Biological Inspiration

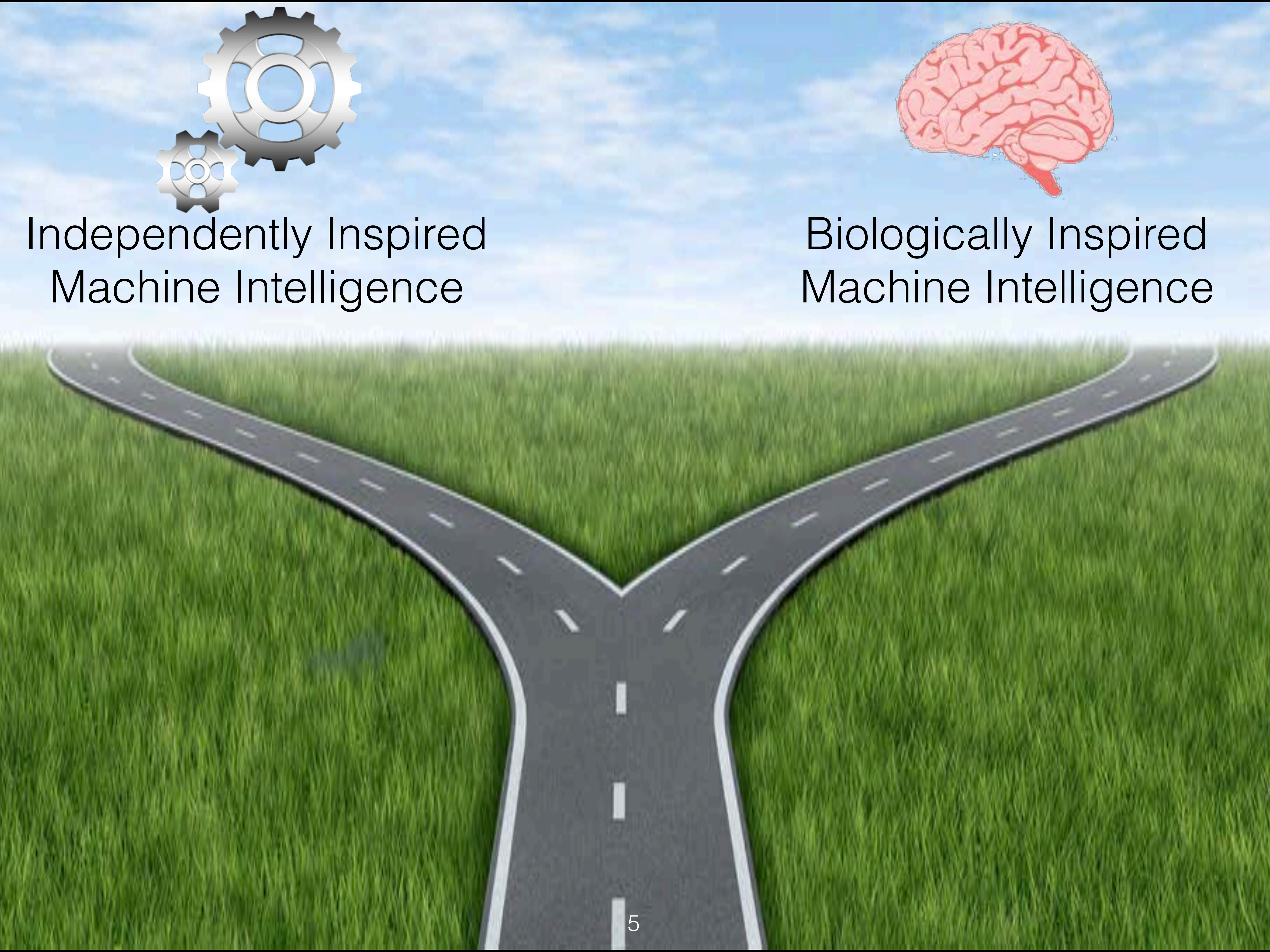
Joe Marino

what is general machine intelligence?

the ability to perform a comprehensive set of ‘intelligent’ tasks

- identifying high level concepts from environmental sensory signals (e.g. objects, words, etc.)
- planning and interacting intelligently in the environment (e.g. walking, driving, talking, etc.)
- remembering recent and past events
- reasoning about high level concepts (logic, math, etc.)

we want to apply the computational scalability of machines to the natural world what will it take to develop general machine intelligence?


Independently Inspired
Machine Intelligence
Biologically Inspired
Machine Intelligence

Marr’s Three Levels

- **Computational**
  - what is the goal of the computation?
- **Algorithmic**
  - what is the strategy to achieve it?
- **Implementational**
  - what is the design of such a system?

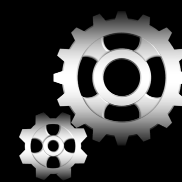
David Marr, 1982

Marr’s Three Levels

- **Computational**
  - object recognition
- **Algorithmic**
  - visual filters, pooling
- **Implementational**
  - biological neurons

