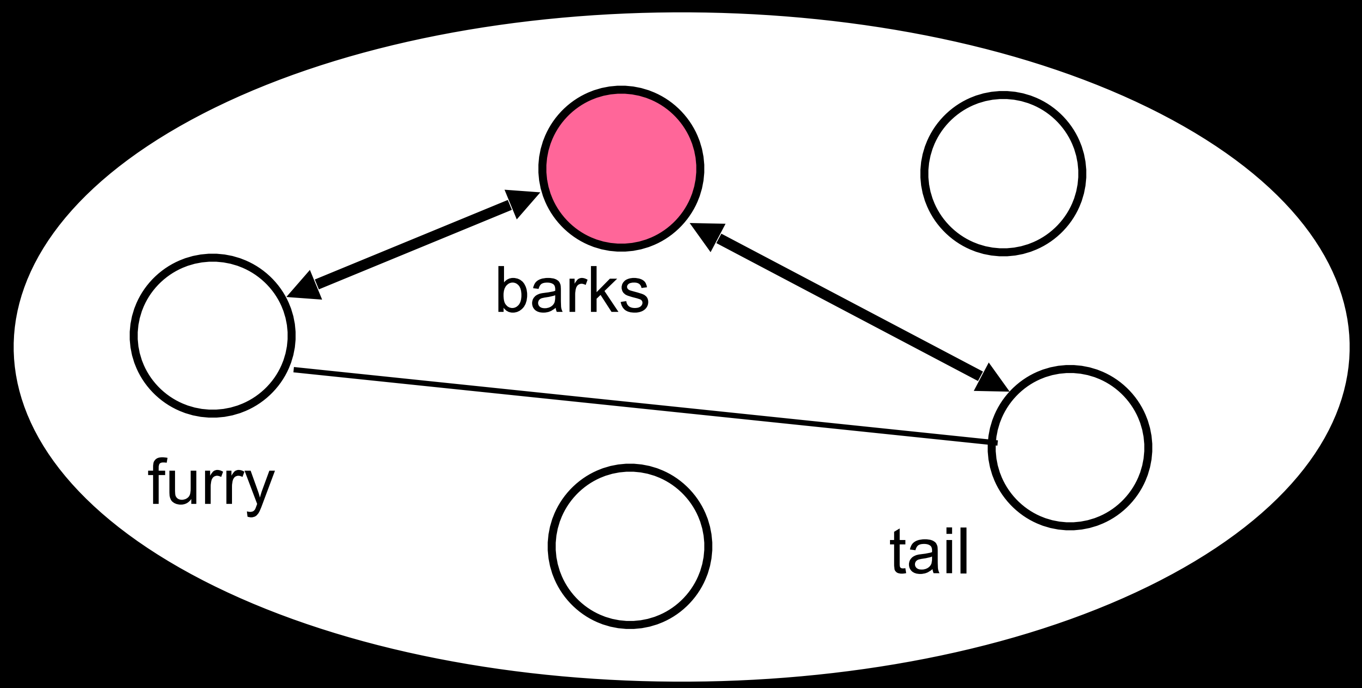


Memory

- Short Term
  - sustained pattern of activity



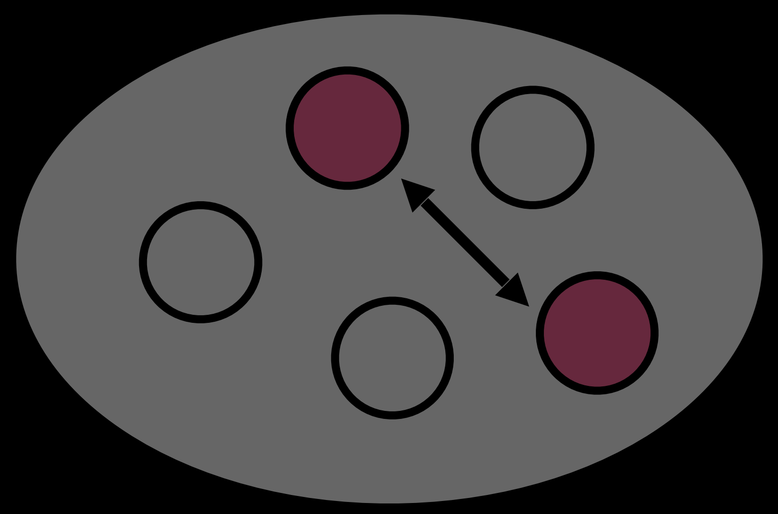
- Long Term
  - retrieval:



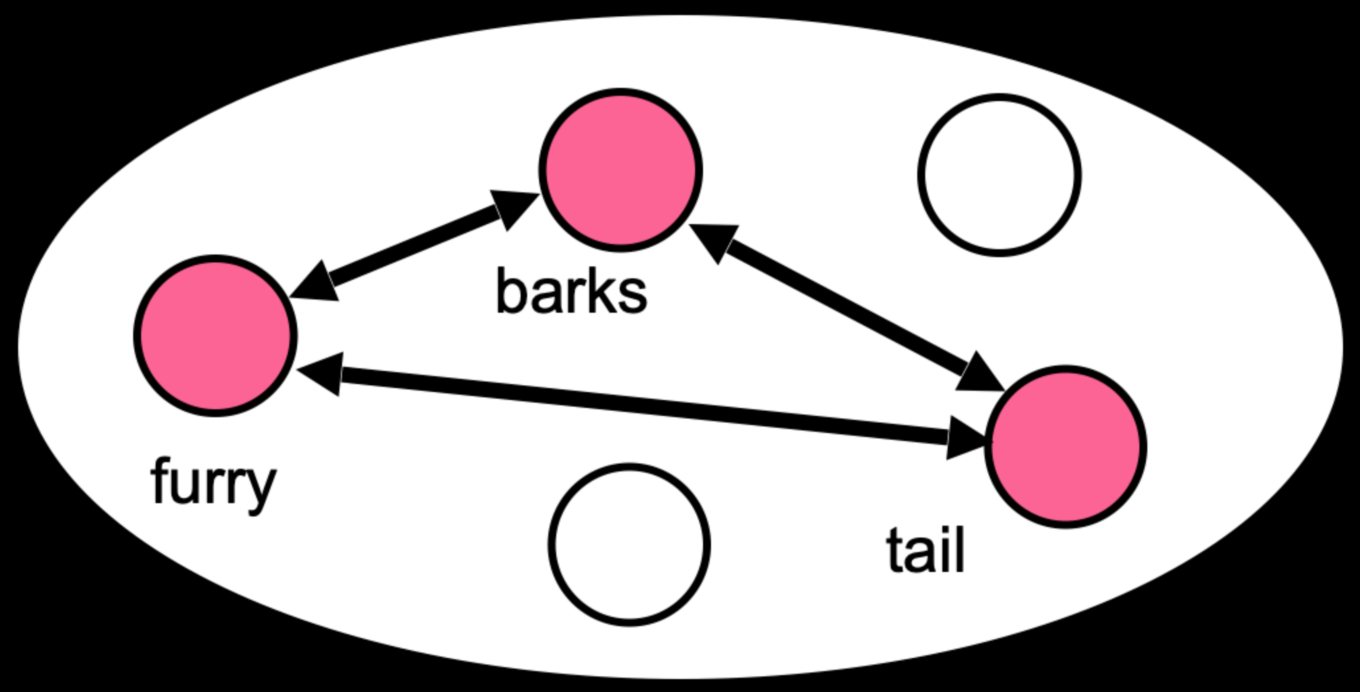
DOG

Memory

- Short Term
  - sustained pattern of activity



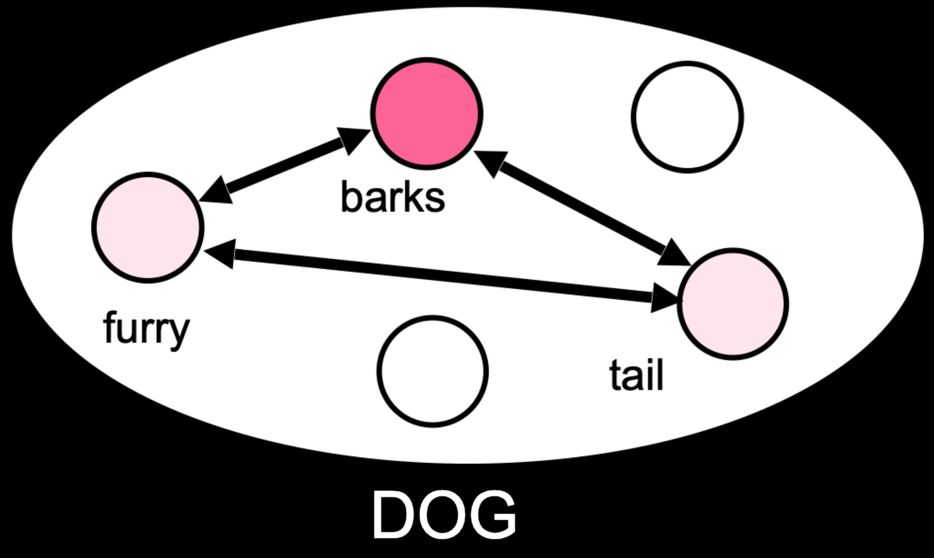
- Long Term
  - retrieval: reactivation of a whole pattern from its parts

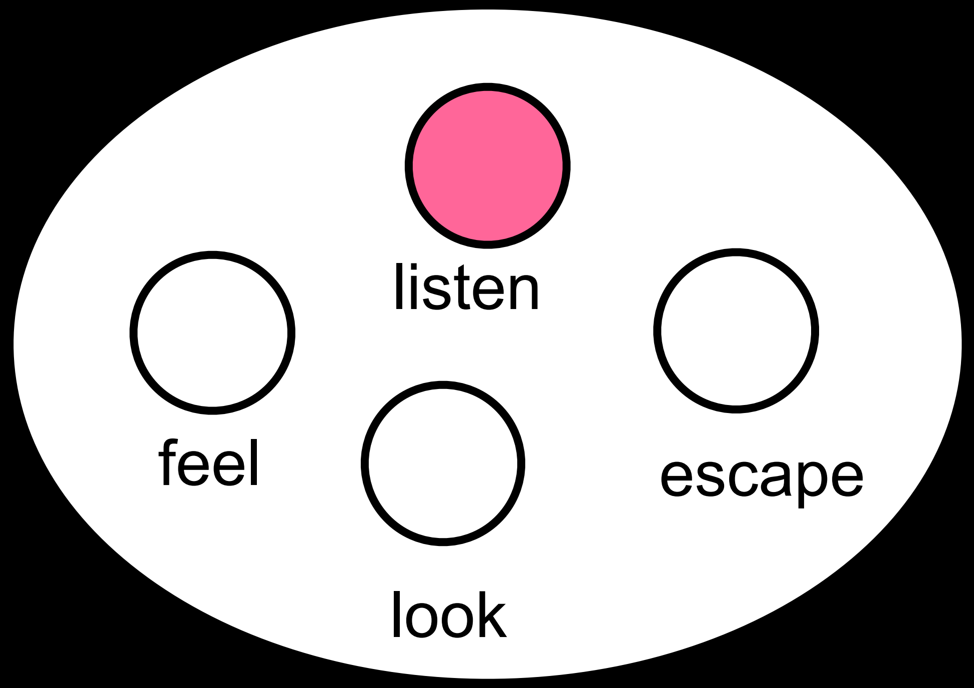


DOG

Control

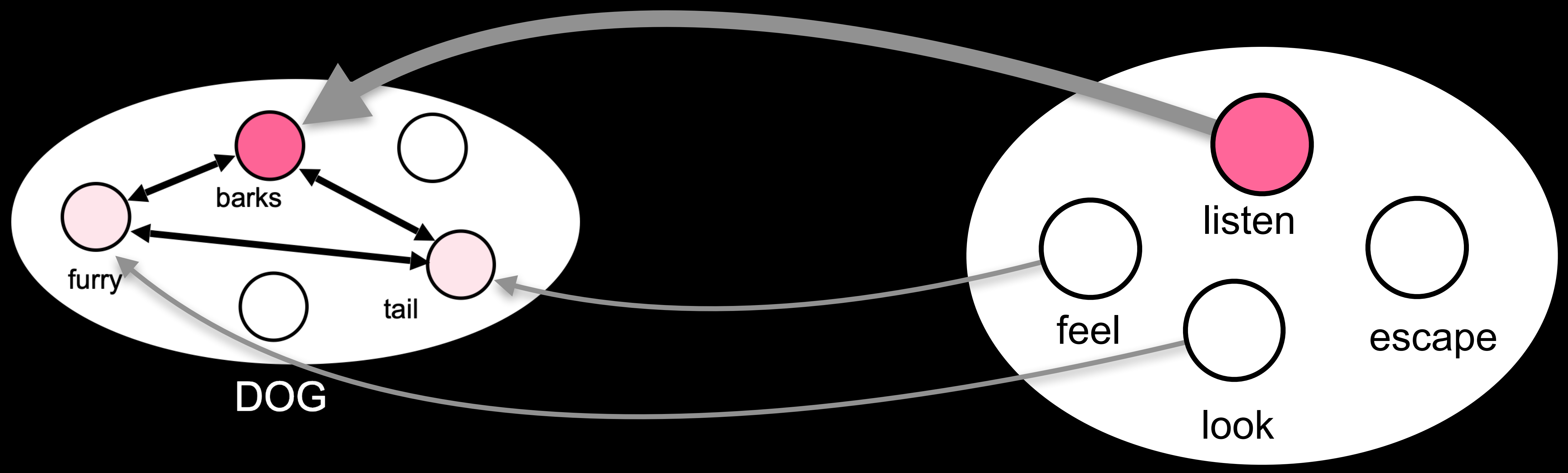
- Attention
  - selection of some features to activate





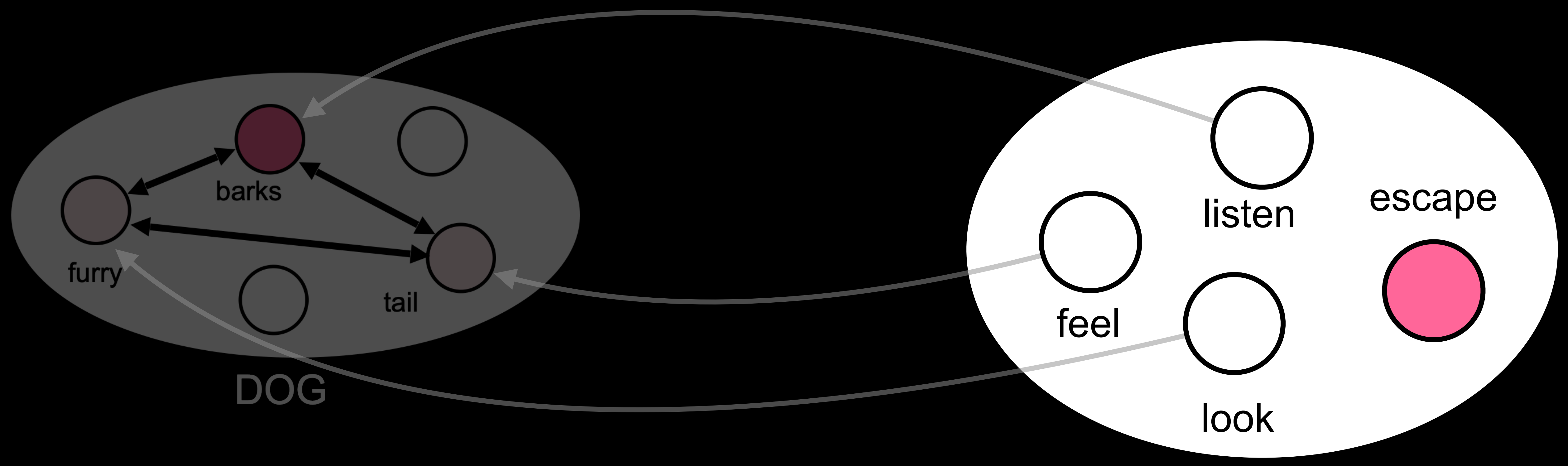
Control

- **Attention**
  - **selection of some features to activate**

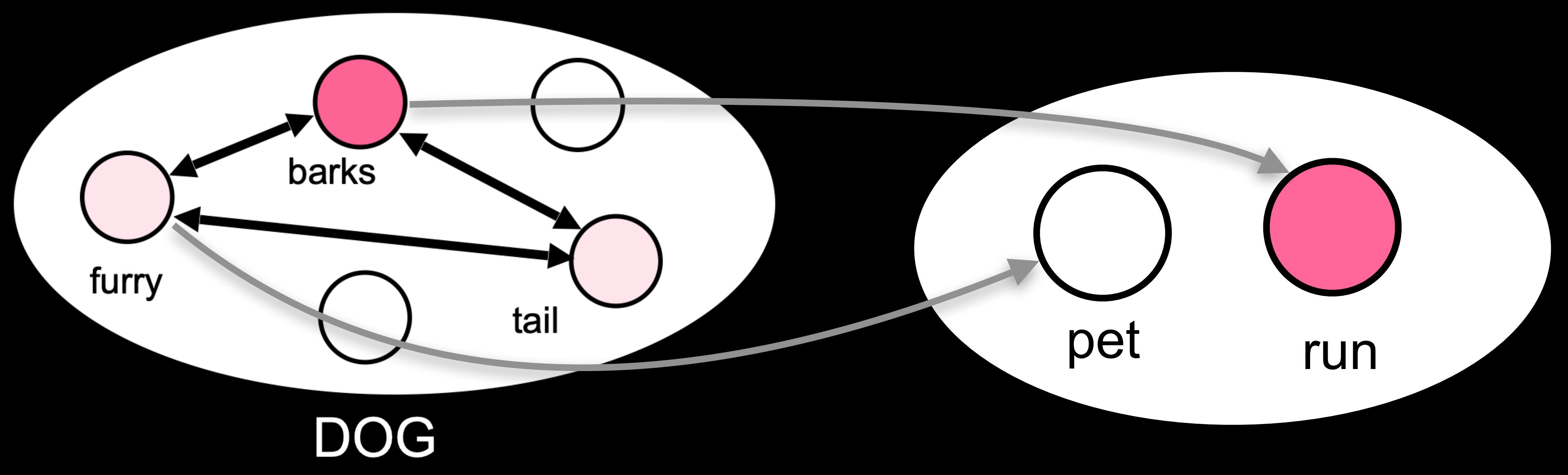


Control

- Attention
  - selection of features to activate

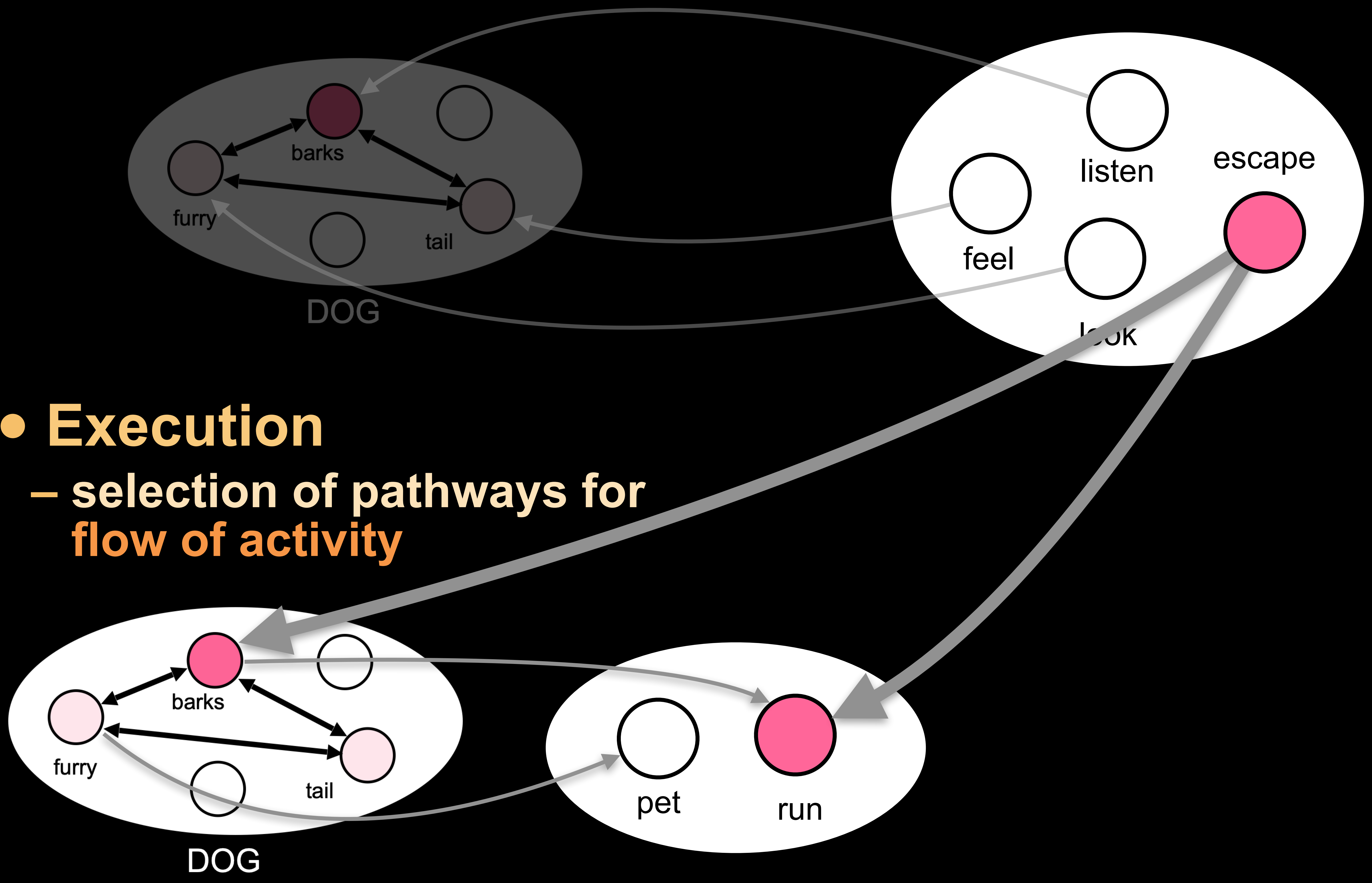


- Execution
  - selection of pathways for flow of activity



Control

- Attention
  - selection of features to activate
- Execution
  - selection of pathways for flow of activity