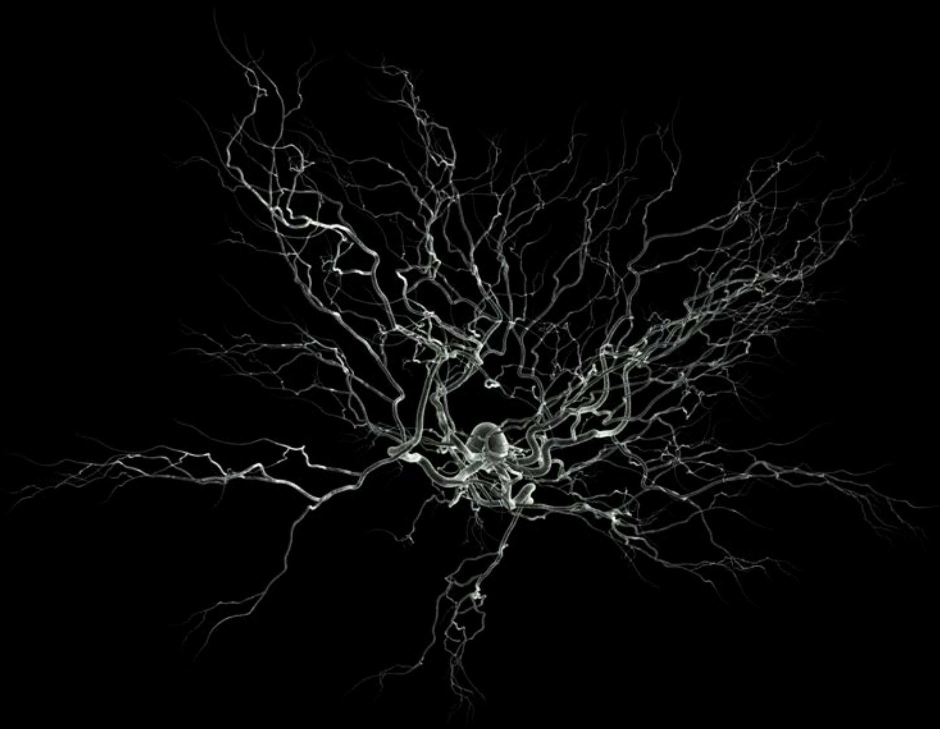
David Marr, 1982

Marr’s Three Levels

- **Computational**  
- **Algorithmic** 
- **Implementational**

Biological Inspiration

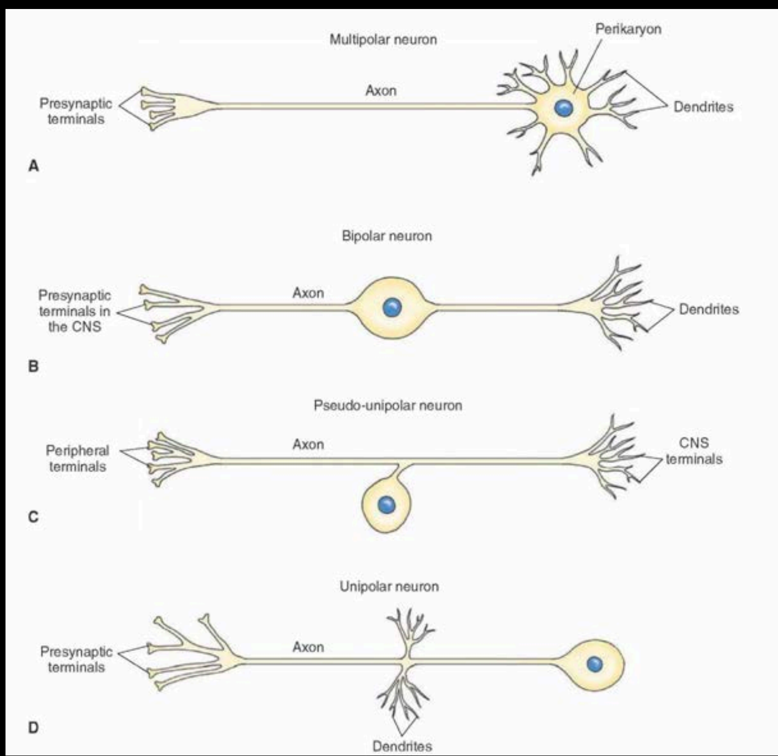
- try to use biological intelligence as a proof of concept model for machine intelligence
- similar algorithmic interpretation, different implementation (hardware/software vs. wetware)
- can make use of evolution's insights
- machine intelligence may be too difficult to develop independently, or may end up leading to same result but through more effort

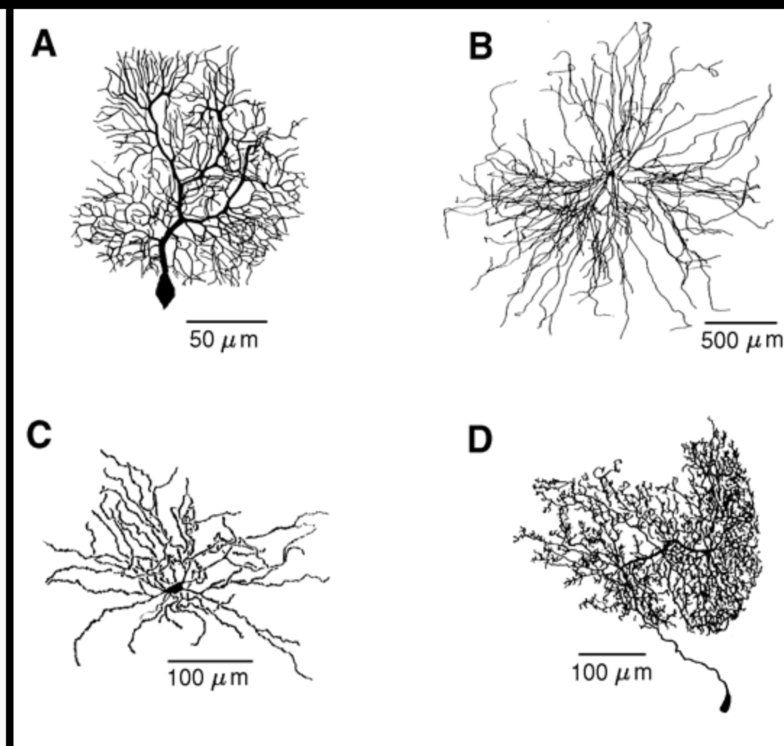


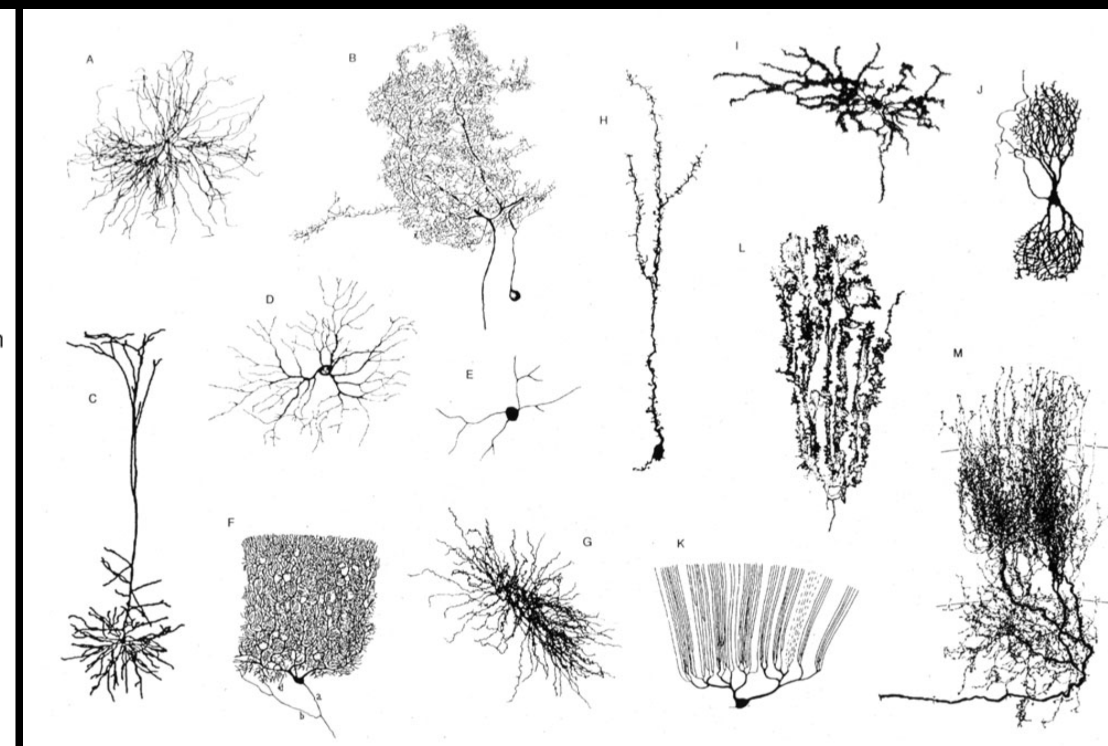
specialized units: neurons

specialized structures: brain stem nuclei, cortex, sensory organs, etc.