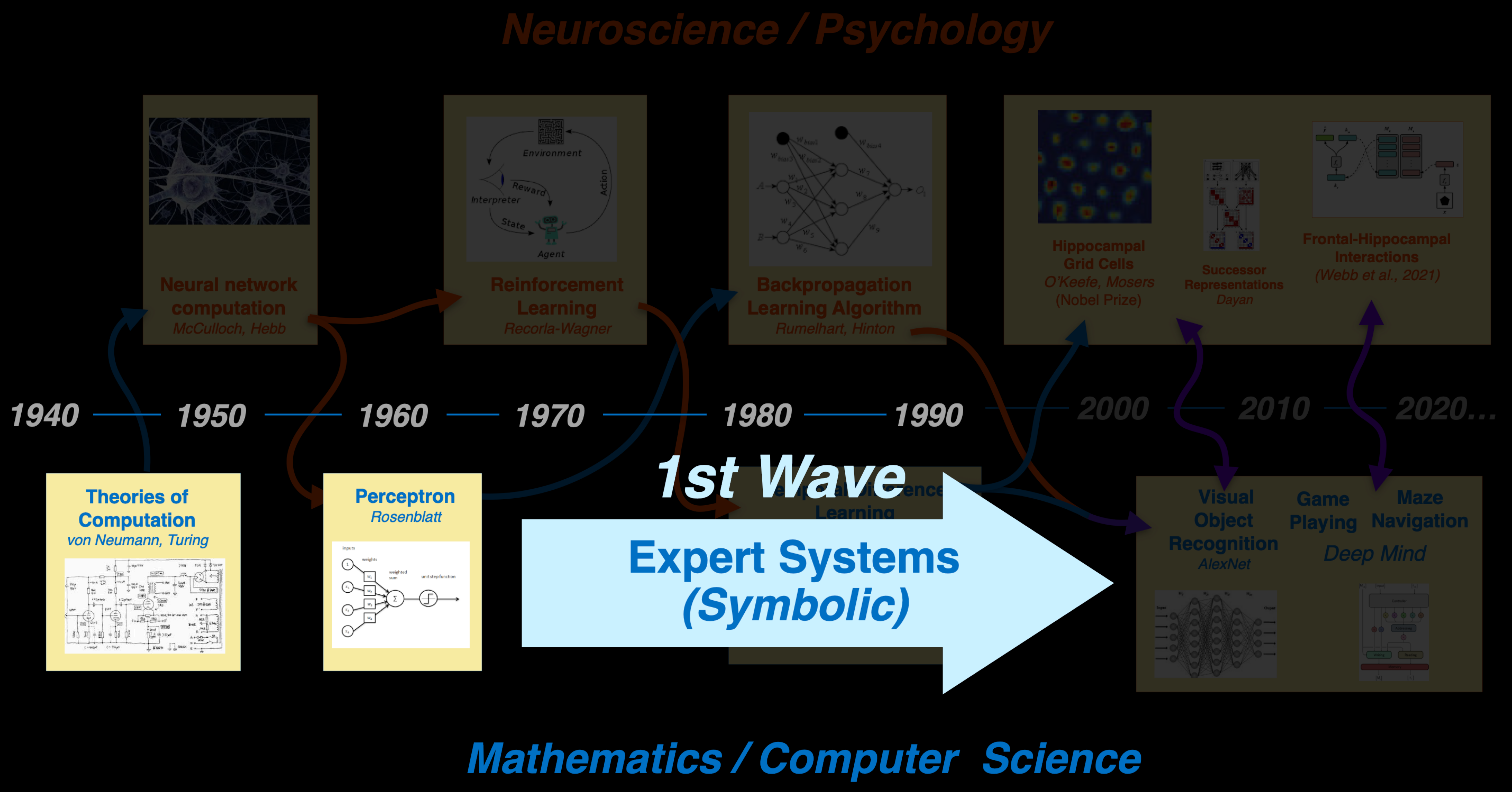
Early Connectionist Models

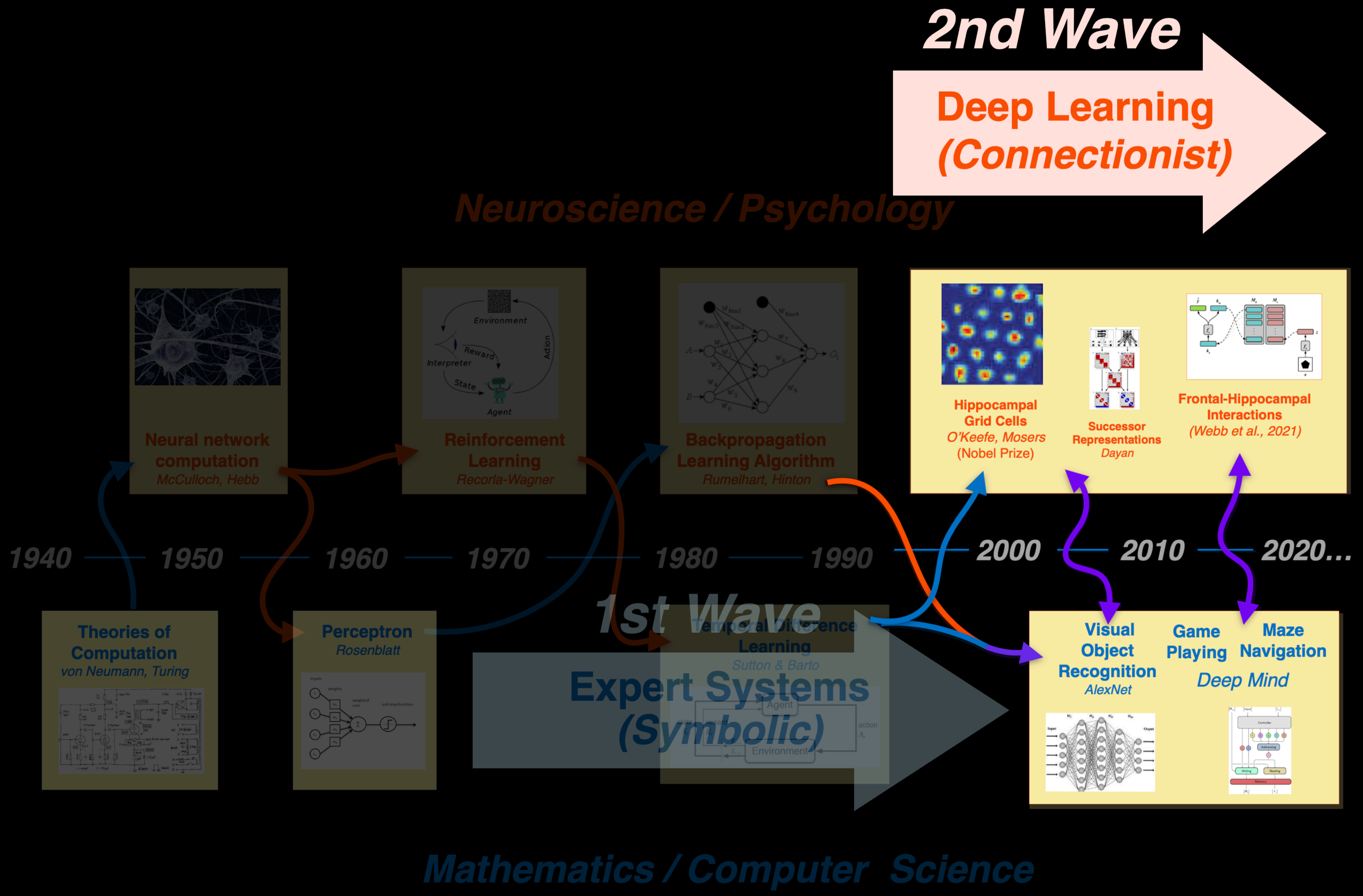
- Good at doing what the brain does easily
  (and what traditional computers do poorly):
  - visual pattern recognition
  - language processing
  - generalization / pattern completion

- Bad at doing what the brain does poorly
  (and what traditional computer programs do easily)
  - complex sequential operations (e.g., arithmetic)
  - rapid repetitive computations

Early Connectionist Models

- Good at doing what the brain does easily (and what traditional computers do poorly):
  - visual pattern recognition
  - language processing
  - generalization / pattern completion
- Bad at doing what the brain does poorly (and what traditional computer programs do easily)
  - complex sequential operations (e.g., arithmetic)
  - rapid repetitive computations
- However, things have changed…


Neuroscience / Psychology
Neural network computation
McCulloch, Hebb
Environment
Reward
Interpreter
State
Agent
Action
Reinforcement Learning
Recorla-Wagner
Backpropagation Learning Algorithm
Rumelhart, Hinton
Hippocampal Grid Cells
O'Keefe, Mosers
(Nobel Prize)
Successor Representations
Dayan
Frontal-Hippocampal Interactions
(Webb et al., 2021)
1940
1950
1960
1970
1980
1990
2000
2010
2020…
Theories of Computation
von Neumann, Turing
Perceptron
Rosenblatt
1st Wave
Expert Systems
(Symbolic)
Visual Object Recognition
AlexNet
Game Playing
Maze Navigation
Deep Mind
Mathematics / Computer Science



2nd Wave
Deep Learning
(Connectionist)
Neuroscience / Psychology
Neural network computation
McCulloch, Hebb
Reinforcement Learning
Recorla-Wagner
Backpropagation Learning Algorithm
Rumelhart, Hinton
Hippocampal Grid Cells
O'Keefe, Mosers
(Nobel Prize)
Successor Representations
Dayan
Frontal-Hippocampal Interactions
(Webb et al., 2021)
1940
1950
1960
1970
1980
1990
2000
2010
2020…
Theories of Computation
von Neumann, Turing
Perceptron
Rosenblatt
1st Wave
Temporal Difference Learning
Sutton & Barto
Expert Systems
(Symbolic)
Visual Object Recognition
AlexNet
Game Playing
Maze Navigation
Deep Mind
Mathematics / Computer Science

“Deep Learning”

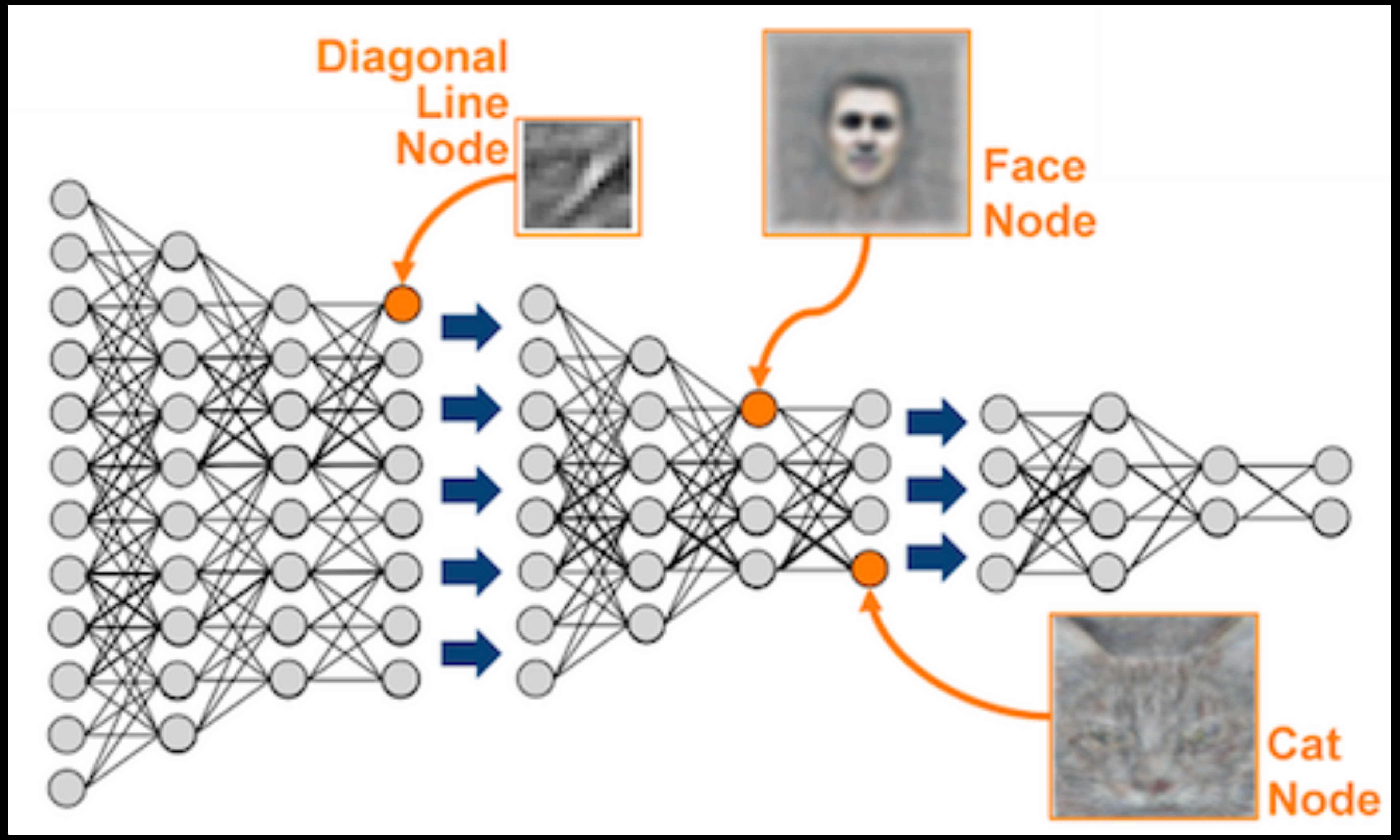
Dave Rumelhart
(UCSD / Stanford)

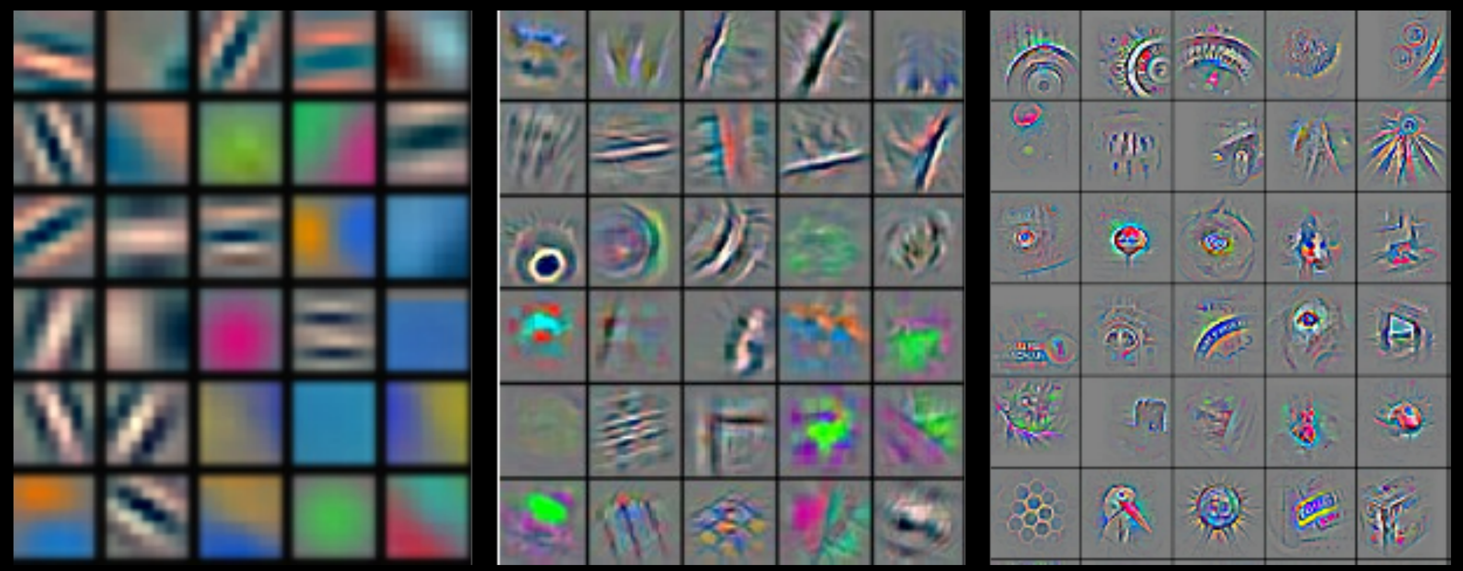
Jann LeCun
(NYU)

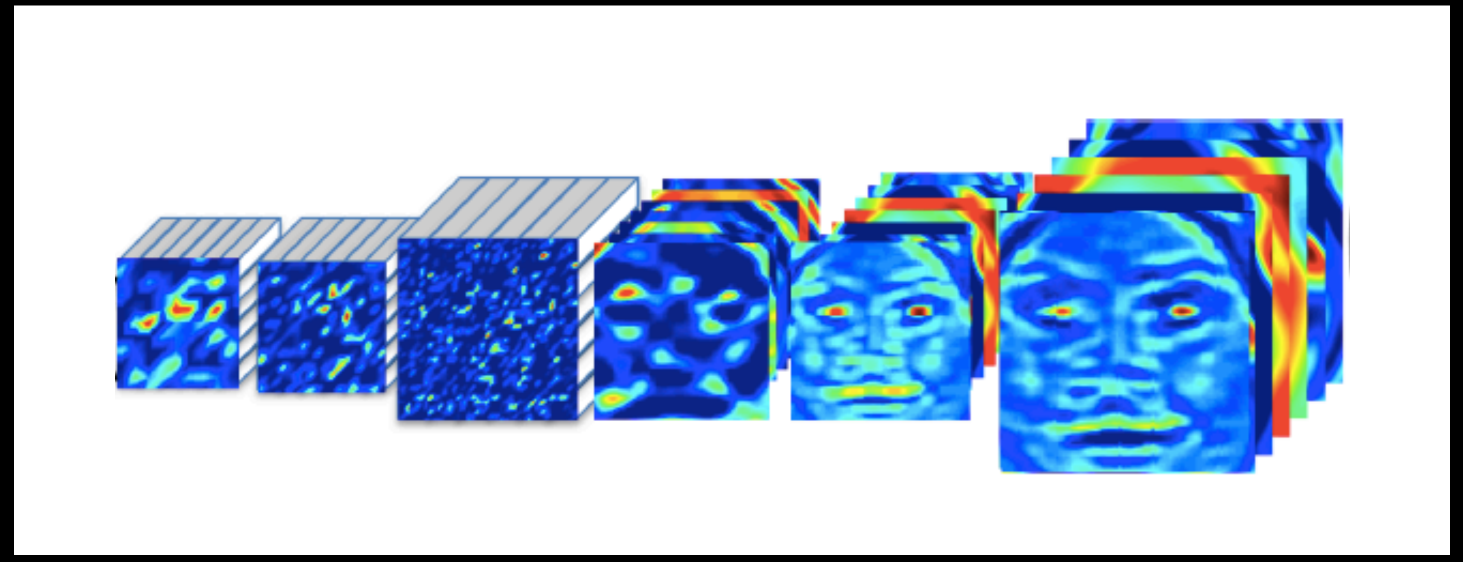
Geoff Hinton
(Toronto / Google)