Biological Neurons



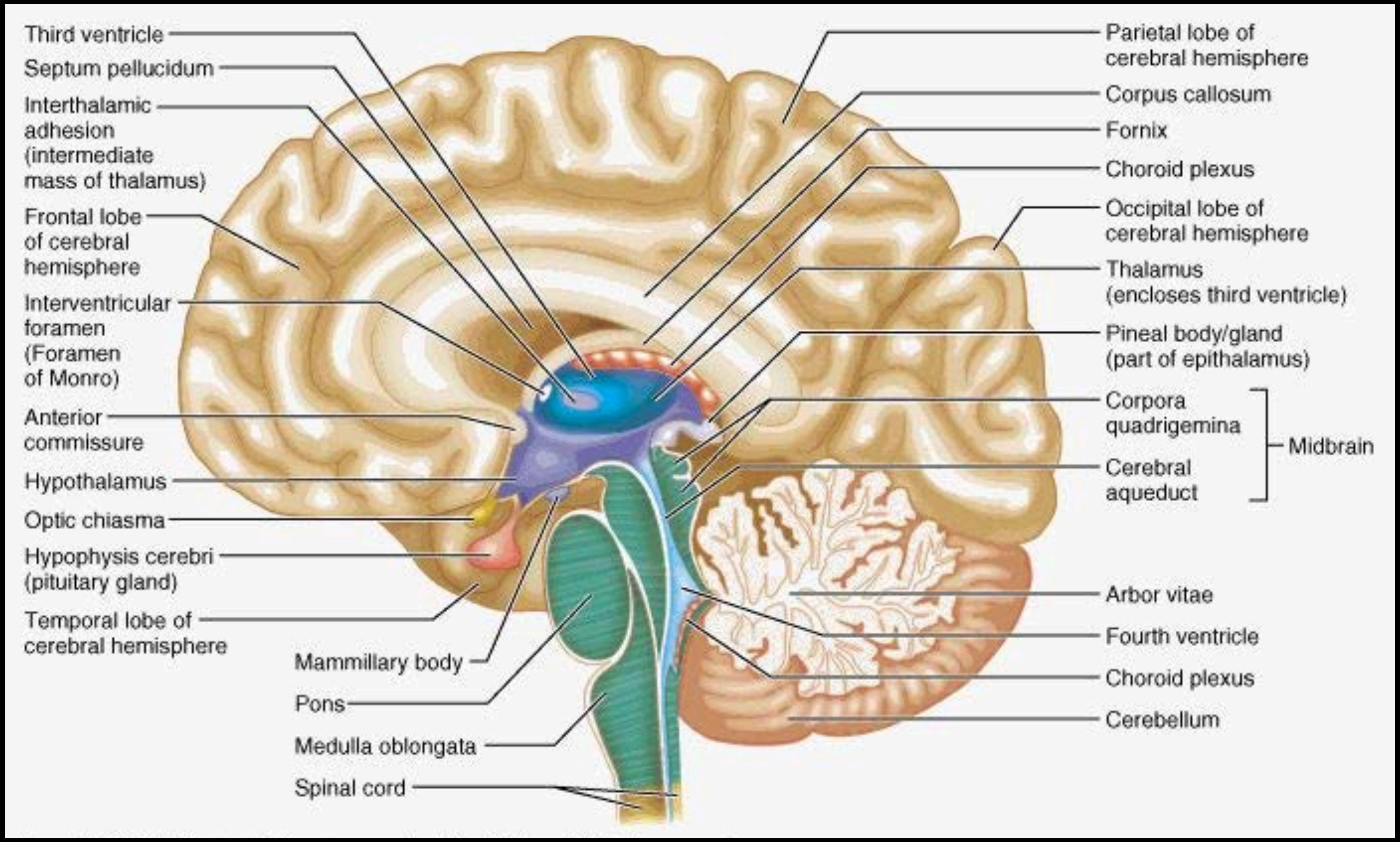




at least ***hundreds*** of distinct neuron types

roughly ***100 billion*** neurons in the human brain

Biological Structures



the central nervous system contains many morphologically and functionally specialized structures

Biological Intelligence

Biological intelligence is not random. A brain is more than just a collection of neurons.

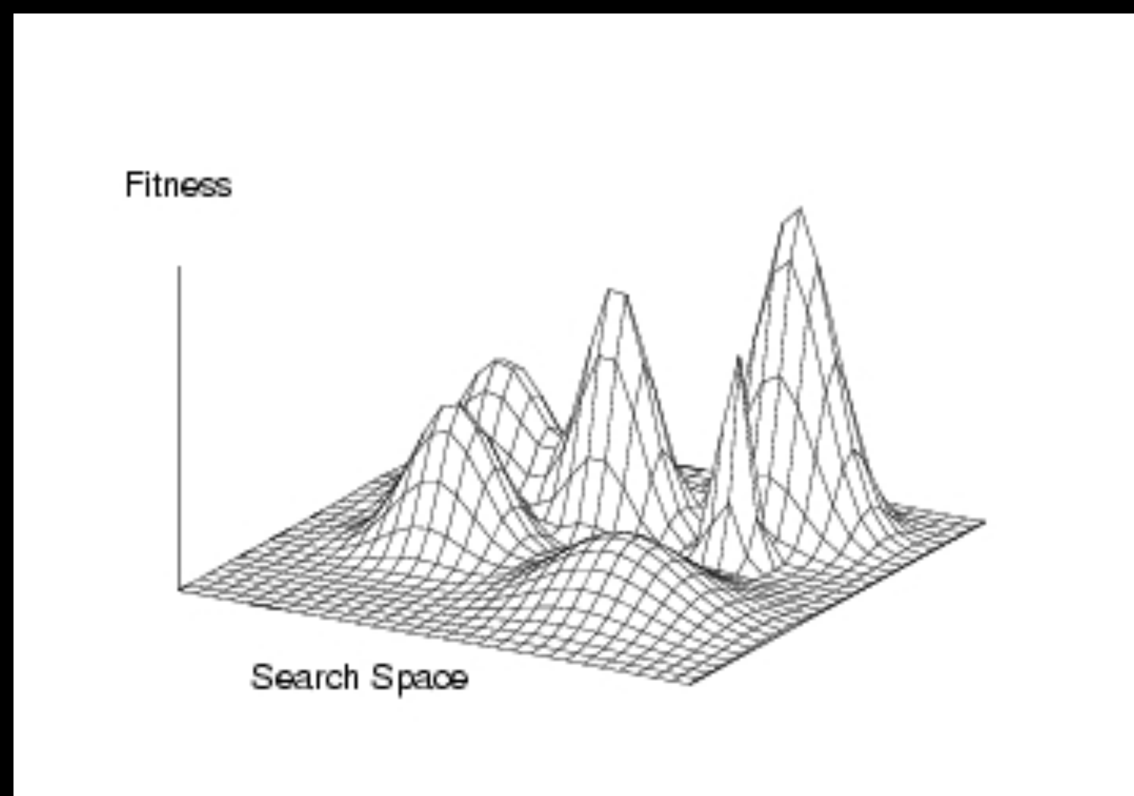
Intelligent biological systems start with a set of genetic ***biological priors*** on their basic units and their overall structure.

Evolution has found these priors to be helpful for survival. Too vital to be learned each time.

Biological intelligence is a result of evolution (nature) and individual learning (nurture).

Evolutionary Learning

Evolution is really an optimization (i.e. learning) algorithm.
Evolution is the ‘outer loop’ for learning in intelligent biological systems.



Evolutionary intelligence is not directly learnable by individuals:
You cannot learn to have a hippocampus in the same way that you cannot learn to have wings.