Bruno Olshausen
(Redwood Institute)

Fei Fei Li
(Princeton/Stanford)

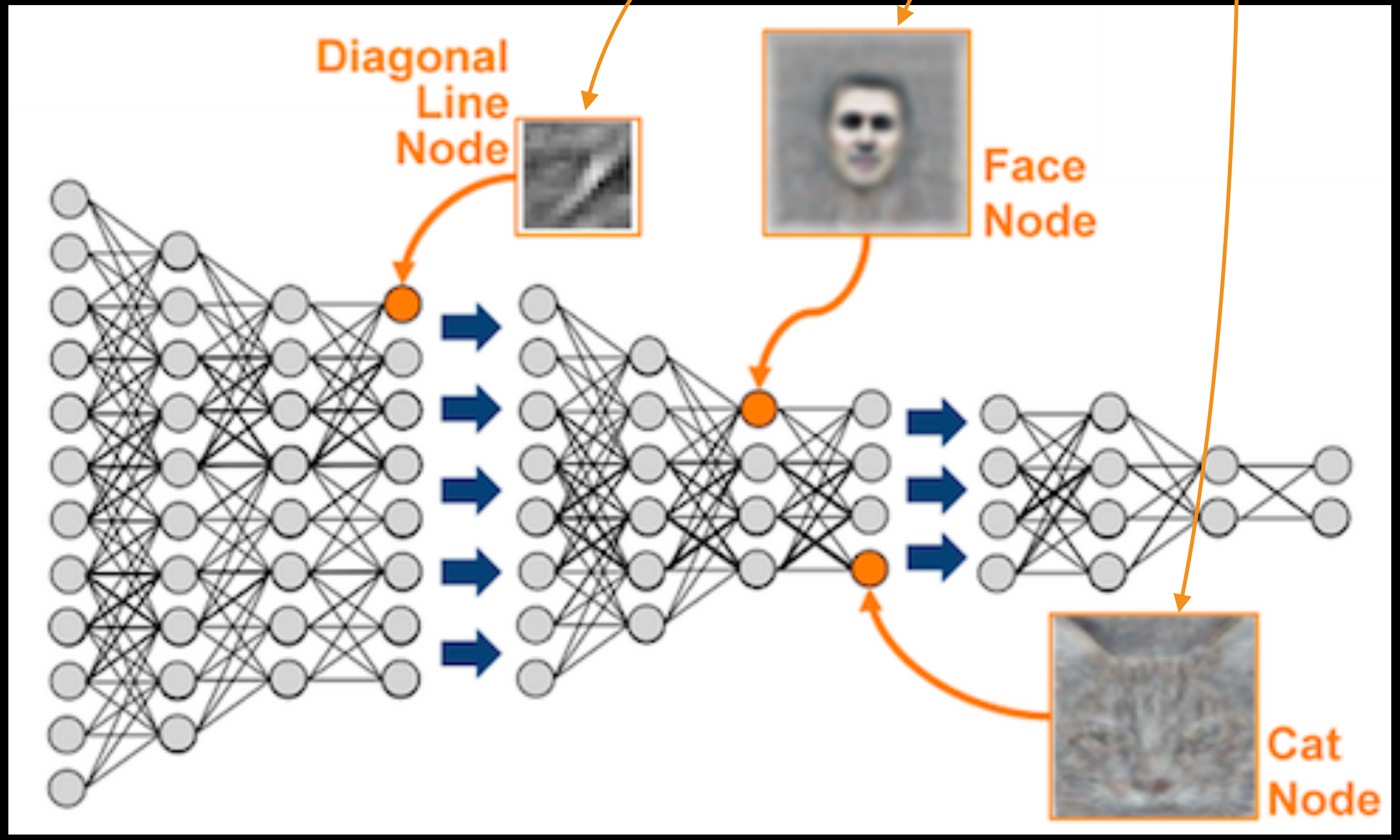


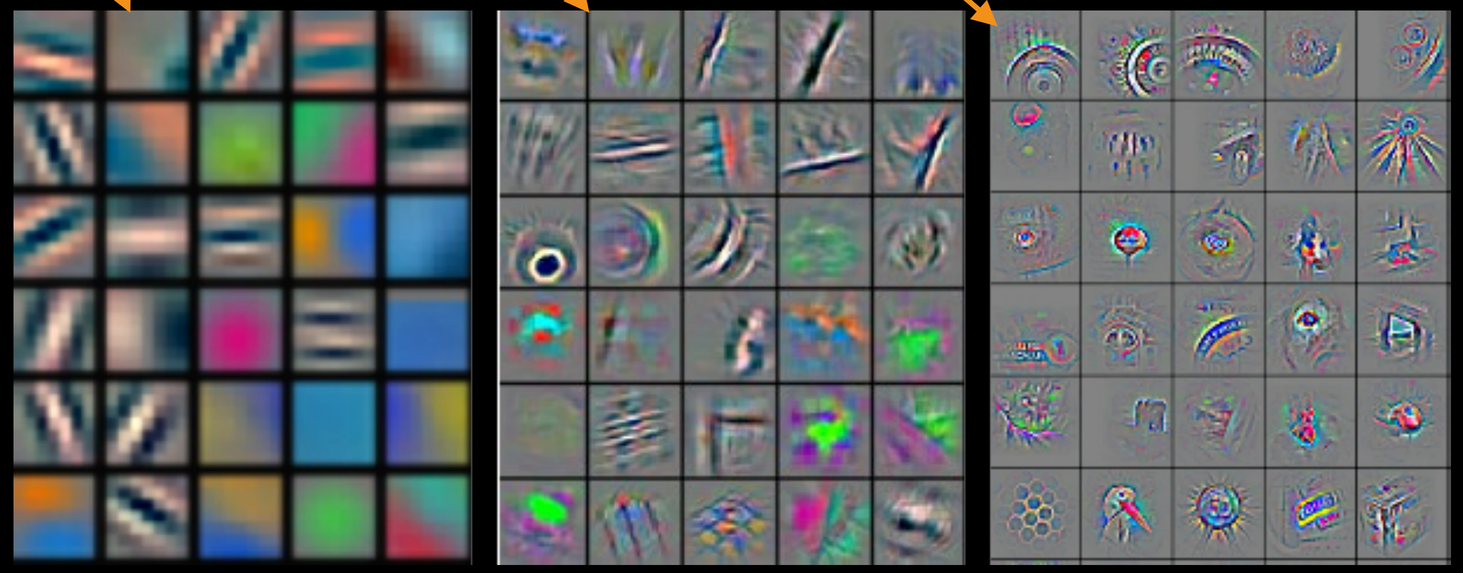


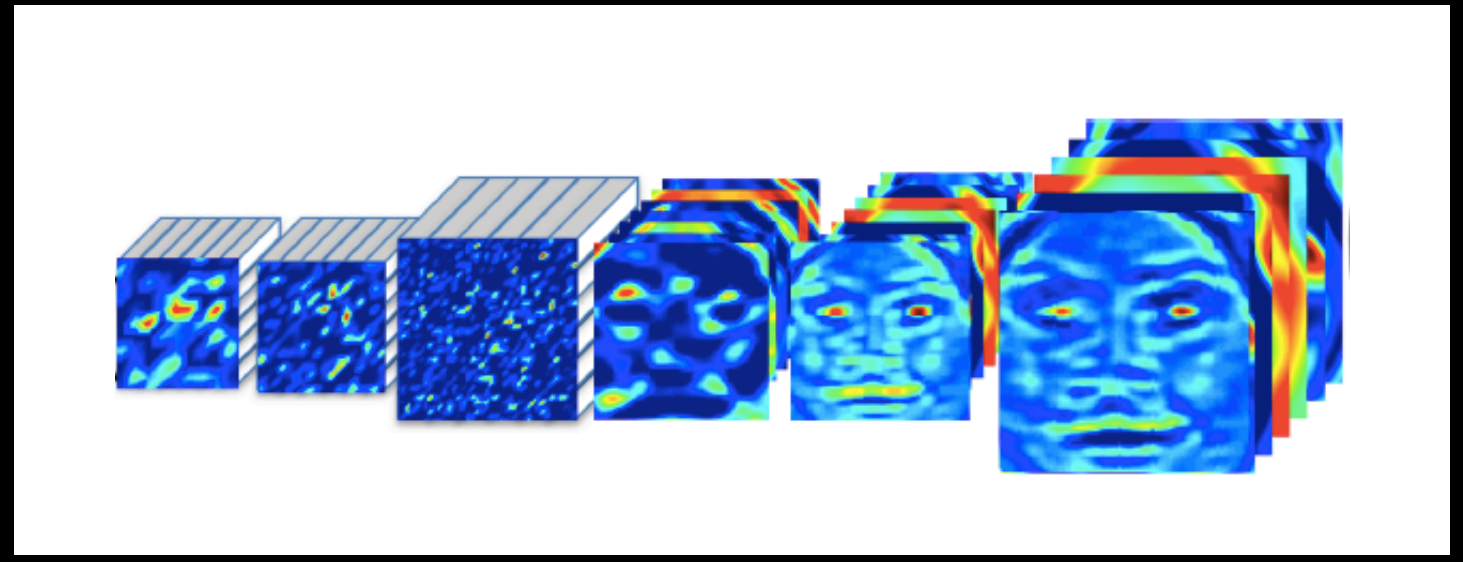


“Deep Learning”

Internal representations

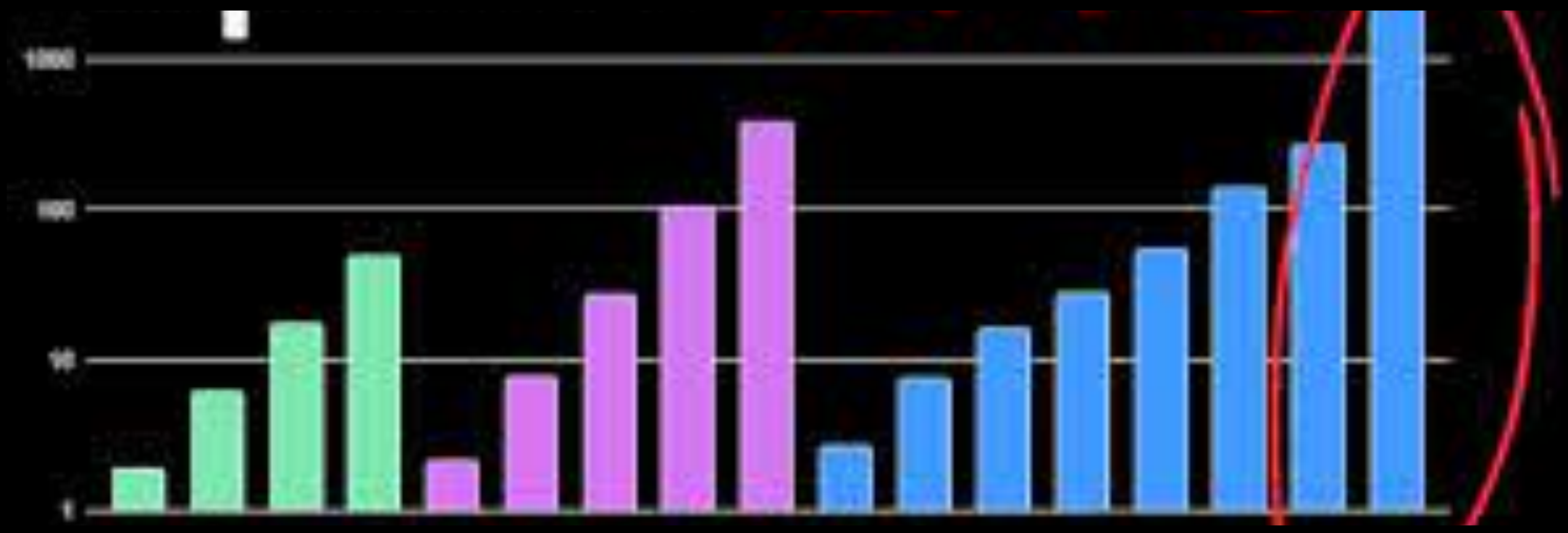






“Deep Learning”

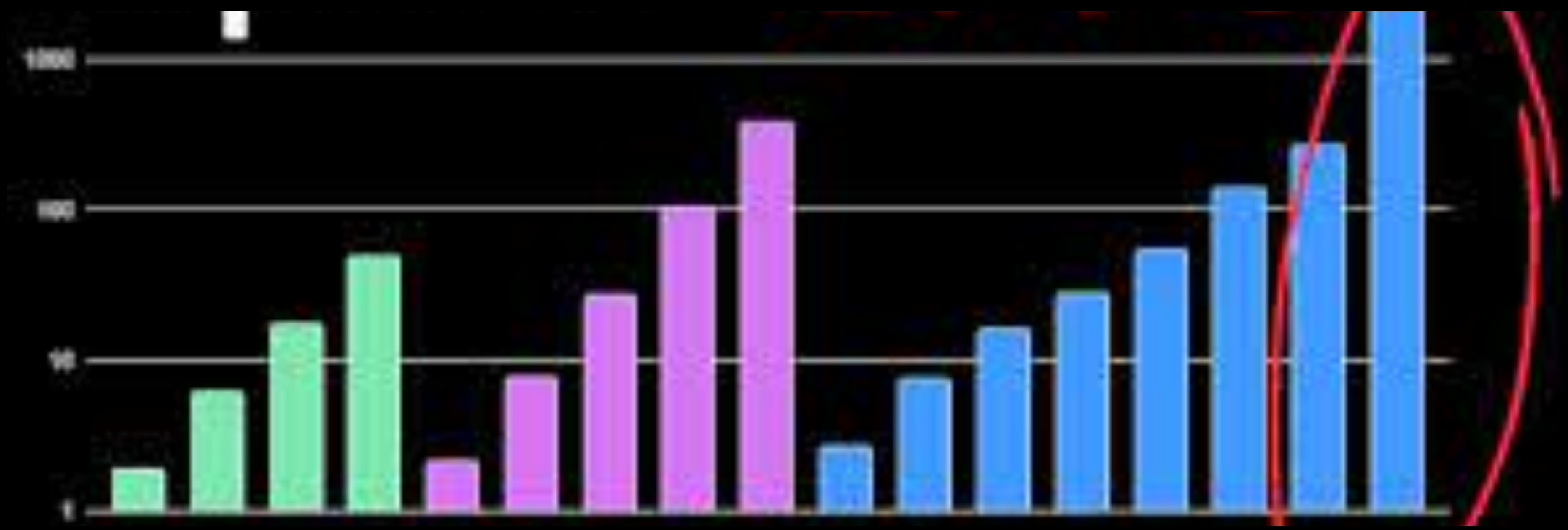
Google GEMINI



1.5 TRILLION Parameters

“Deep Learning”