Evolutionary Learning

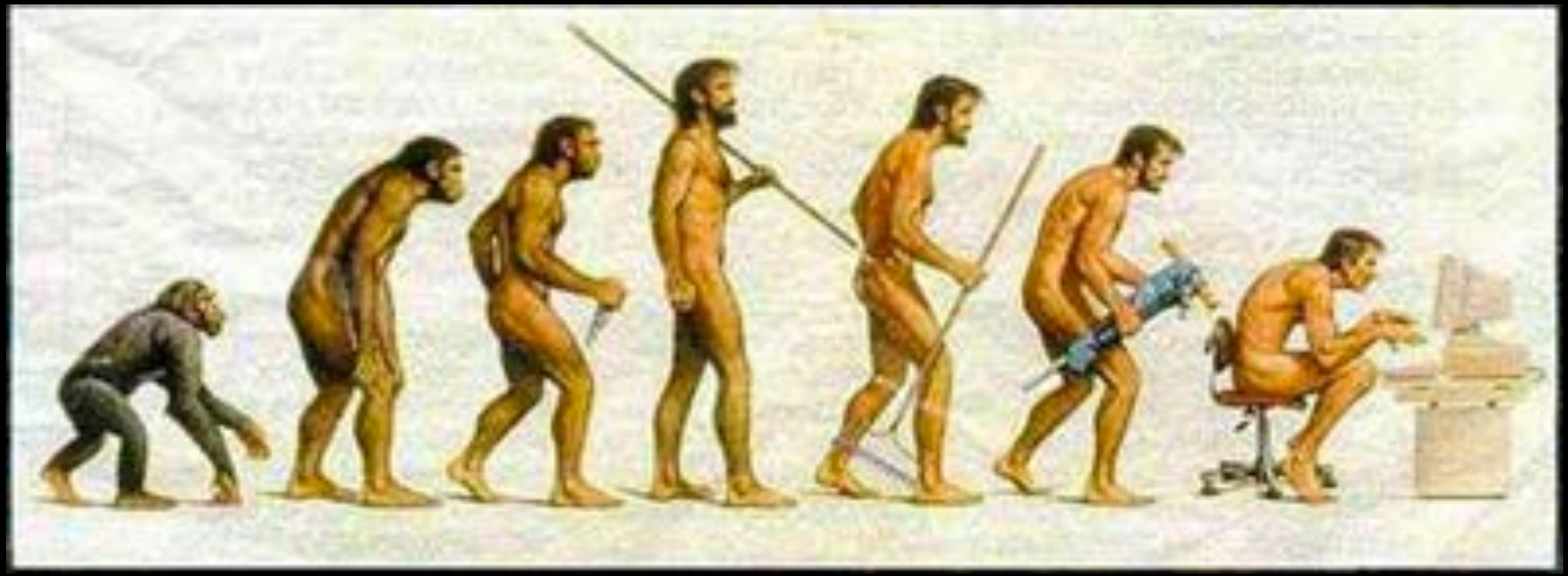
evolution

individual

Biological Intelligence

Evolutionary priors, which allow for pattern recognition, memory, planning, reasoning, communication, etc. are useful for survival.

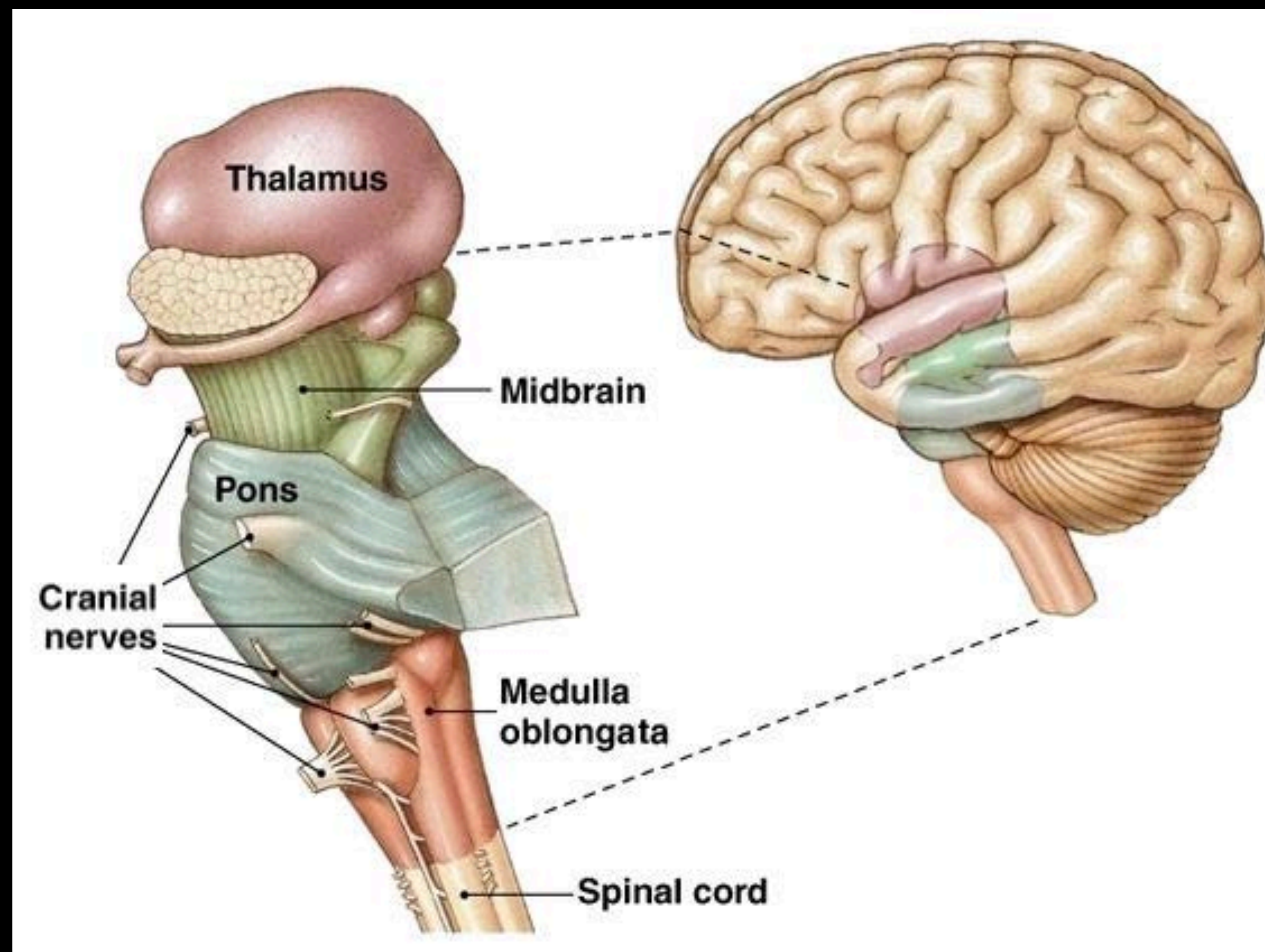
Apparently also useful for discovering the laws of nature.



Evolution has found a computing architecture that generalizes beyond tasks directly relevant for survival.

Overview of Biological Intelligence

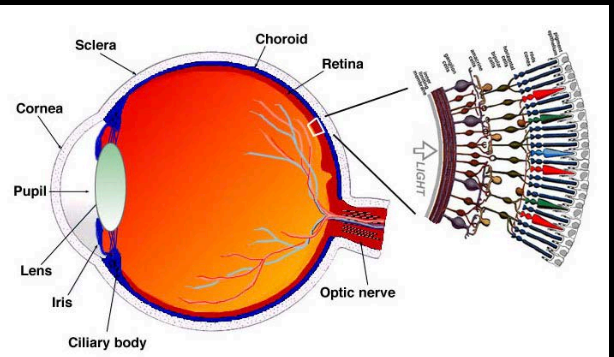
Brain Stem



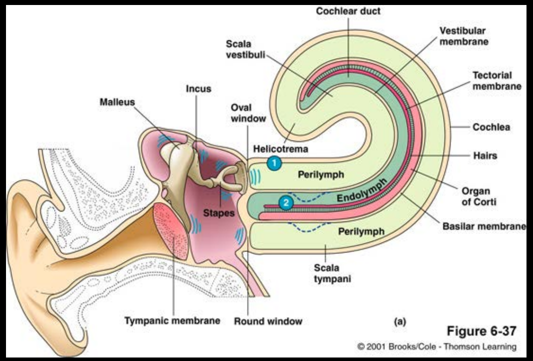
Medulla, Pons, Midbrain

Regulates basic body functions: heart rate, breathing, eating, sleeping, etc.