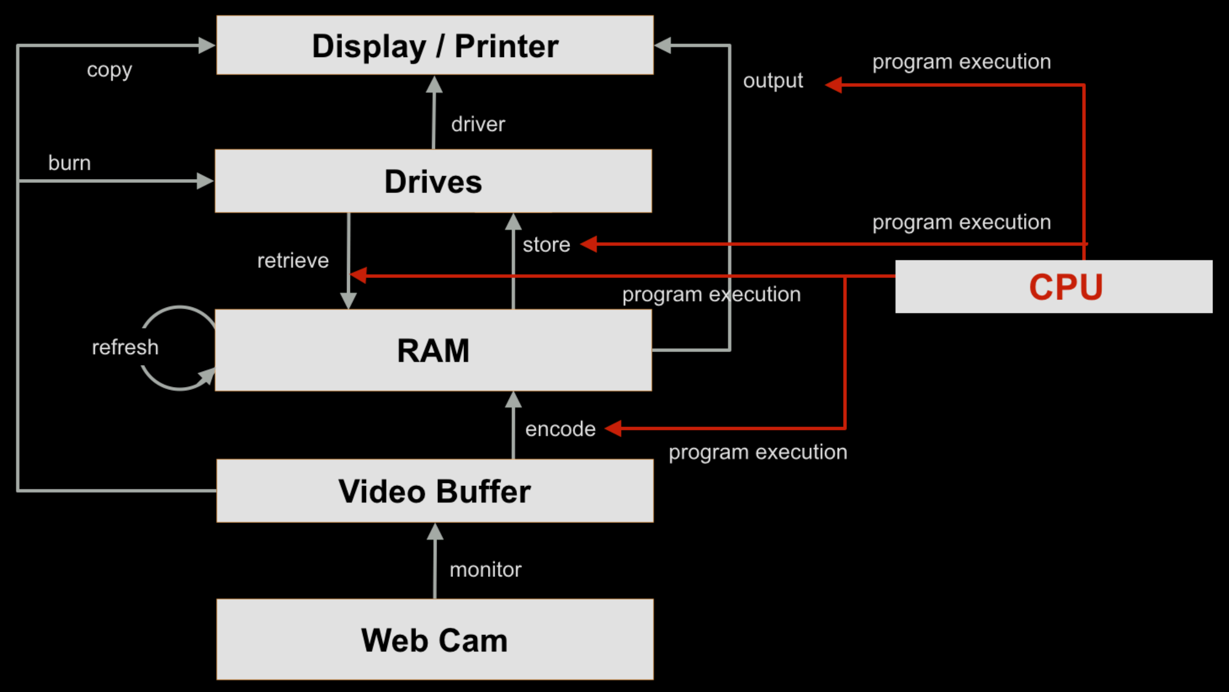
Google GEMINI



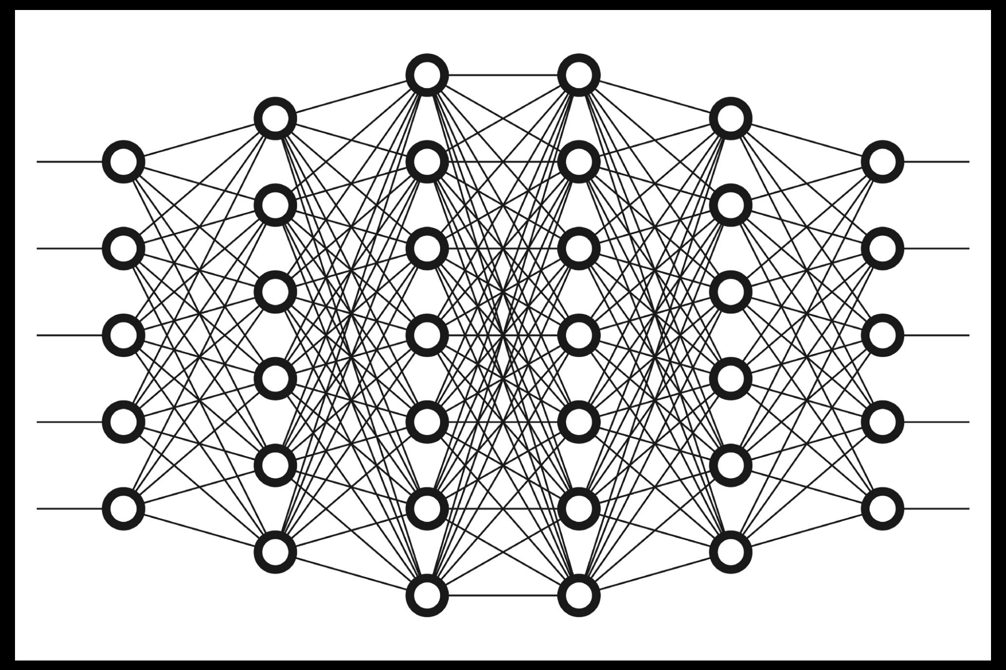
1.5 TRILLION Parameters

Extremes of a Continuum

Symbolic

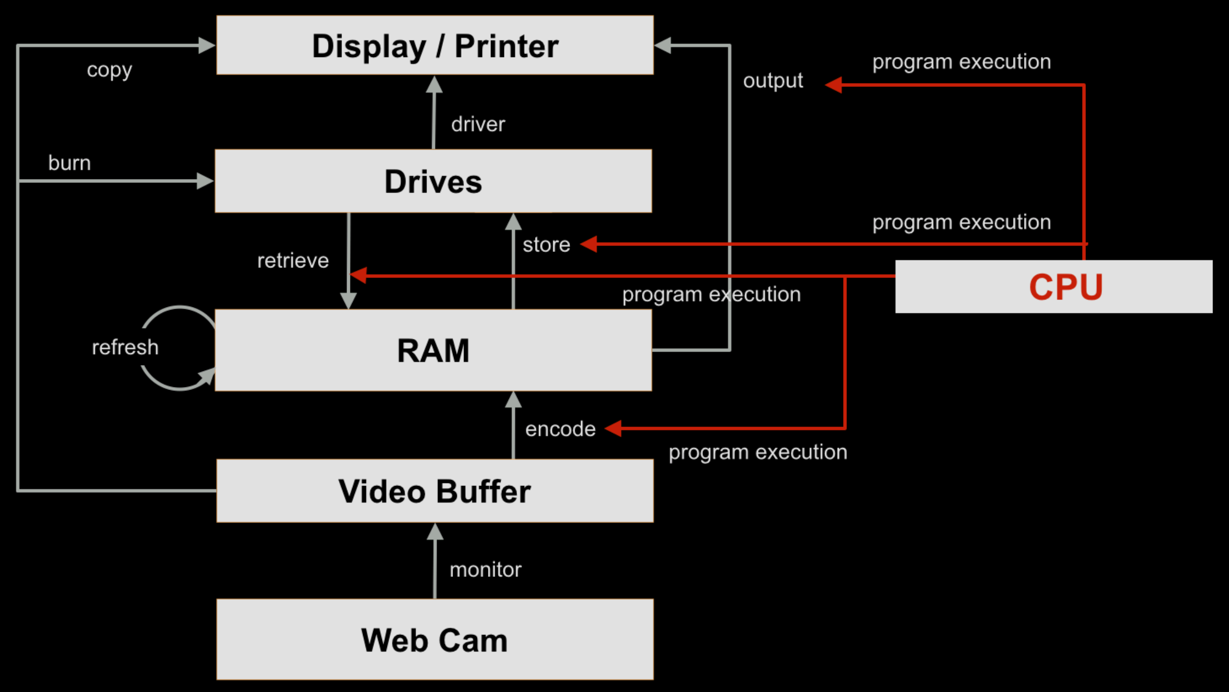


Connectionist



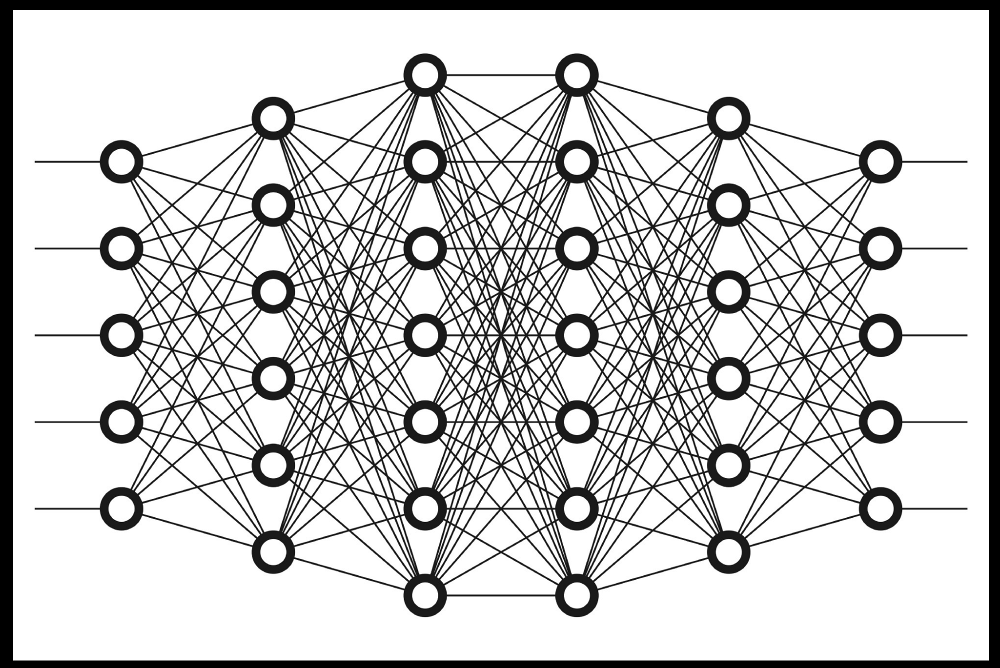
Extremes of a Continuum

Symbolic



logical

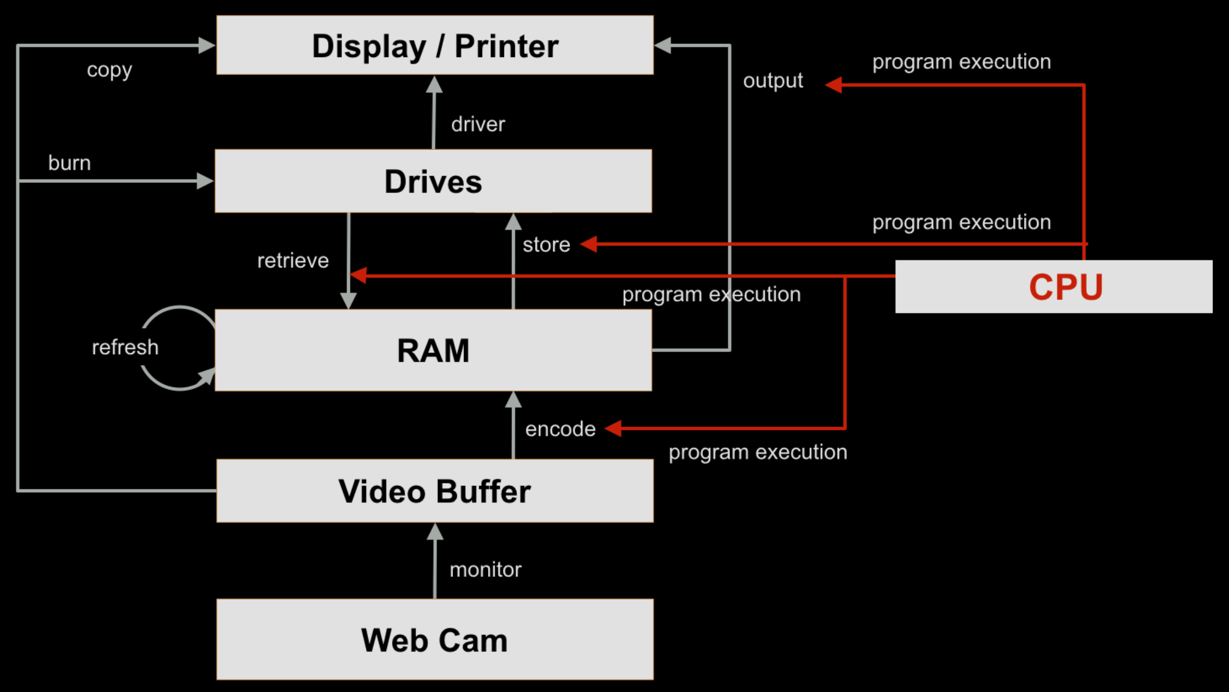
Connectionist



statistical

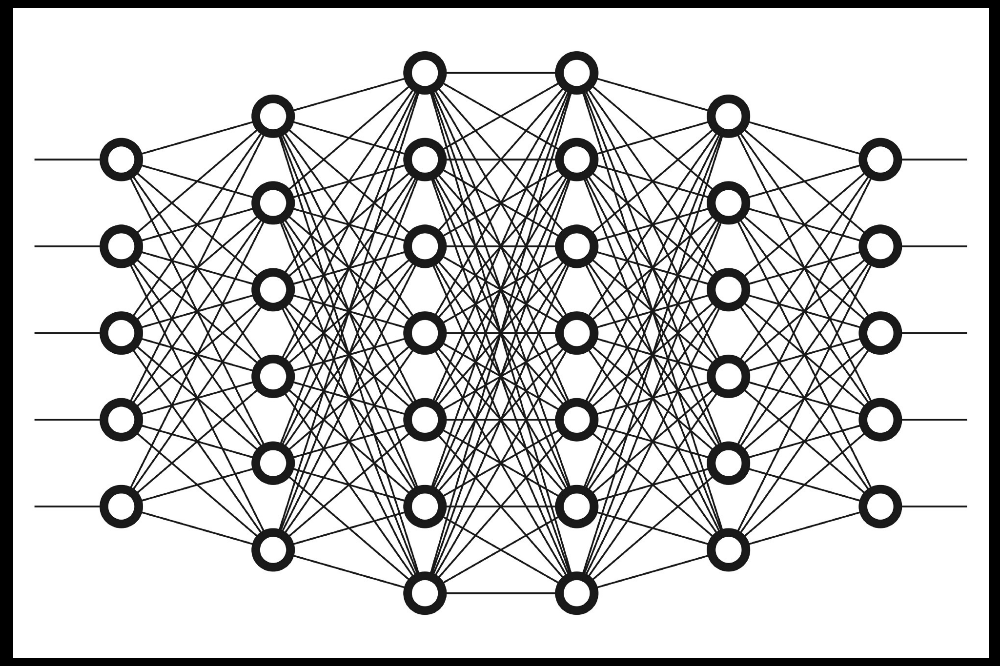
Extremes of a Continuum

Symbolic



serial

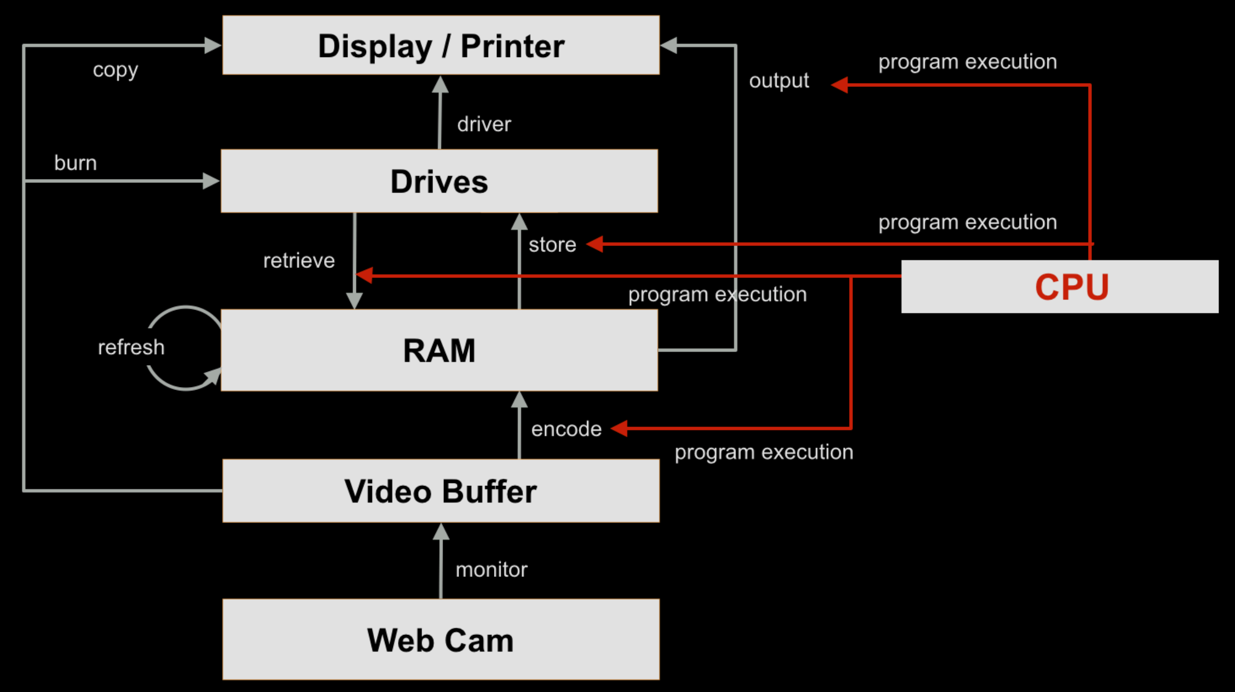
Connectionist



parallel

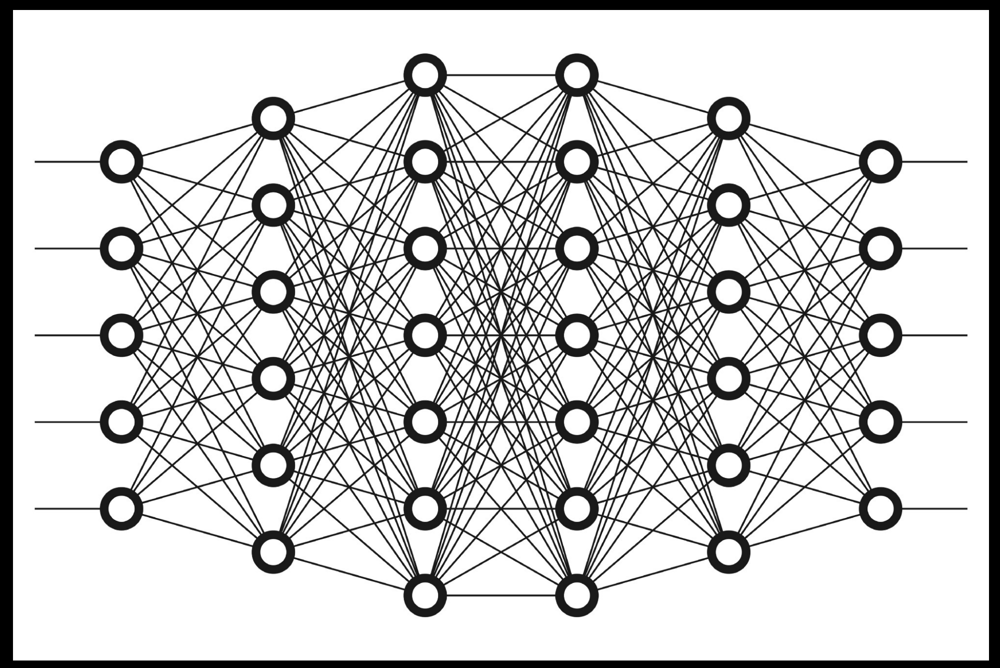
Extremes of a Continuum

Symbolic



discrete

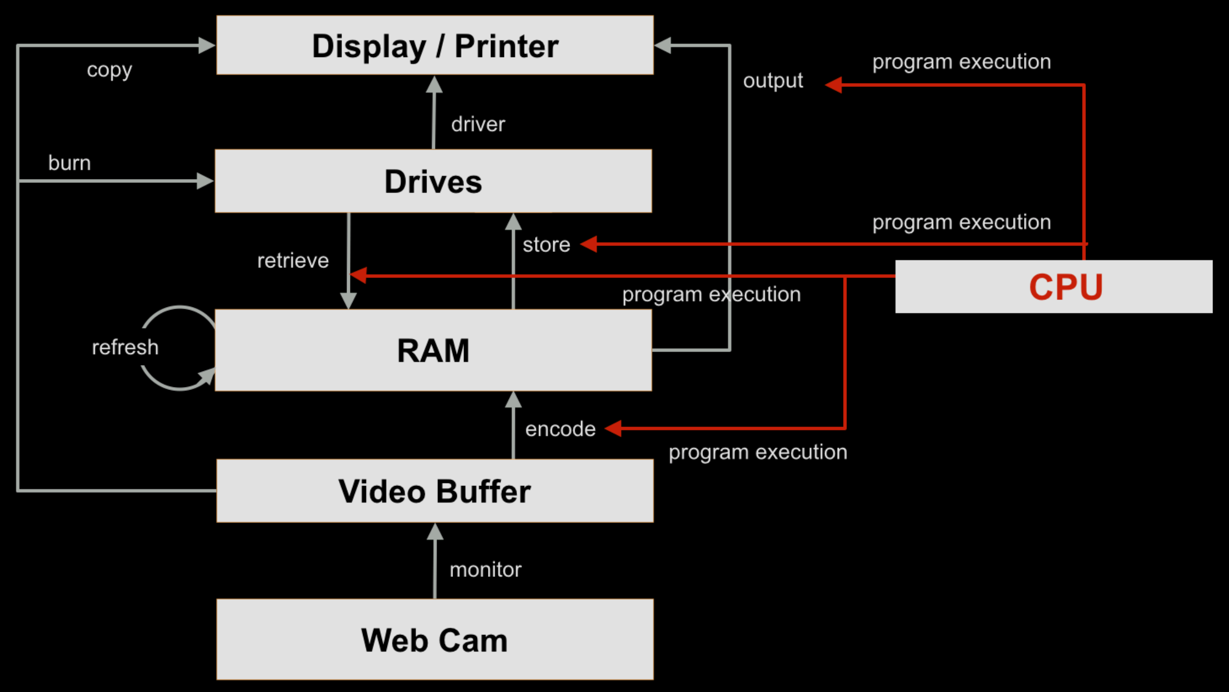
Connectionist



continuous

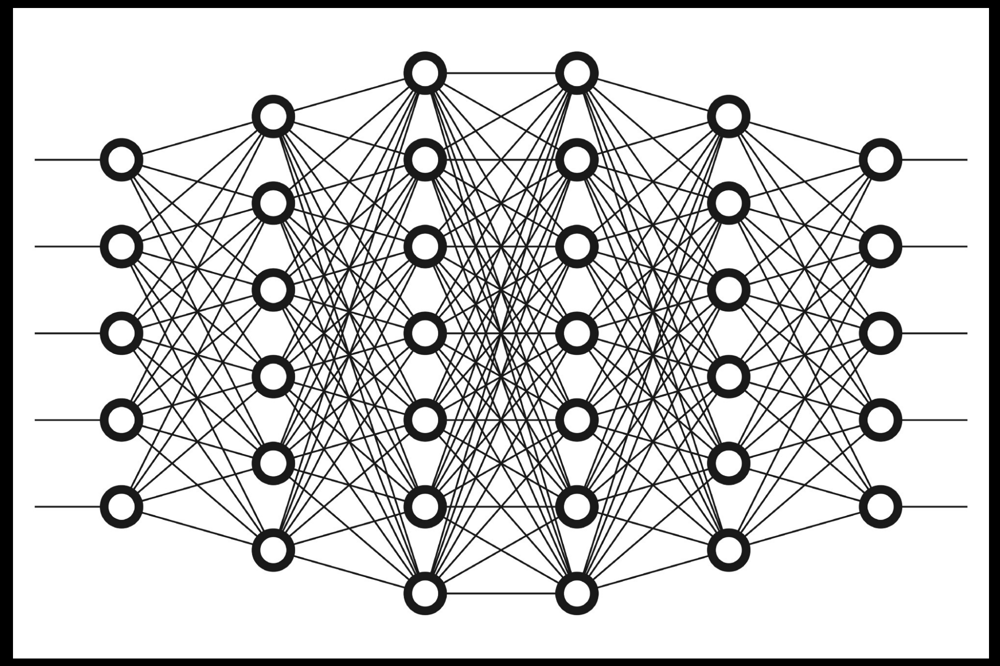
Extremes of a Continuum

Symbolic



localized

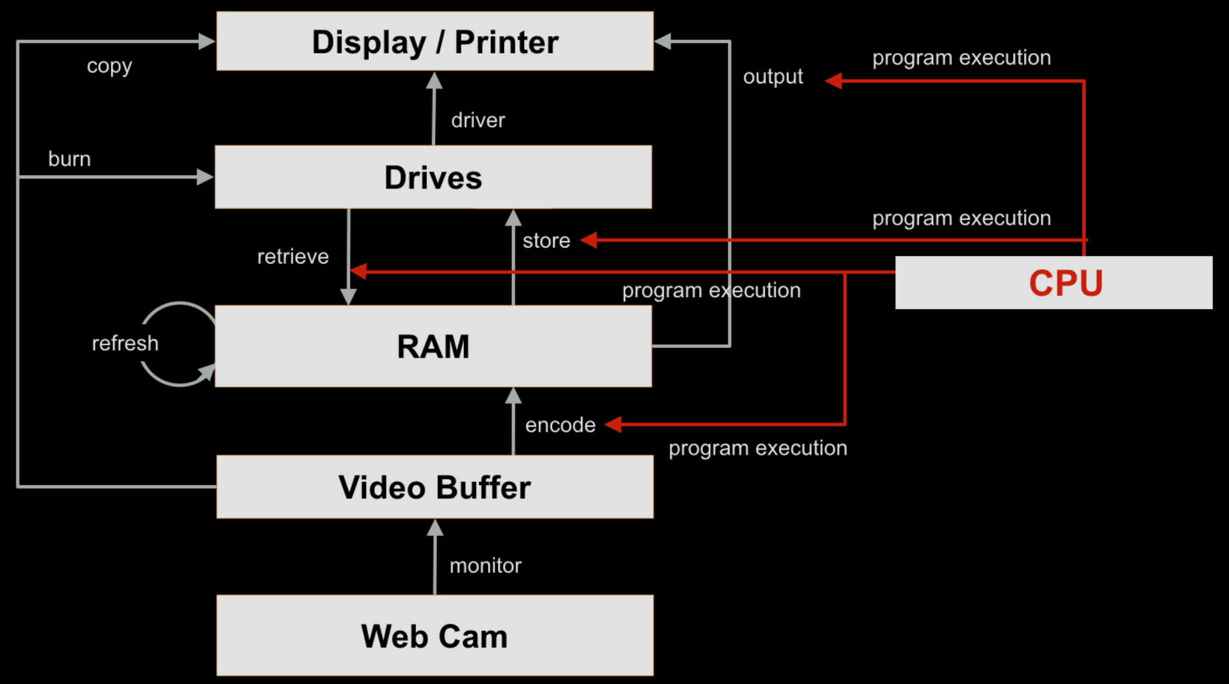
Connectionist



distributed

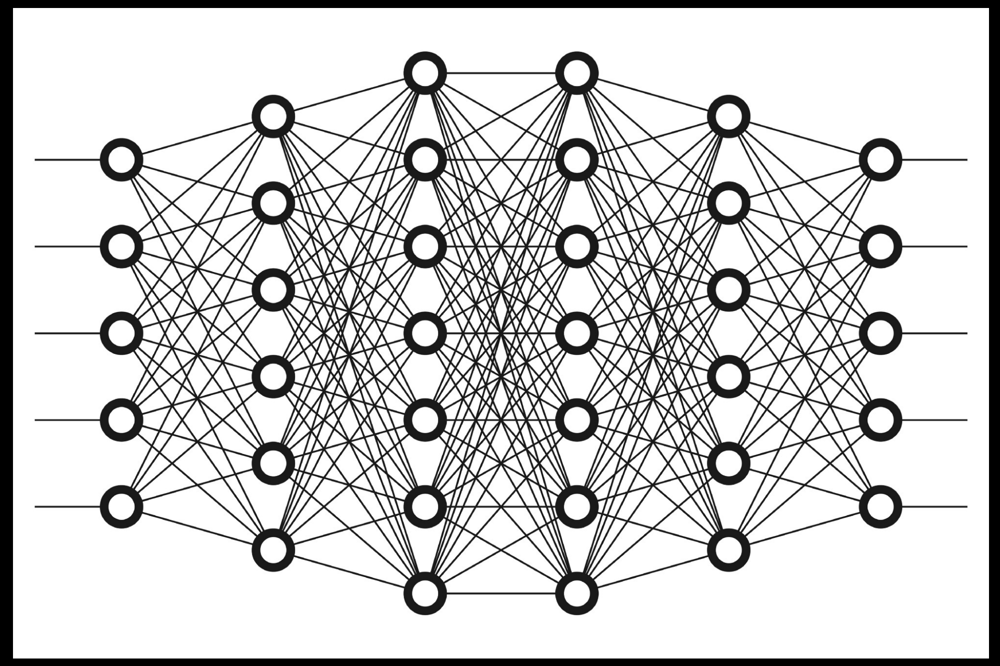
Extremes of a Continuum

Symbolic



episodic

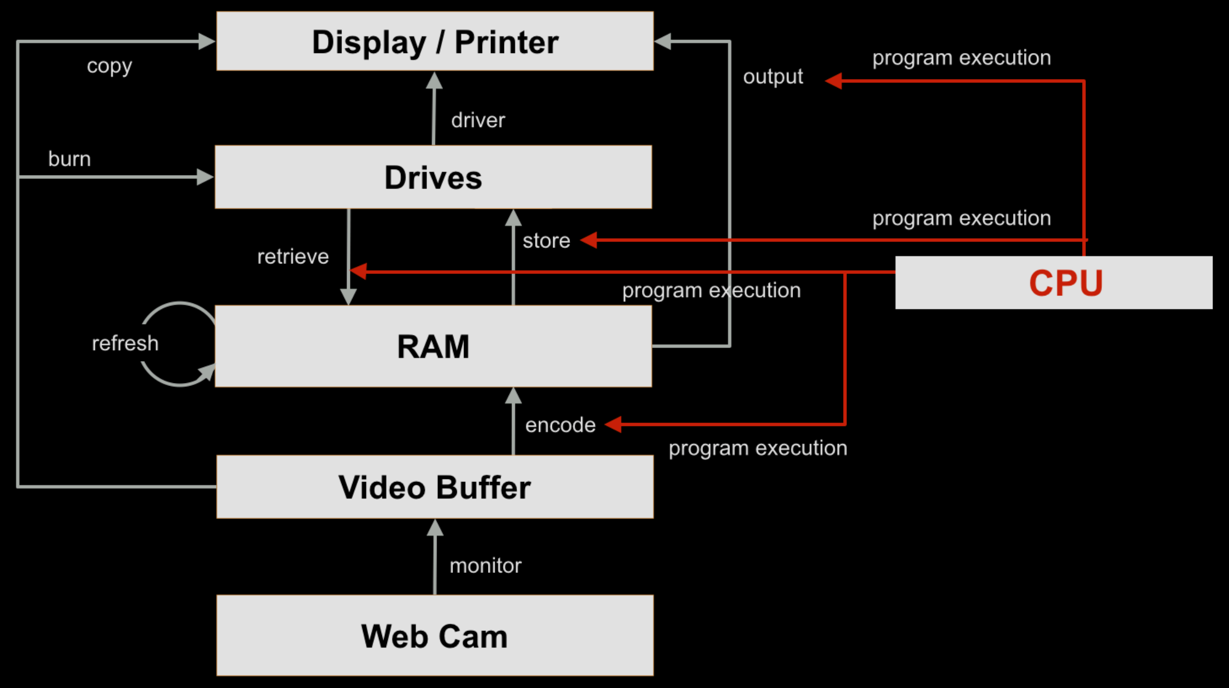
Connectionist



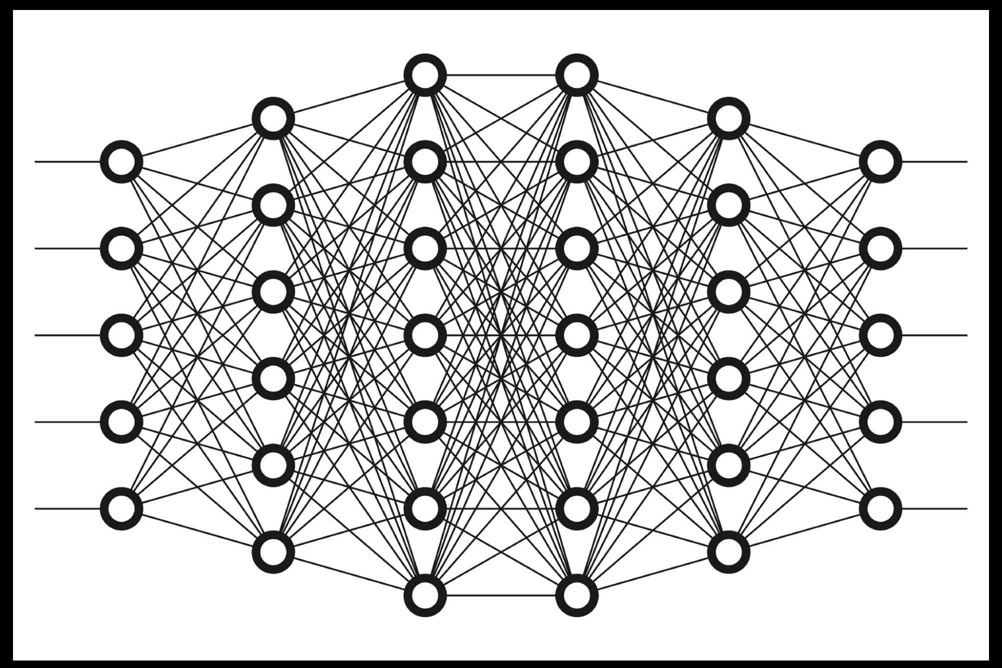
semantic

Extremes of a Continuum

Symbolic



Connectionist

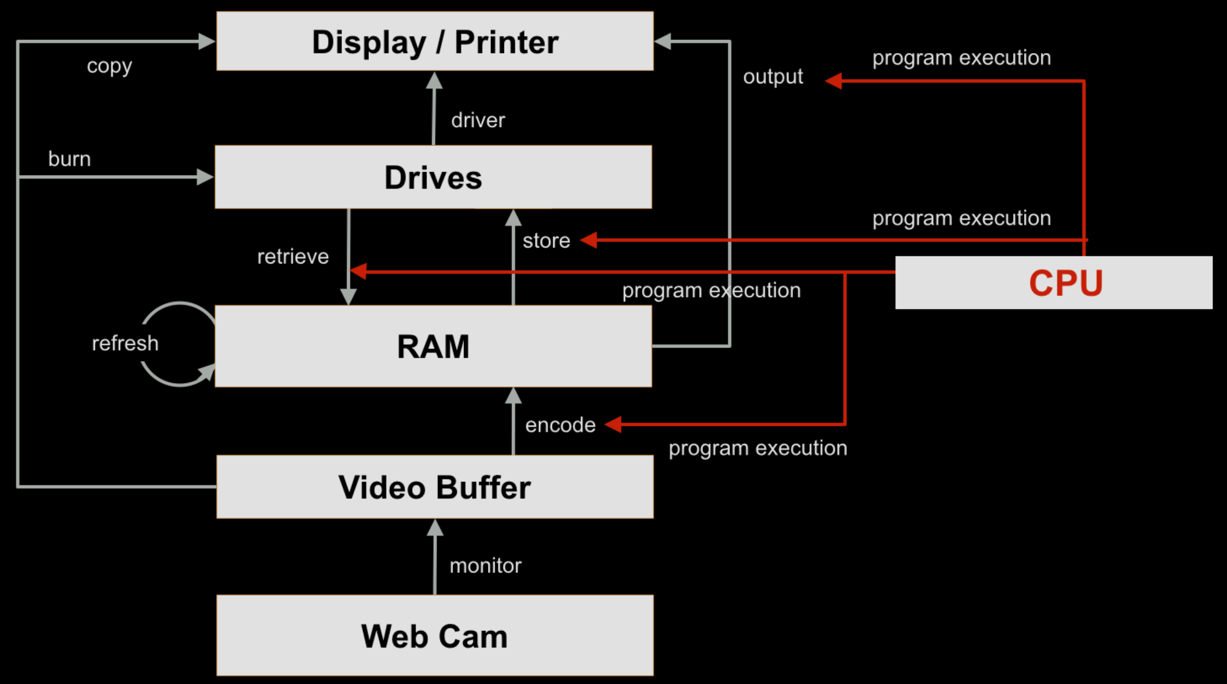


variance

bias

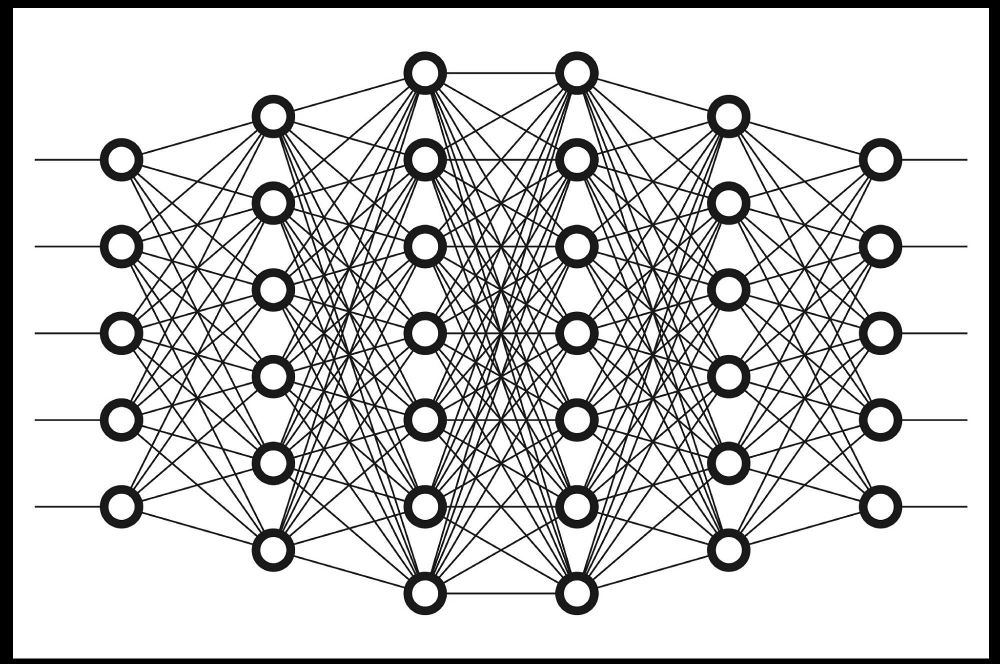
Extremes of a Continuum