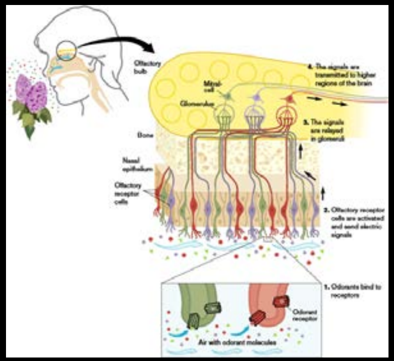
Sensory Inputs



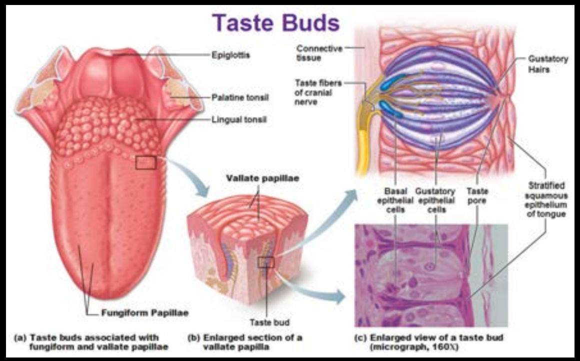
Vision



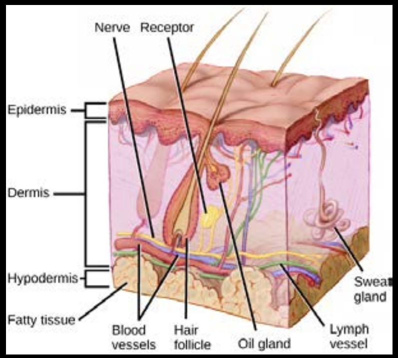
Audition



Olfaction



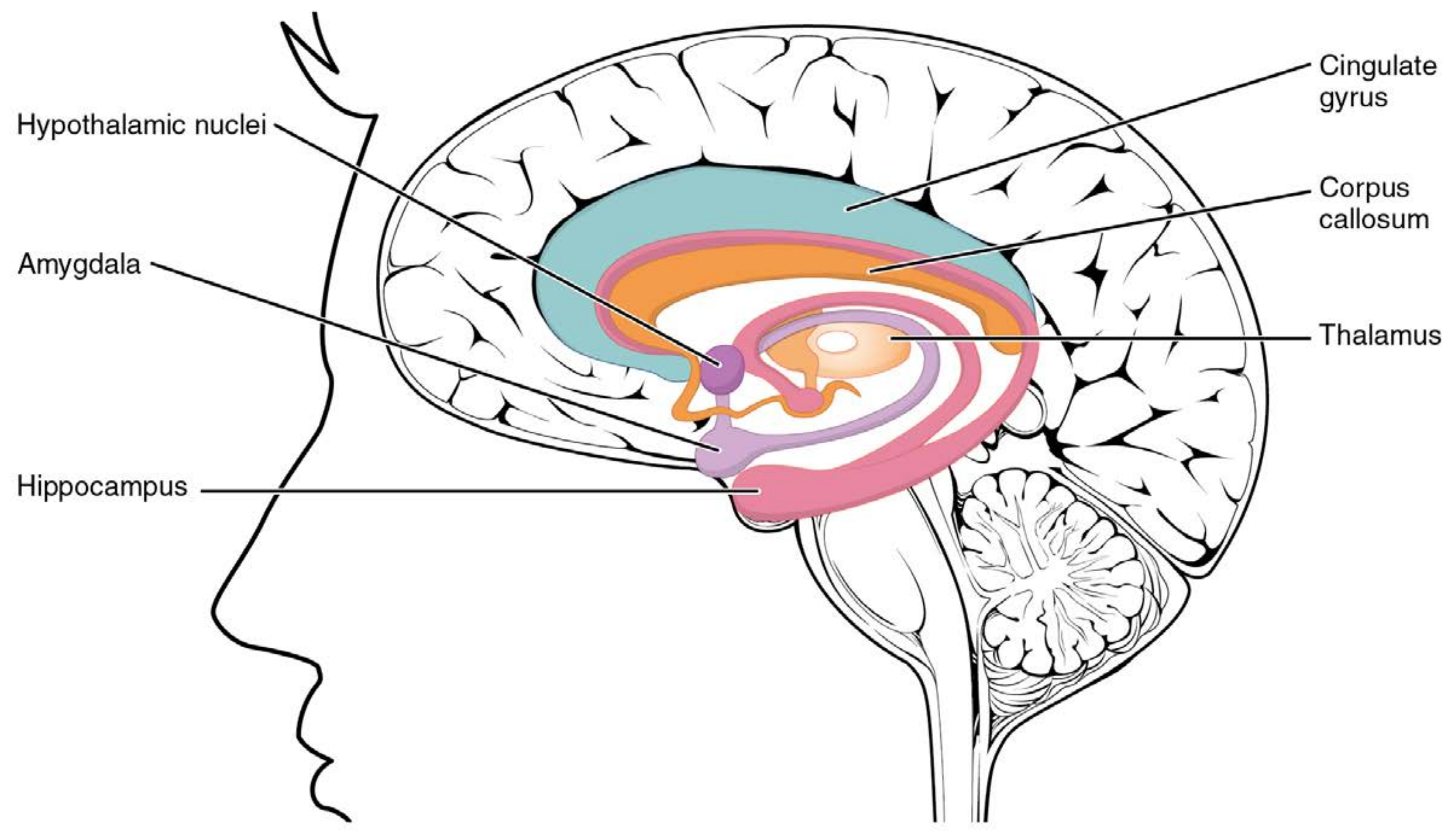
Gustation



Somatosensation

+ Proprioception, Thermoception, Nociception, Mechanoreception, Equilibrioception, Chemoreception, …