Symbolic



symbol manipulation

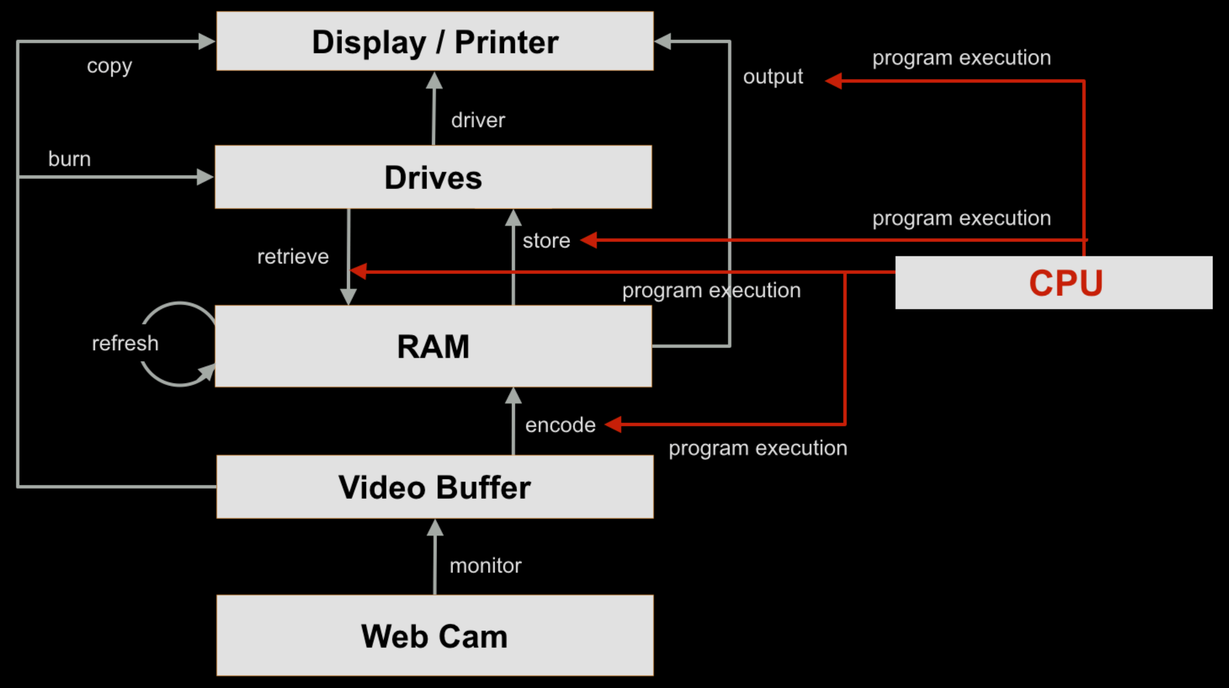
Connectionist



function approximation

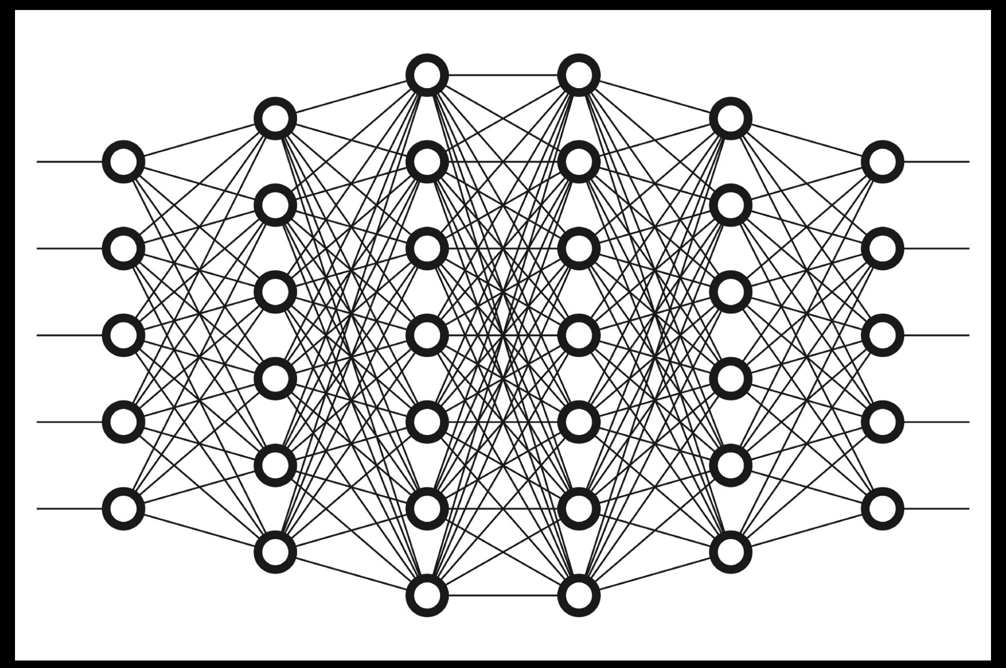
Extremes of a Continuum

Symbolic



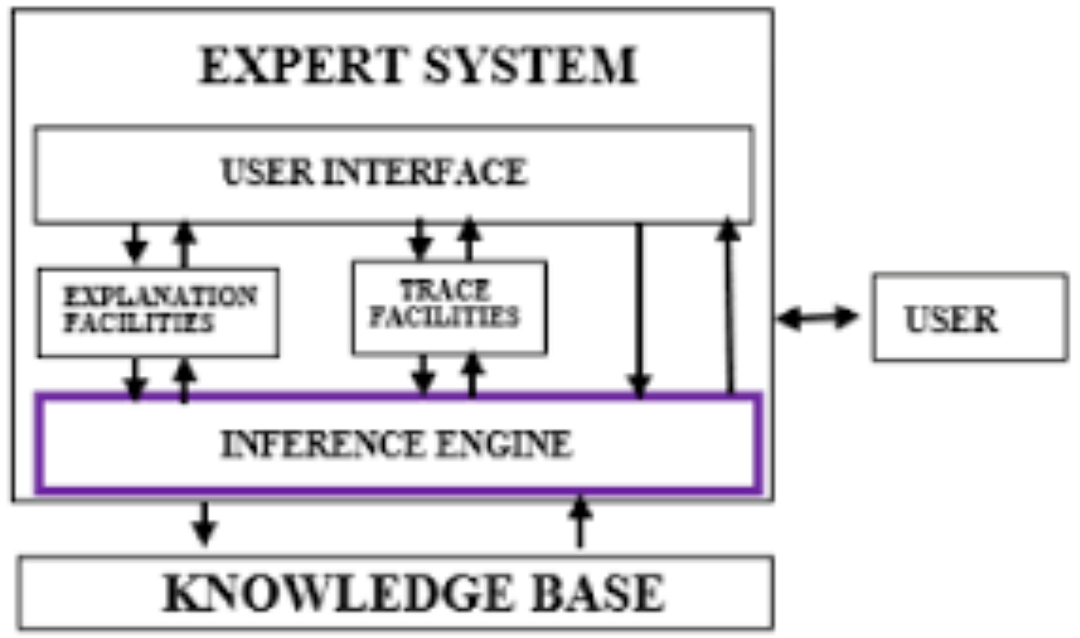
flexible

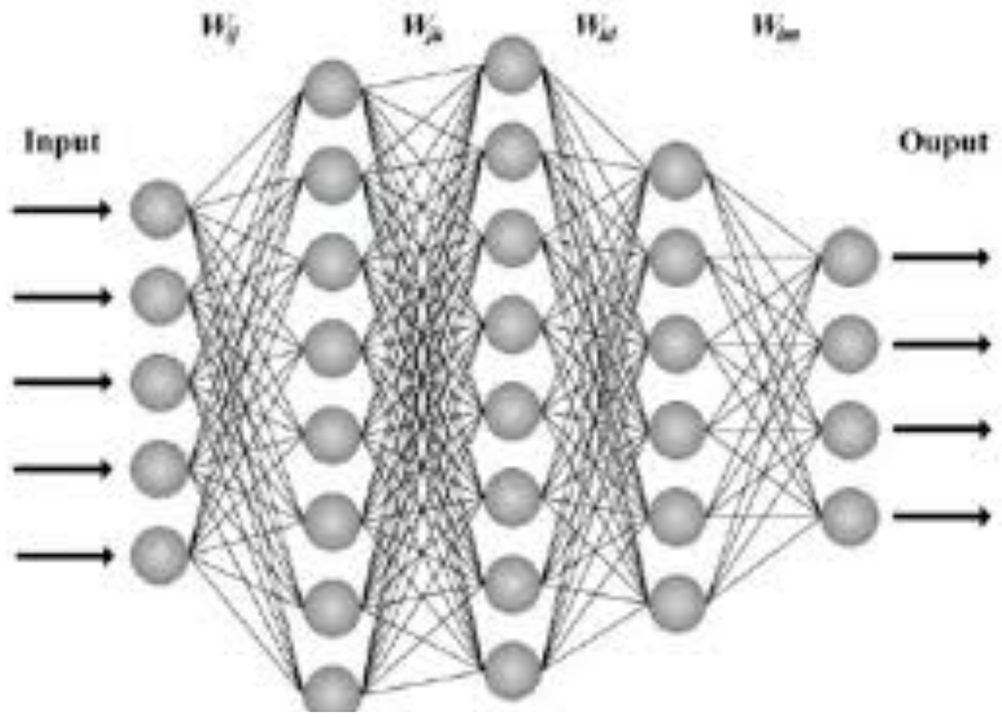
Connectionist



efficient

Artificial Intelligence





	Symbolic	**Connectionist**
Knowledge:	**explicitly represented** *expressions and procedures* ✓ *explanation* ✗ *domain specific*	**implicitly represented** *connection weights* ✓ *efficient* ✗ *domain specific*
Configuration:	**programming** ✓ *flexible* ✗ *hand-coded*	**learning** ✓ *learns from experience,* ✗ *but only when trained*

Clash of the Titans

Complementary Approaches

Complementary Approaches

- Symbolic approach
  - compositional, *context-free:*

Complementary Approaches

- **Symbolic approach**
  - **compositional,** ***context-free:***

$$
\begin{array}{lcll}
2 & + & 2 & \\
47 & + & 2 & \\
n & + & m & \leftarrow \textit{the interpretation of } m \textit{ is not affected by } n
\end{array}
$$

Complementary Approaches

- **Symbolic approach**
  - **compositional, *context-free:***

    $2 + 2$
    $47 + 2$
    $n + m$ ← *the interpretation of m is not affected by n*