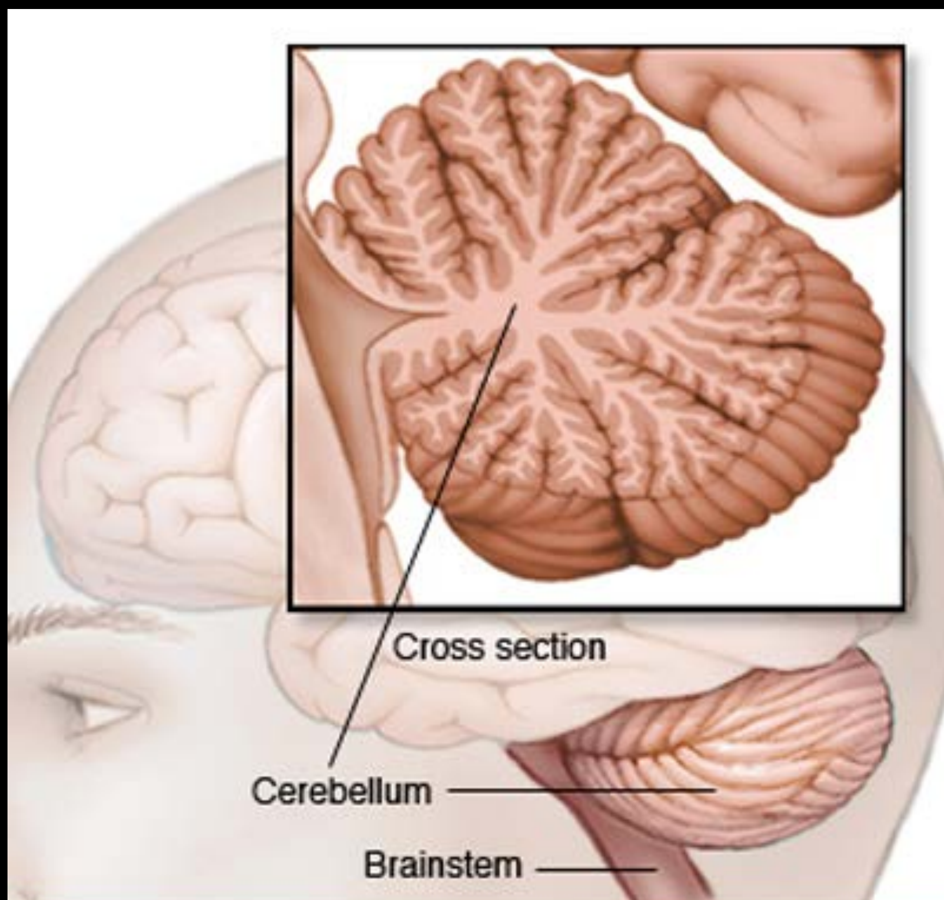
Limbic System

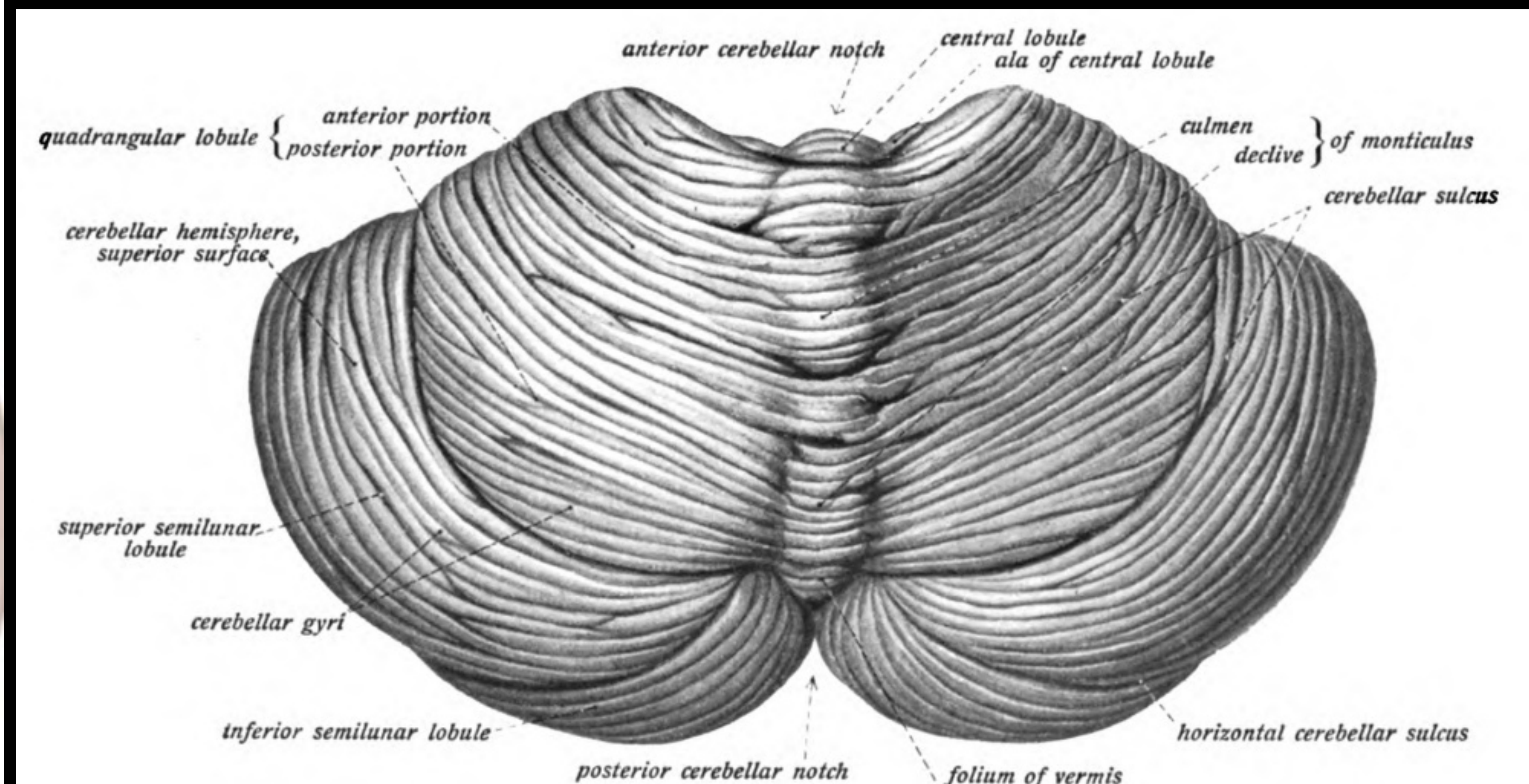


Collection of areas between brain stem and cerebrum: thalamus, hypothalamus, hippocampus, amygdala.

Involved in processing motivation, emotion, learning, senses, memory.

Cerebellum

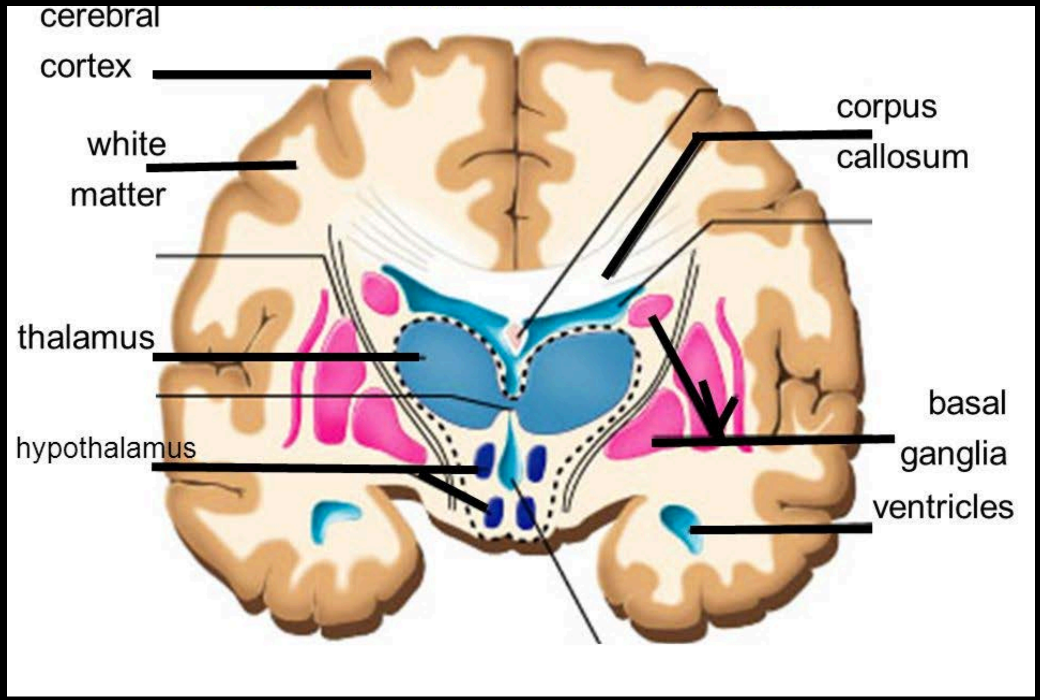




Highly folded cortical sheet, dense with neurons.

Involved in posture and movement.