- **PDP / connectionist approach**

Complementary Approaches

- **Symbolic approach**
  - **compositional, *context-free:***

    2 + 2
    47 + 2
    n + m ← *the interpretation of m is not affected by n*

- **PDP / connectionist approach**
  - ***context-sensitive*:**

    river + bank

Complementary Approaches

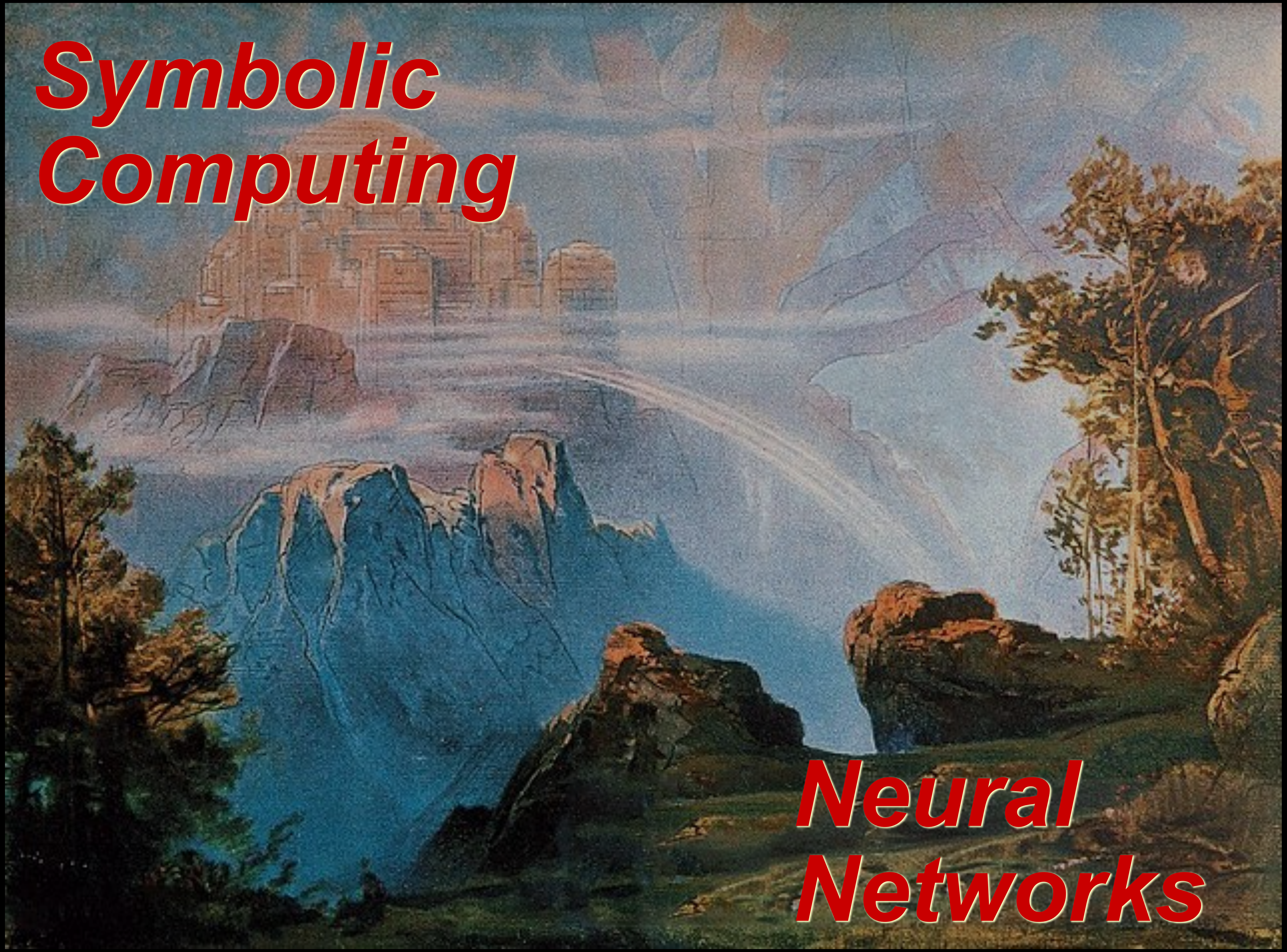
- **Symbolic approach**
  - **compositional, *context-free:***

    2 + 2
    47 + 2
    n + m ← *the interpretation of m is not affected by n*

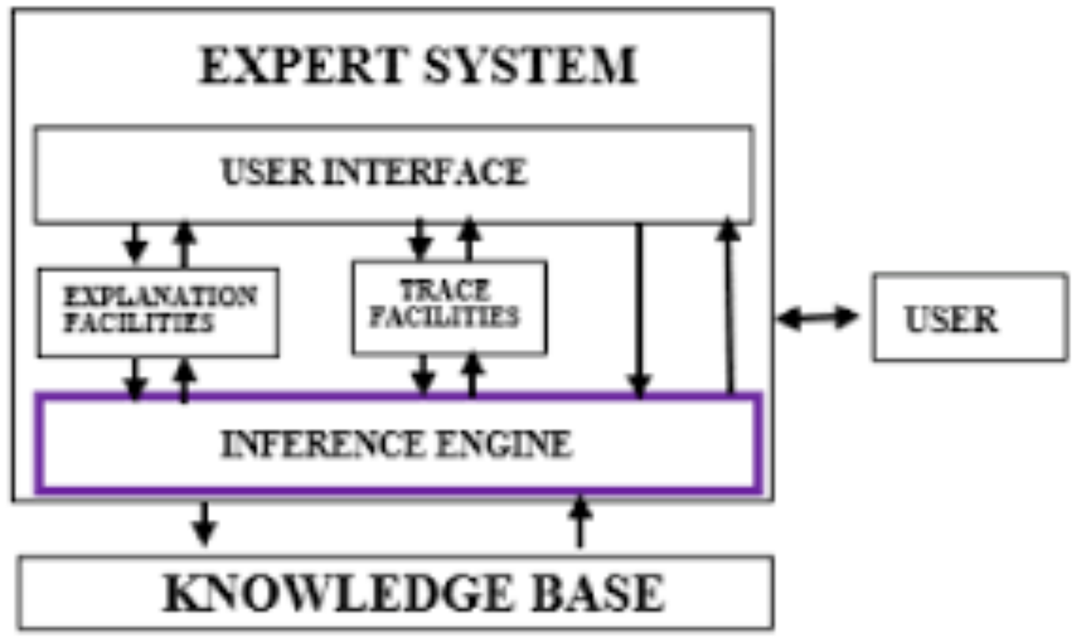
- **PDP / connectionist approach**
  - ***context-sensitive*:**

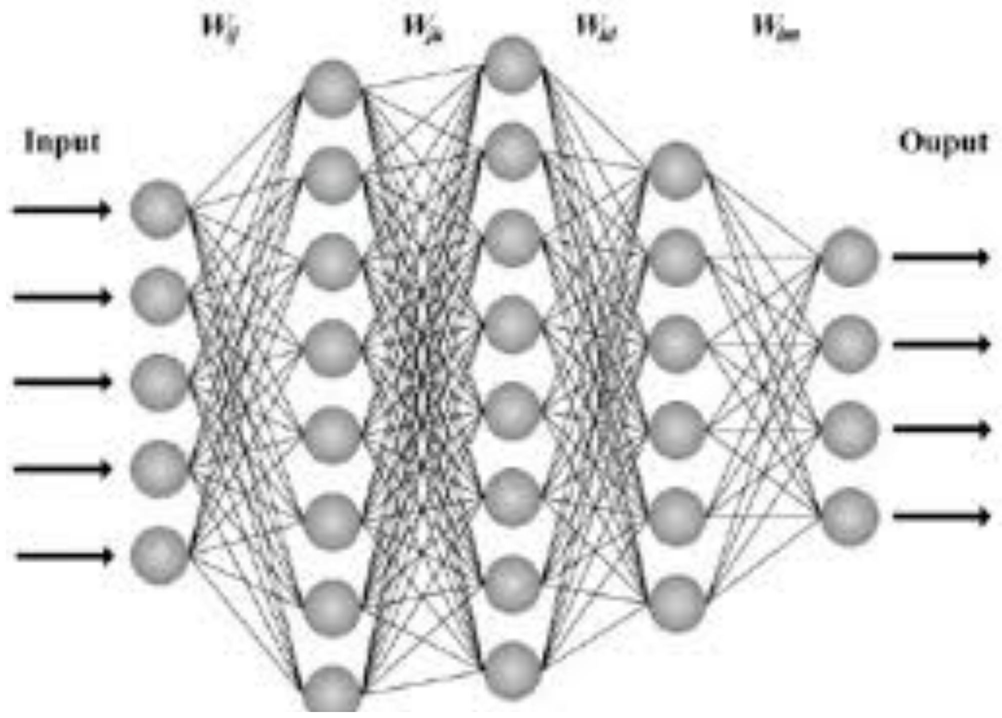
    river + bank
    savings + bank
    n + m ← *the interpretation of m depends on n*

Shangri-La



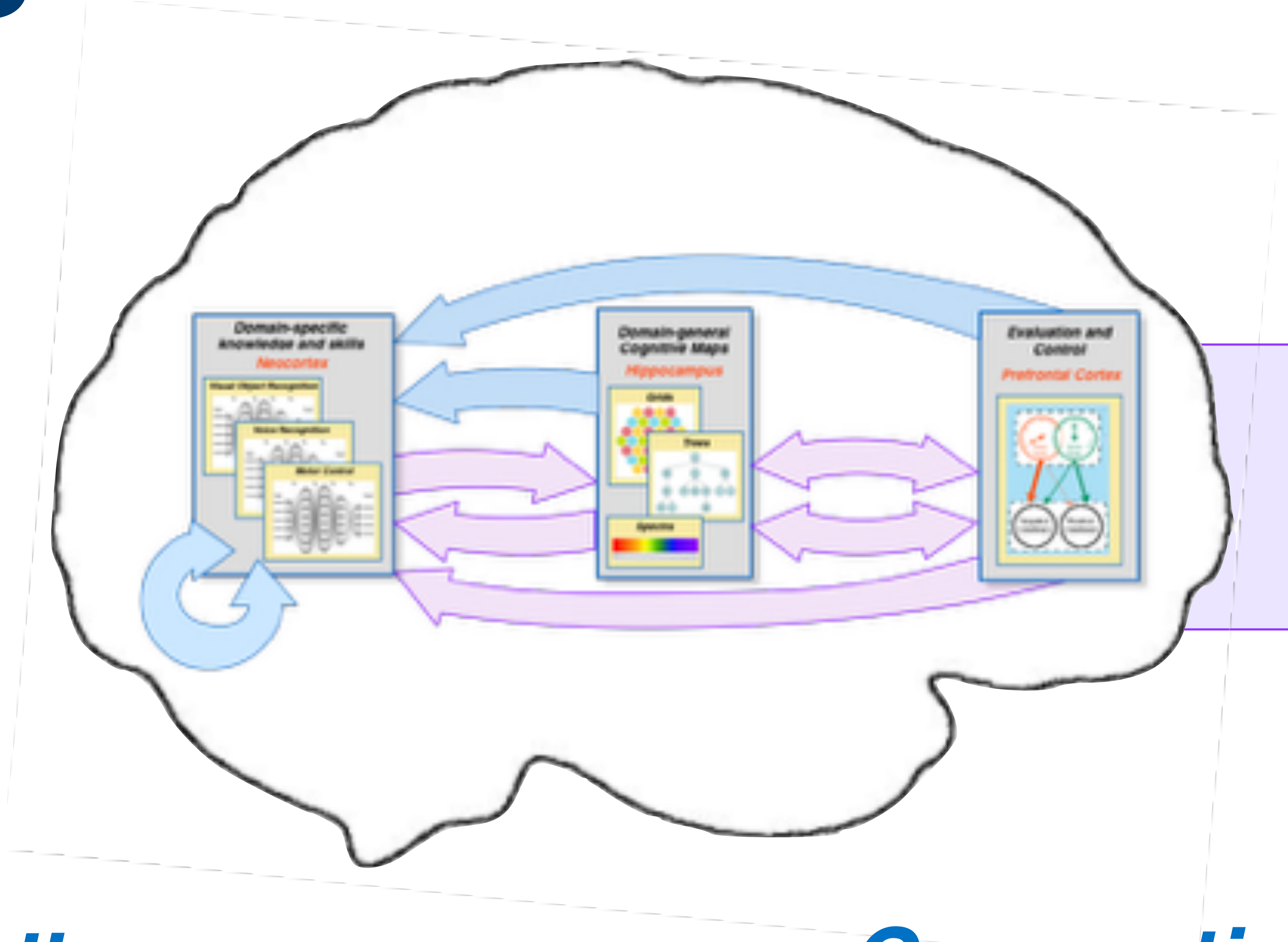
Artificial Intelligence





	<u>***Symbolic***</u>	<u>***Connectionist***</u>
Knowledge:	**explicitly represented** *expressions and procedures* ✓ *explanation* ✗ *domain specific*	**implicitly represented** *connection weights* ✓ *efficient* ✗ *domain specific*
Configuration:	**programming** ✓ *flexible* ✗ *hand-coded*	**learning** ✓ *learns from experience,* ✗ *but <u>only</u> when trained*