Cerebrum

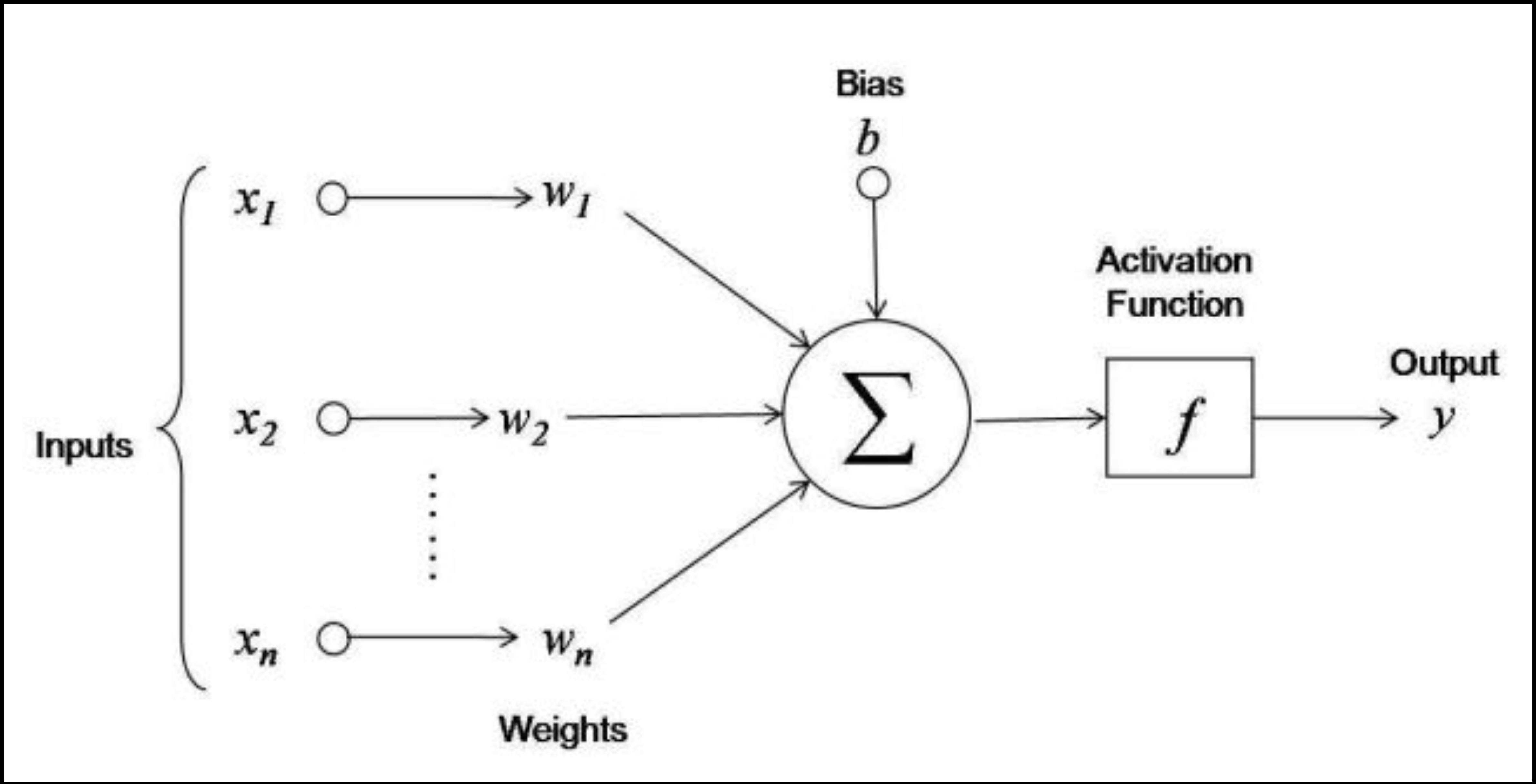


Highly folded cortical sheet. Consists of four lobes in each hemisphere.

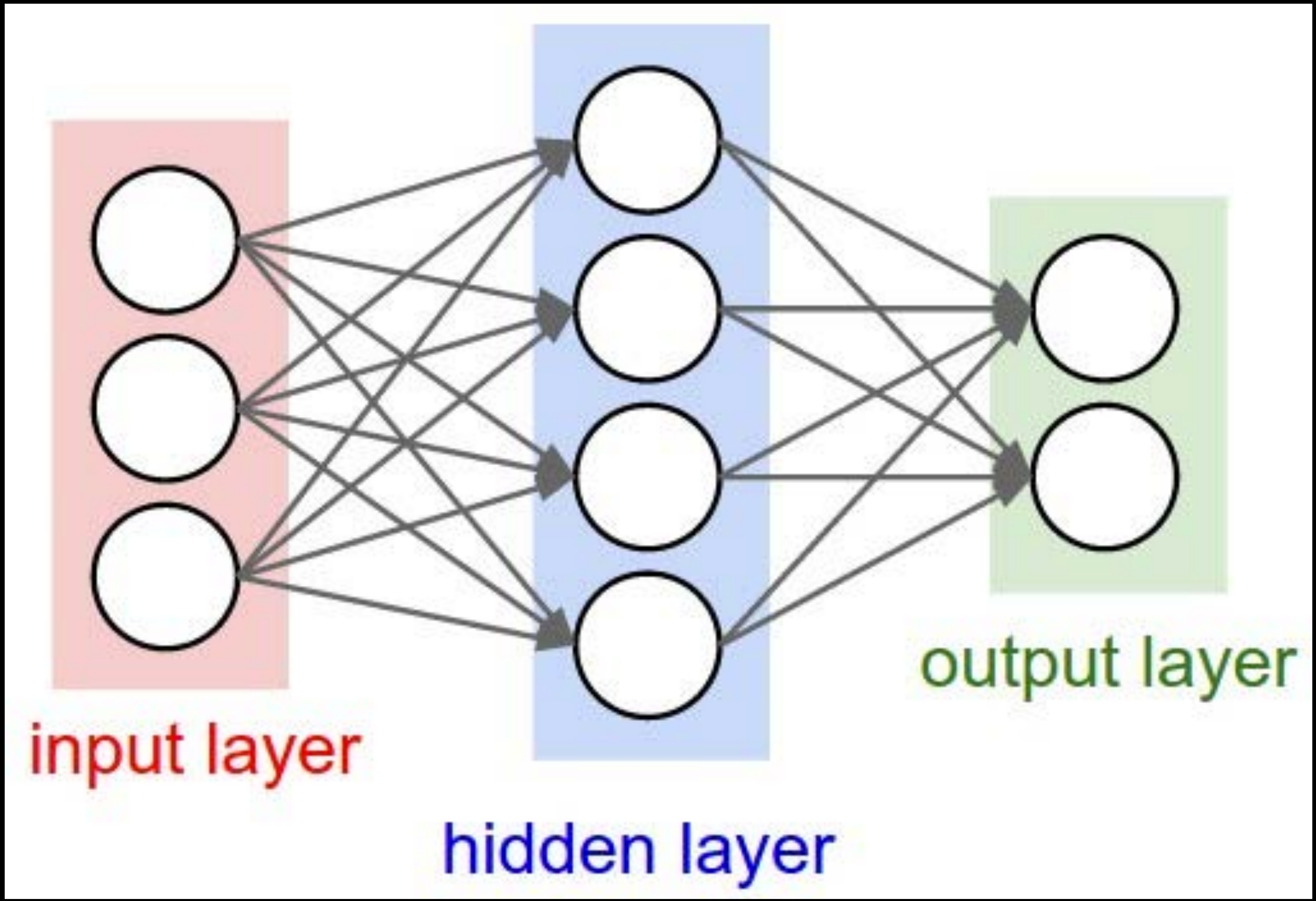
Involved in sensory processing, memory, language, movement, planning, reasoning.