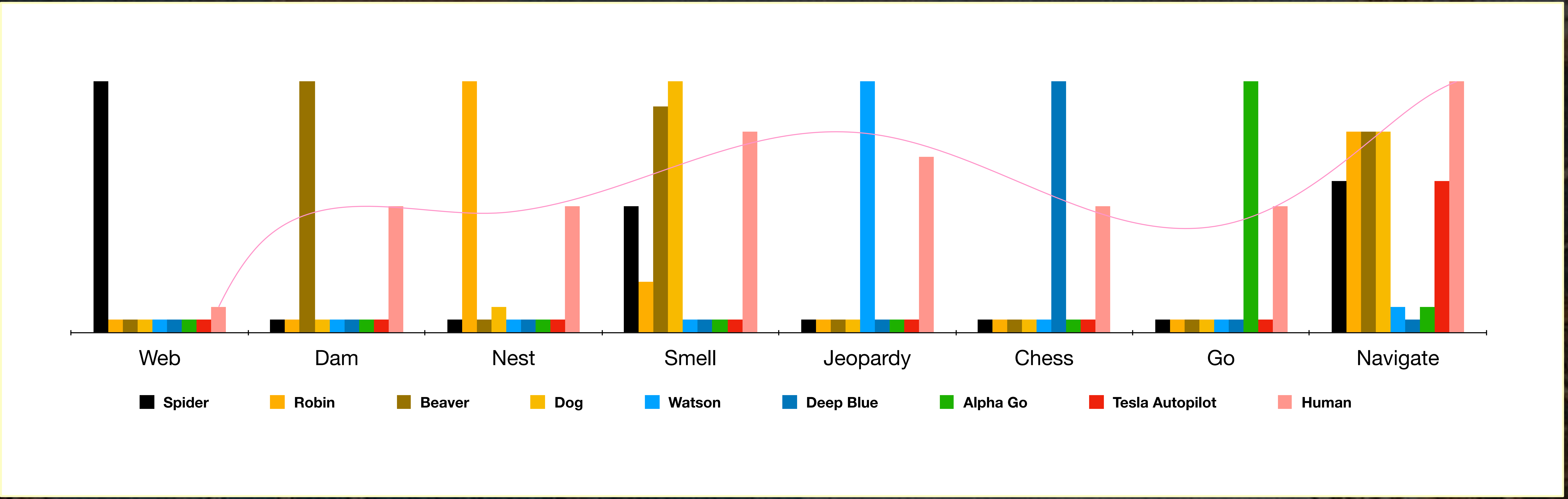
Natural Intelligence



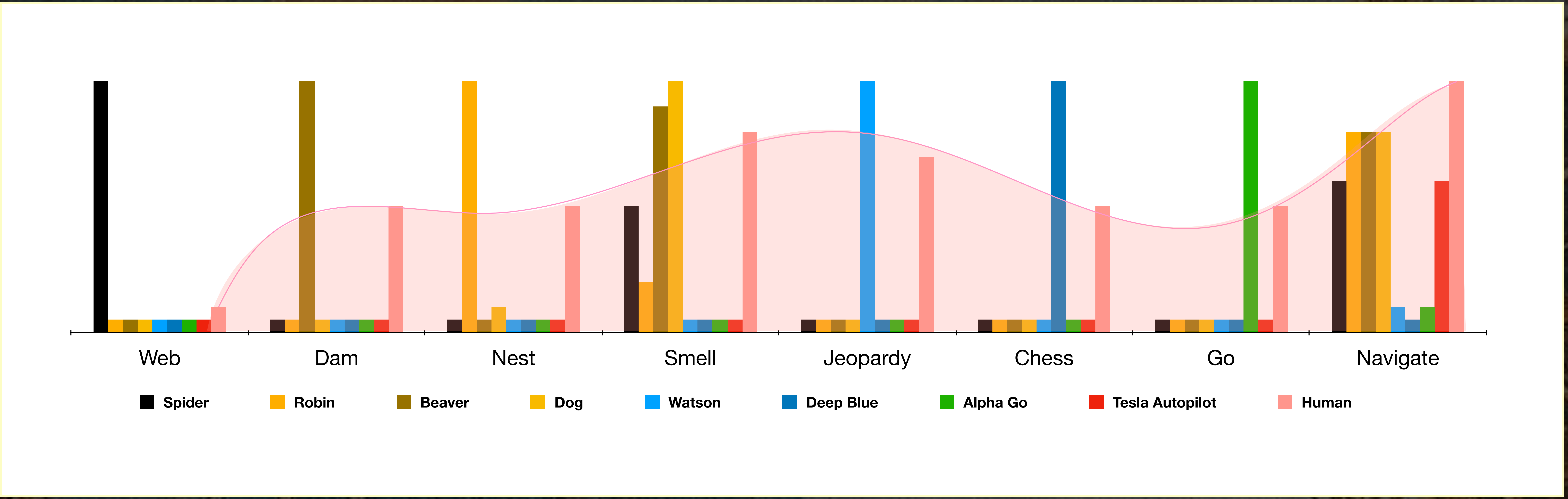
3rd Wave

	Symbolic	Connectionist	Cog Neuro / System Level Brain Modeling
***Knowledge*:**	**explicitly represented** *expressions and procedures* ✓ *explanation* ✗ *domain specific*	**implicitly represented** *connection weights* ✓ *efficient* ✗ *domain specific*	***efficient generalization***
***Configuration*:**	**programming** ✓ *flexible* ✗ *hand-coded*	**learning** ✓ *learns from experience,* ✗ *but only when trained*	***autonomous adaptation***

Human Brain = Existence Proof



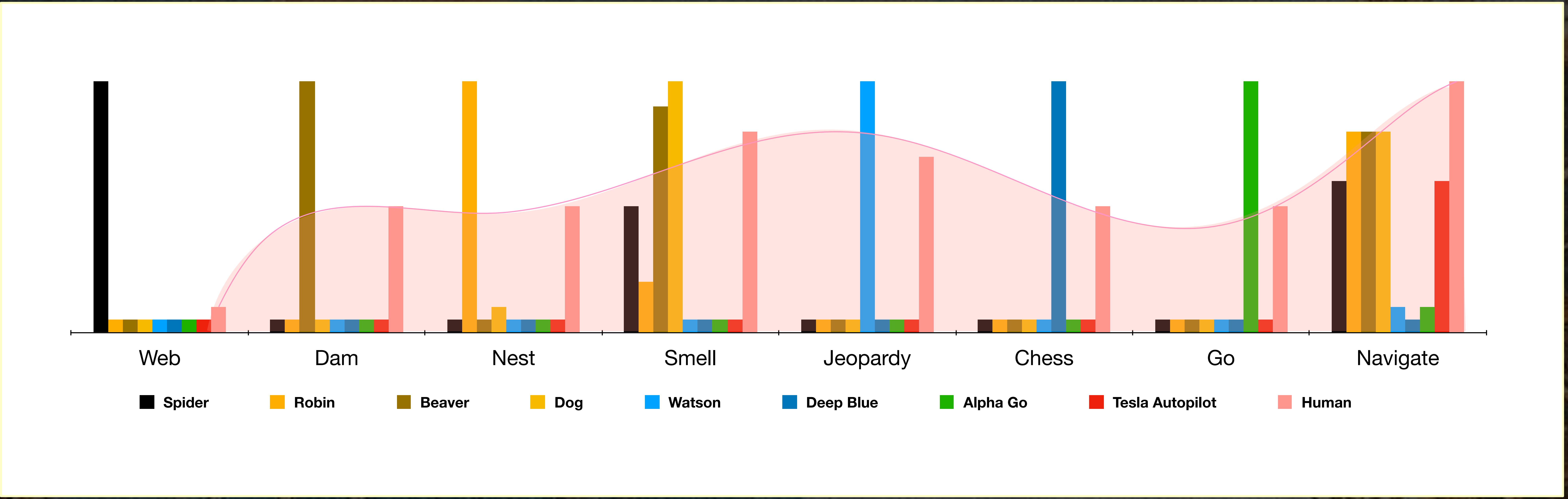
Human Brain = Existence Proof



"*Sweet spot*" between flexibility and efficiency

- Near limitless range of tasks at adequate performance - <u>flexibility</u>

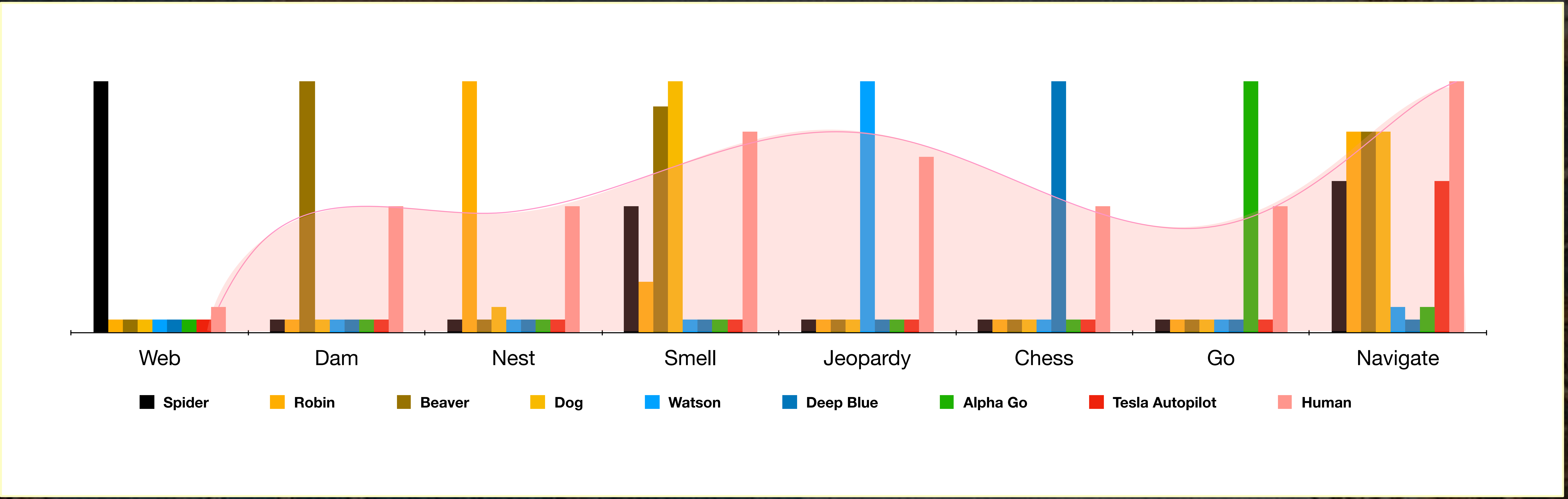
Human Brain = Existence Proof



"*Sweet spot*" between flexibility and efficiency

- With *reasonable* amounts, and often little or *no* training - sample <u>efficiency</u>

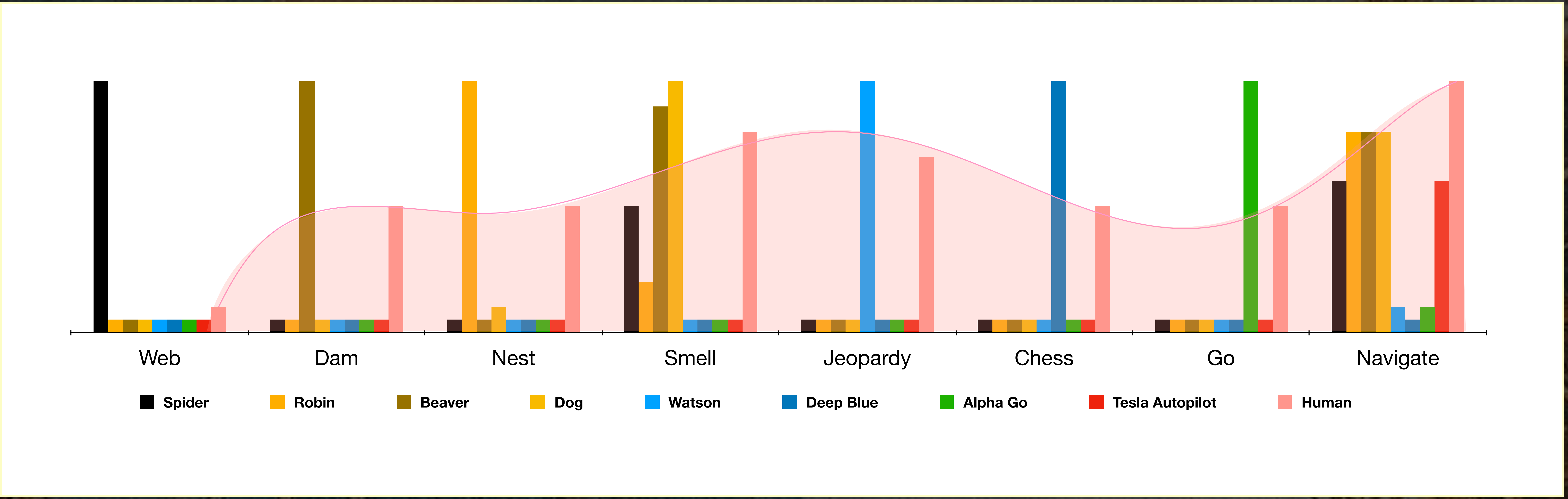
Human Brain = Existence Proof



"*Sweet spot*" between flexibility and efficiency

- ~20 watts, often with parallel performance - processing <u>efficiency</u>

Human Brain = Existence Proof



How does it accomplish this?

Shangri-La?

Shangri-La?

- **Challenge:**
  - Integrate ***flexibility*** of *symbolic processing* in traditional architectures
  - with ***efficiency*** of *function approximation* in neural networks

Shangri-La?

- **Current efforts:**
  - **Neuro-symbolic approaches:**
    - start with pre-specified **symbolic primitives** (“core knowledge”)
    - use deep learning to **combine these** (e.g., “program induction”)

Shangri-La?

- **Current efforts:**
  - **“Neo-connectionist" approaches:**
    - use deep learning for “end-to-end” training of neural networks

Shangri-La?

- **Current efforts:**
  - **"Neo-connectionist" approaches:**
    - <u>inductive biases</u> that favor <u>**abstraction**</u>
      - *training:* curricular learning, meta learning
      - *architecture & processing:* attention, external memory

Shangri-La?

- **Current efforts:**
  - **“Neo-connectionist" approaches:**
    - ♦ <u>inductive biases</u> that favor <u>**abstraction**</u>
      - *training:* curricular learning, meta learning
      - *architecture & processing:* attention, external memory
- **Still not there…**

Shangri-La?

- Still not there…
  - How does symbolic computing and the capacity for intelligence arise in the brain?

Building Bridges

1980-2000

Neurobiology

Psychology

Building Bridges

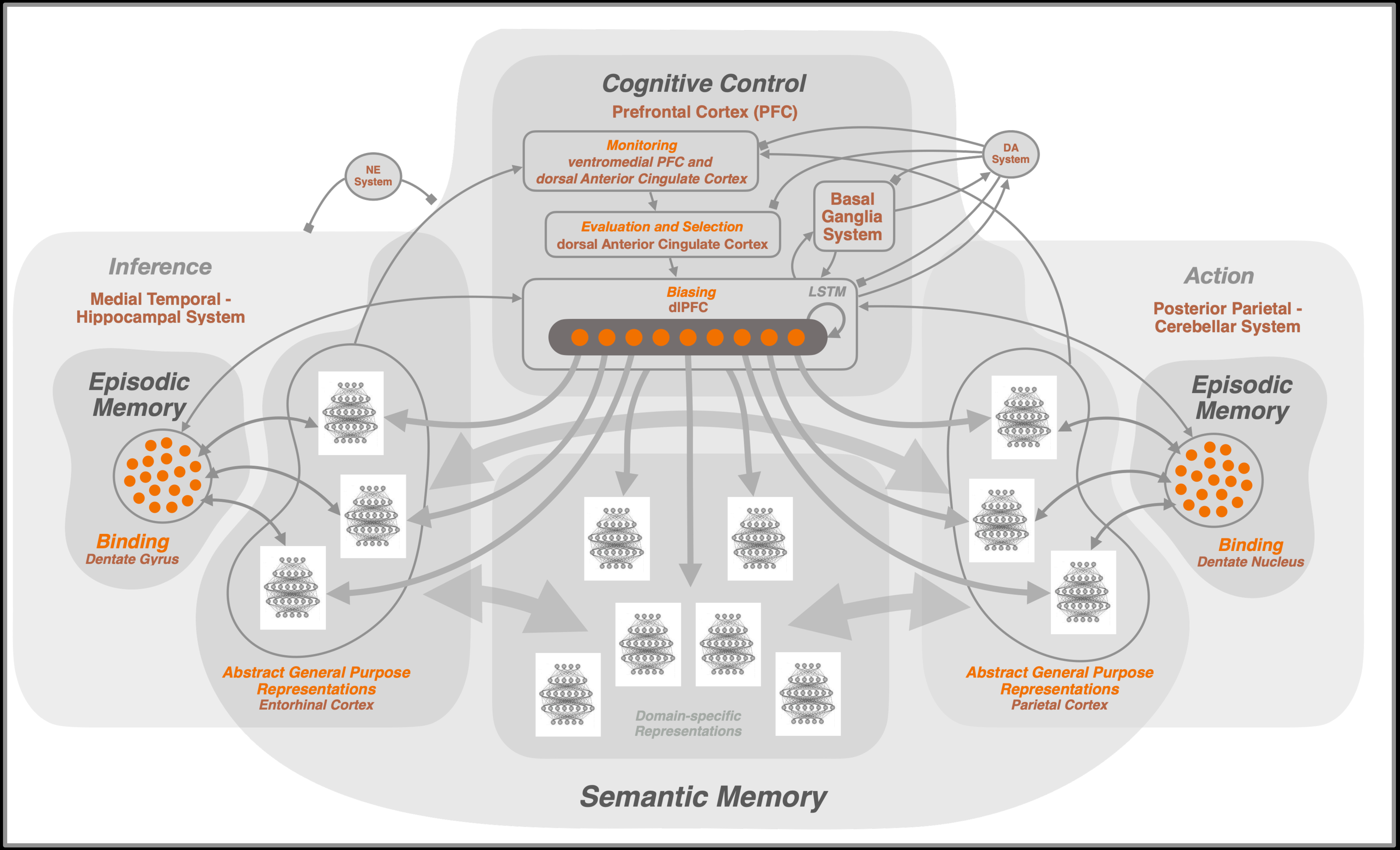
2025 →

Psychology

Neurobiology

Computer Science

2025 and beyond…



Levels of Analysis (Marr)

Levels of Analysis (Marr)

- **Computational** *(cognitive science)*
  - What is the overall goal?

Levels of Analysis (Marr)

- Computational *(cognitive science)*
  - What is the overall goal?

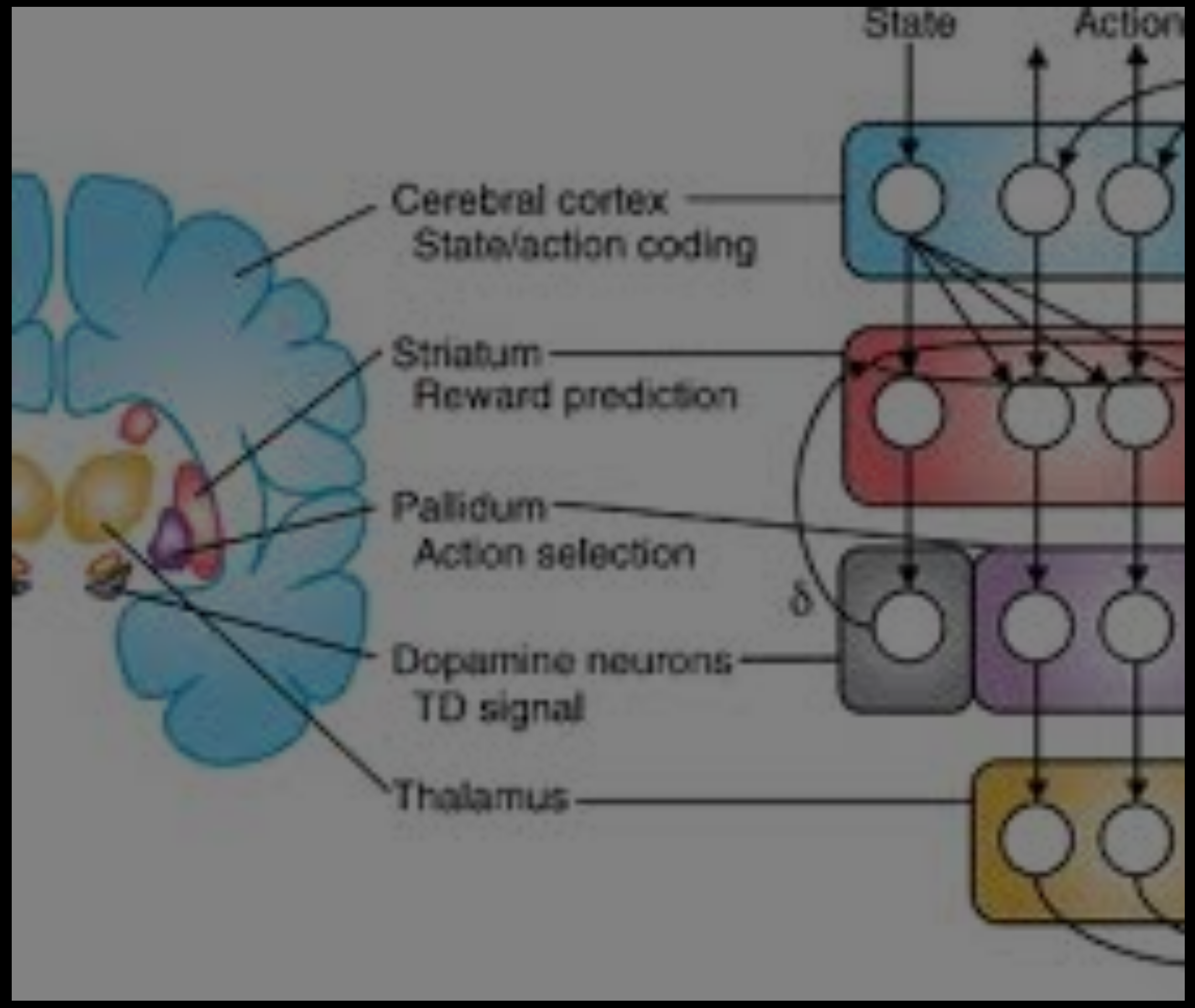
- Algorithmic *(cognitive psychology / cognitive neuroscience)*
  - What strategy is used?

Levels of Analysis (Marr)

- **Computational** *(cognitive science)*
  - What is the overall goal?

- **Algorithmic** *(cognitive psychology / cognitive neuroscience)*
  - What strategy is used?

- **Implementational** *(neuroscience)*
  - How is it physically realized?



Levels of Analysis (in reality)

Levels of Analysis (in reality)

- Behavior of a coin:

Levels of Analysis (in reality)

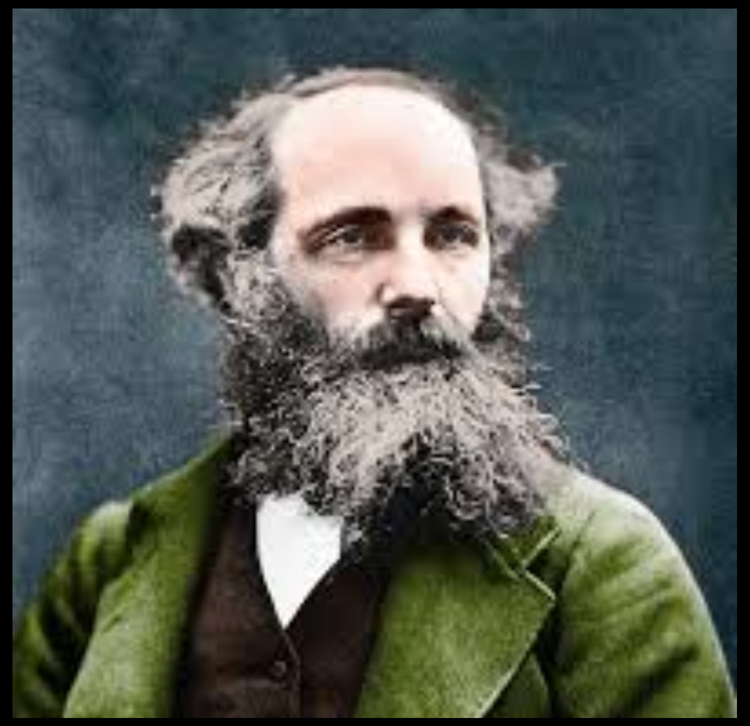
- Behavior of a coin:
  - Trajectory?

Levels of Analysis (in reality)

- Behavior of a coin:
  - Trajectory?

Newton

or



Maxwell

?

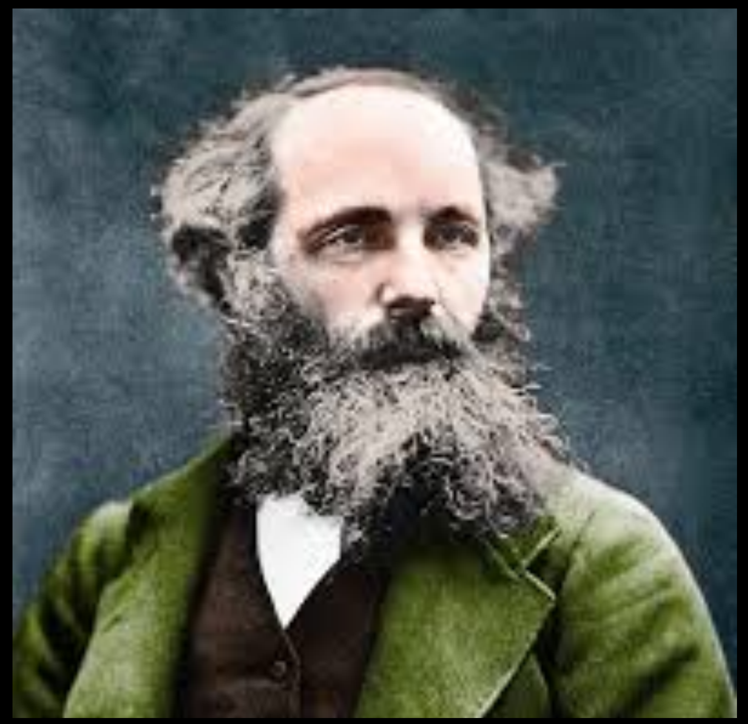
Levels of Analysis (in reality)

- Behavior of a coin:
  - Trajectory?

Newton

$$
\begin{aligned}
v &= u + at \\
v^2 &= u^2 + 2as \\
s &= ut + \tfrac{1}{2}at^2
\end{aligned}
$$

or



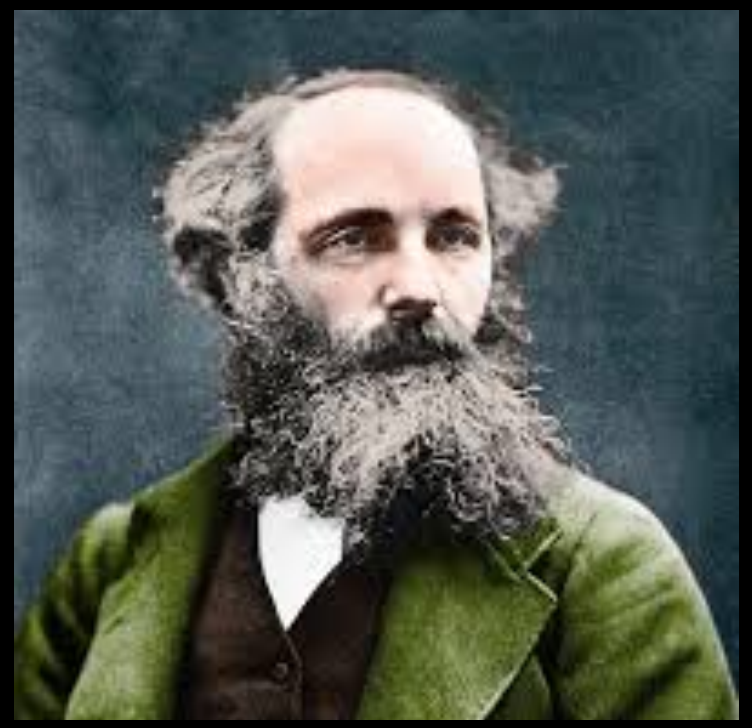
Maxwell

Levels of Analysis (in reality)

- Behavior of a coin:
  - Trajectory?
  - Melting point?

Newton

or



Maxwell

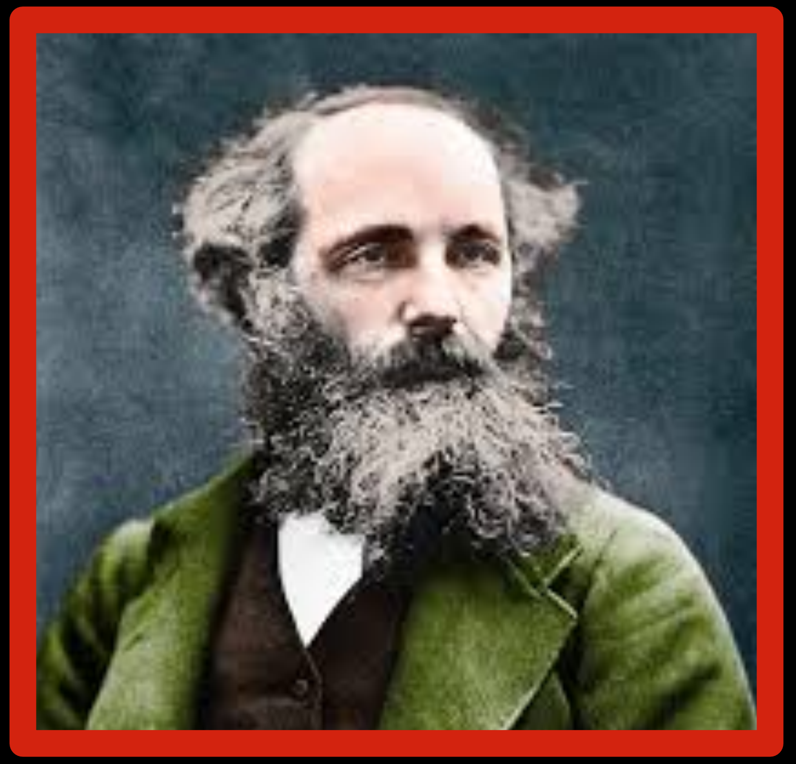
?

Levels of Analysis (in reality)

- Behavior of a coin:
  - Trajectory?
  - Melting point?

Newton

or



Maxwell

$$\Gamma_{ij} = \left(\delta_{ij} + x_i \frac{\partial \ln \gamma_i}{\partial x_j} \right)$$

Levels of Analysis (in reality)

Levels of Analysis (in reality)

- **“More is different”** *(Anderson, 1972)*

Levels of Analysis (in reality)

- “More is different” *(Anderson, 1972)*
- Emergent properties

Levels of Analysis (in reality)

- "More is different" *(Anderson, 1972)*
- Emergent properties
- Instrumentalism:
  - what is the observational variance you wish to capture?
  - what is the best way to capture that?

Levels of Organization

Physical

- Organism
  - Animal
- Organ
  - Brain
- Components
  - Lobes
- Maps & Zones
  - Areas, layers & columns
- Cells
  - Neurons
- Organelles
  - Membranes & Synapses
- Molecules
  - Transmitters & Receptors

Functional

Behavior

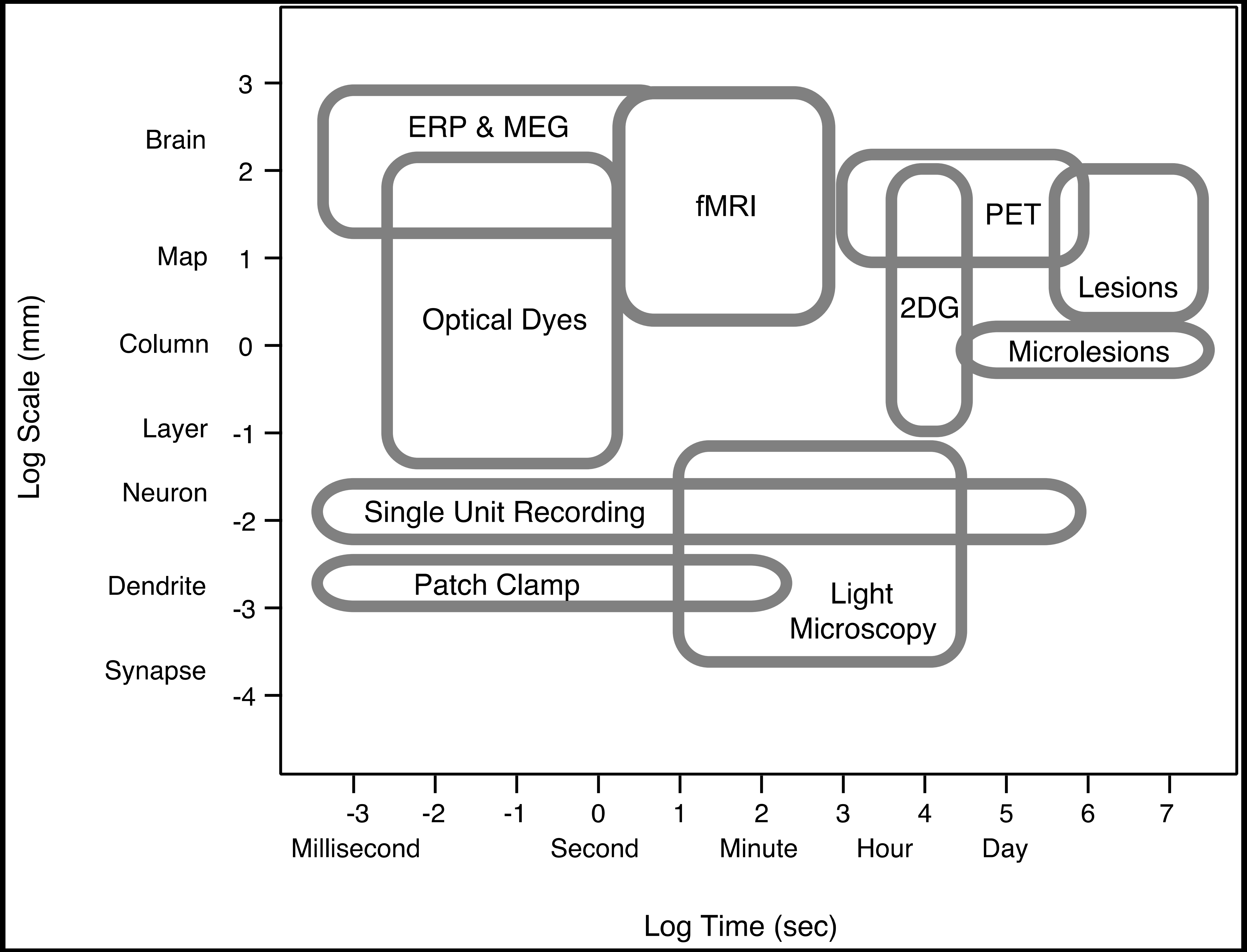
Computation

Functions

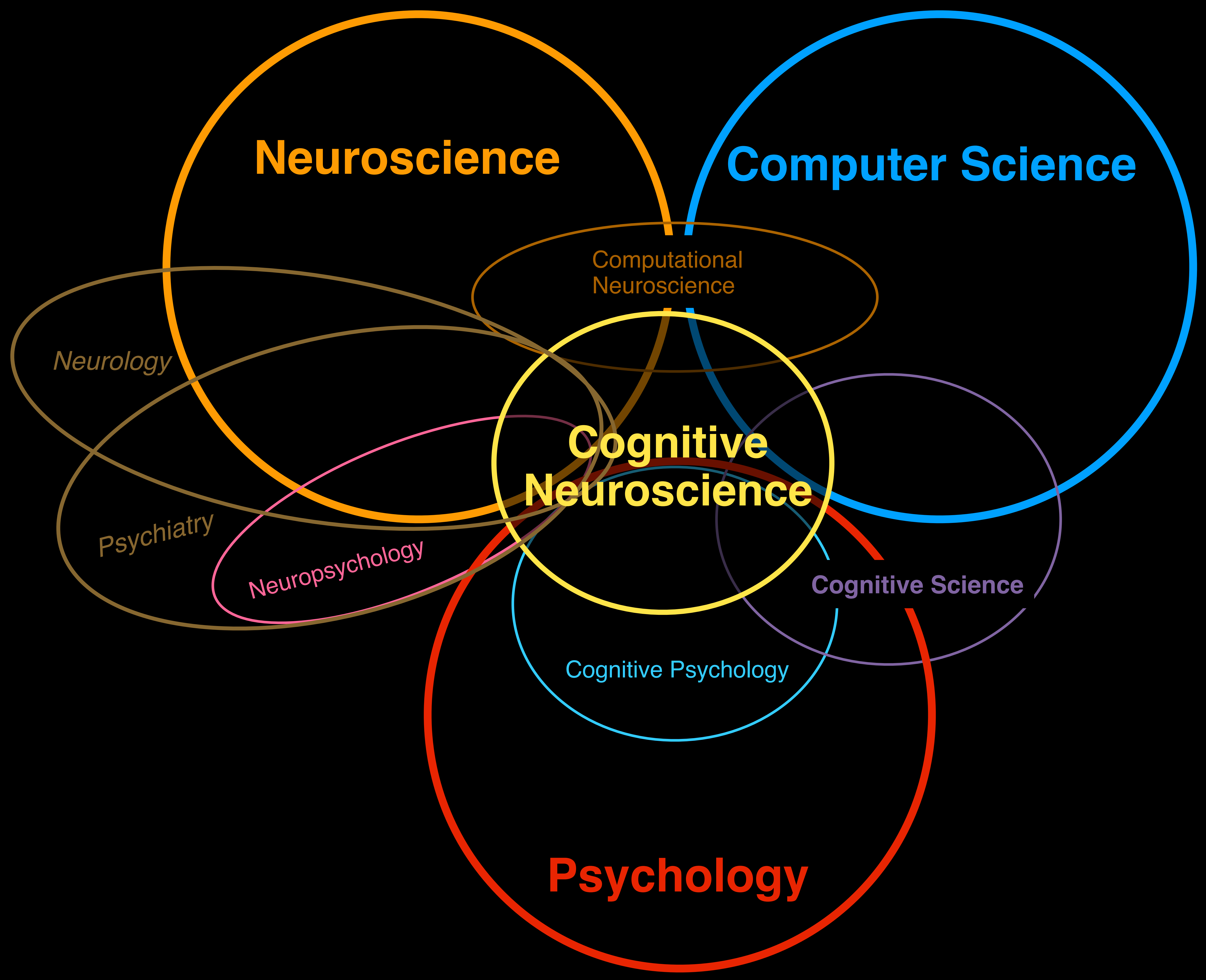
Processes / Representations

Processing units

Levels of Observation



Topography of Fields



Forms of Formalism

Mathematical	Computational
Formulas	Programs
Solutions	Simulations
Simple	Realistic
Precise	Accurate

Contrasts and Tradeoffs

Contrasts and Tradeoffs

- speed vs. accuracy

Contrasts and Tradeoffs

- parallel vs. serial

Contrasts and Tradeoffs

- compiled vs. interpreted

Contrasts and Tradeoffs

- exploit vs. explore

Contrasts and Tradeoffs

- durability vs. accessibility

Contrasts and Tradeoffs

- deterministic vs. statistical

Contrasts and Tradeoffs

- hierarchical vs. heterarchical

Contrasts and Tradeoffs

- programming vs. learning

Contrasts and Tradeoffs

- engineering vs. evolution

Contrasts and Tradeoffs

- hardware vs. software

Contrasts and Tradeoffs

- centralization vs. autonomy

Contrasts and Tradeoffs

- flexibility vs. efficiency…

	Flexibility							
	Structural				**Compositional**			
	Within Domains		**Across Domains**		**Within Domains**		**Across Domains**	
	Acquisition	**Processing**	**Acquisition**	**Processing**	**Acquisition**	**Processing**	**Acquisition**	**Processing**
Traditional Computers	Instruction	Deductive Inference / Symbols	Instruction	Deductive Inference / Symbols	Instruction	Deductive Inference / Symbols	Instruction	Deductive Inference / Symbols
Neural Networks	Learning	Function Approximation	None	None	Learning	Function Approximation	None	None
Non-humans	Learning	Function Approximation	None	None	Learning	Function Approximation	None	None
Humans	Learning Instruction	Deductive Inference / Function Approximation	Learning Instruction	Deductive Inference / Function Approximation	Learning Instruction	Deductive Inference / Function Approximation	Learning Instruction	Deductive Inference / Function Approximation

	Efficiency			
	Time		**Energy**	
	Acquisition	**Processing**	**Acquisition**	**Processing**
Symbol Processing	Instruction	Moore’s Law Deductive Inference Independent Parallelism	Instruction	Moore’s Law
Neural Networks	Reinforcement Learning Curruicular Learning Metalearning	Moore’s Law Independent Parallelism Interactive Parallelism	Learning	Moore’s Law
Non-humans	Unsupervised Learning Reinforcement Learning	Independent Parallelism Interactive Parallelism	Learning	Biological Computation
Humans	Unsupervised Learning Reinforcement Learning Curricular Learning Metalearning Active Learning Instruction	Deductive Inference Independent Parallelism Interactive Parallelism	Learning Instruction	Biological Computation

The Bitter Lesson

Rich Sutton, ~2020

“The second general point to be learned from the bitter lesson is that the actual contents of minds are tremendously, irredeemably complex; we should stop trying to find simple ways to think about the contents of minds, such as simple ways to think about space, objects, multiple agents, or symmetries. All these are part of the arbitrary, intrinsically-complex, outside world. They are not what should be built in, as their complexity is endless; **instead we should build in only the meta-methods that can find and capture this arbitrary complexity.** Essential to these methods is that they can find good approximations, but ***the search for them should be by our methods, not by us. We want AI agents that can discover like we can, not which contain what we have discovered.*** Building in our discoveries only makes it harder to see how the discovering process can be done.“

(http://www. incompleteideas.net/IncIdeas/BitterLesson.html)

"Unexplainable AI"

Turing, 1950

"We also wish to allow the possibility than an engineer or team of engineers may construct a machine which works, but whose manner of operation cannot be satisfactorily described by its constructors because they have applied a method which is largely experimental. "