Overview of
(Biologically Inspired)
Machine Intelligence

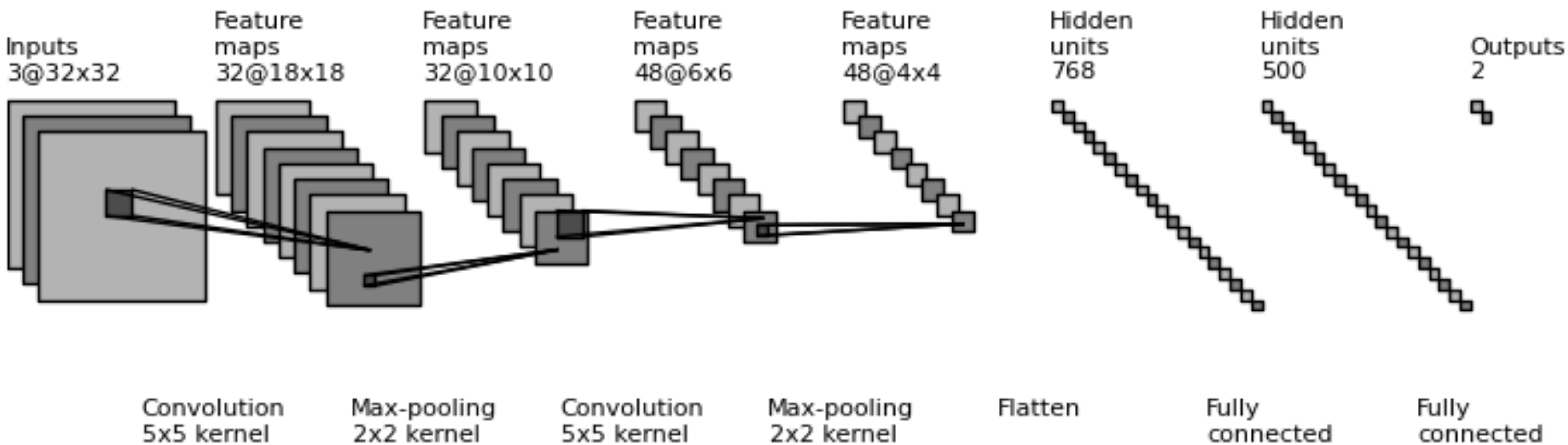
Artificial Neurons



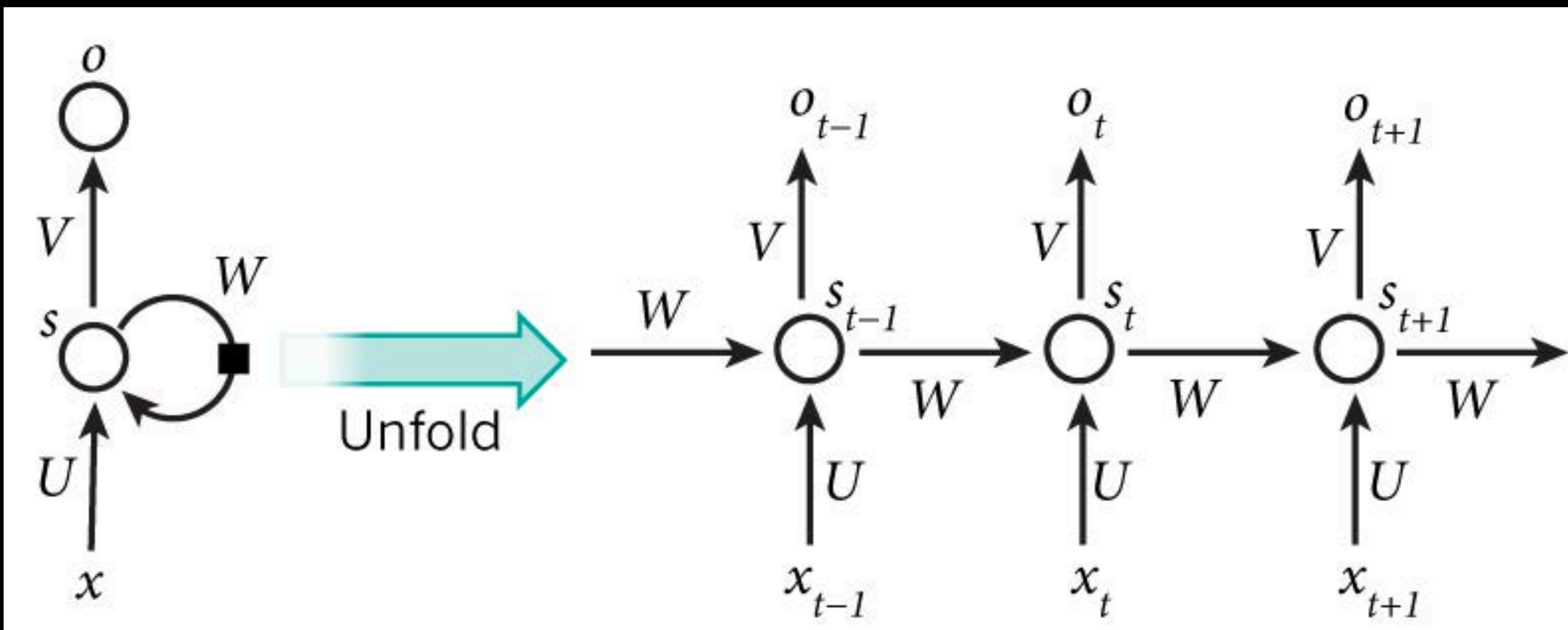
Feedforward Networks



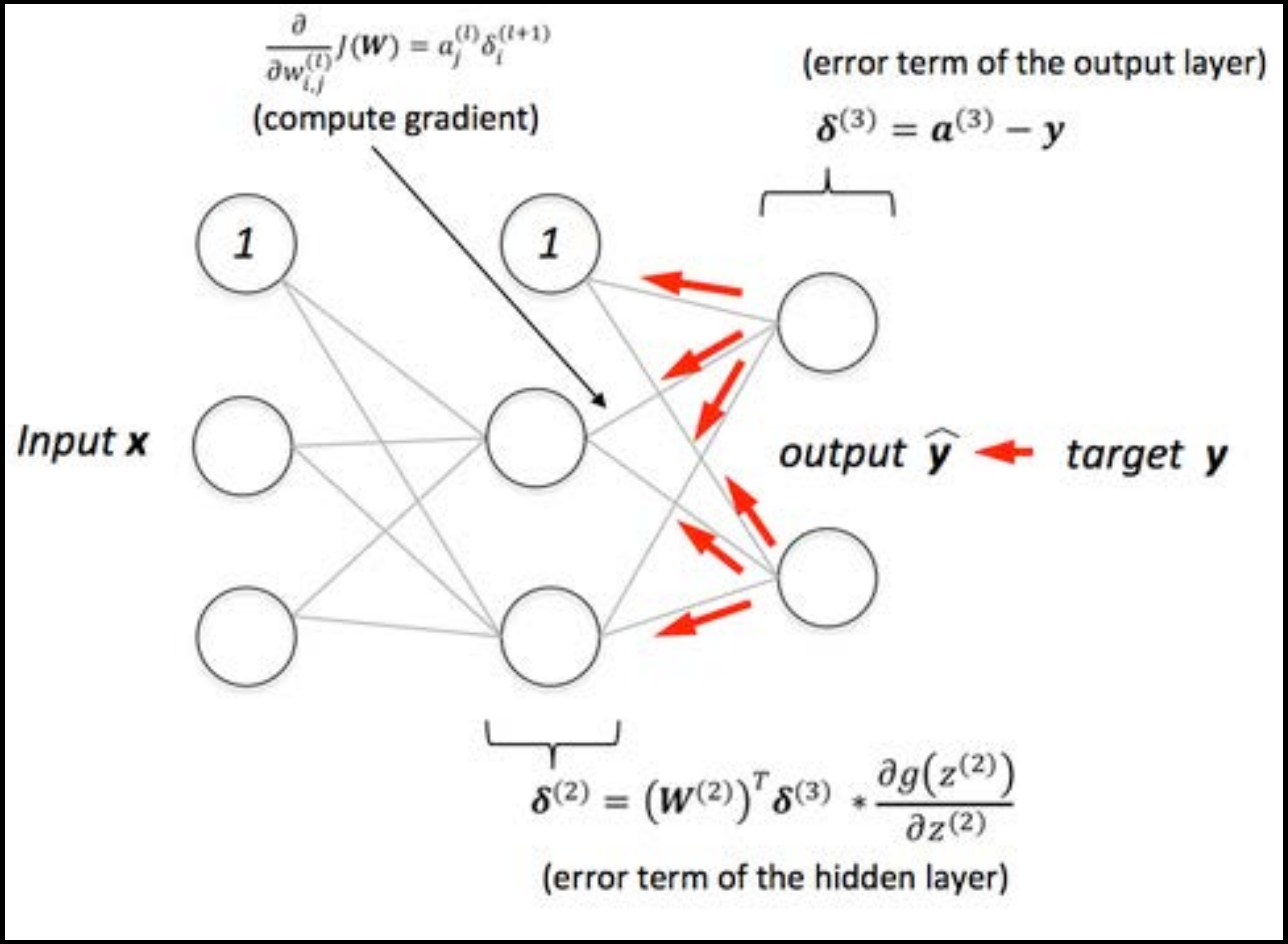
Convolutional Networks



Recurrent Networks

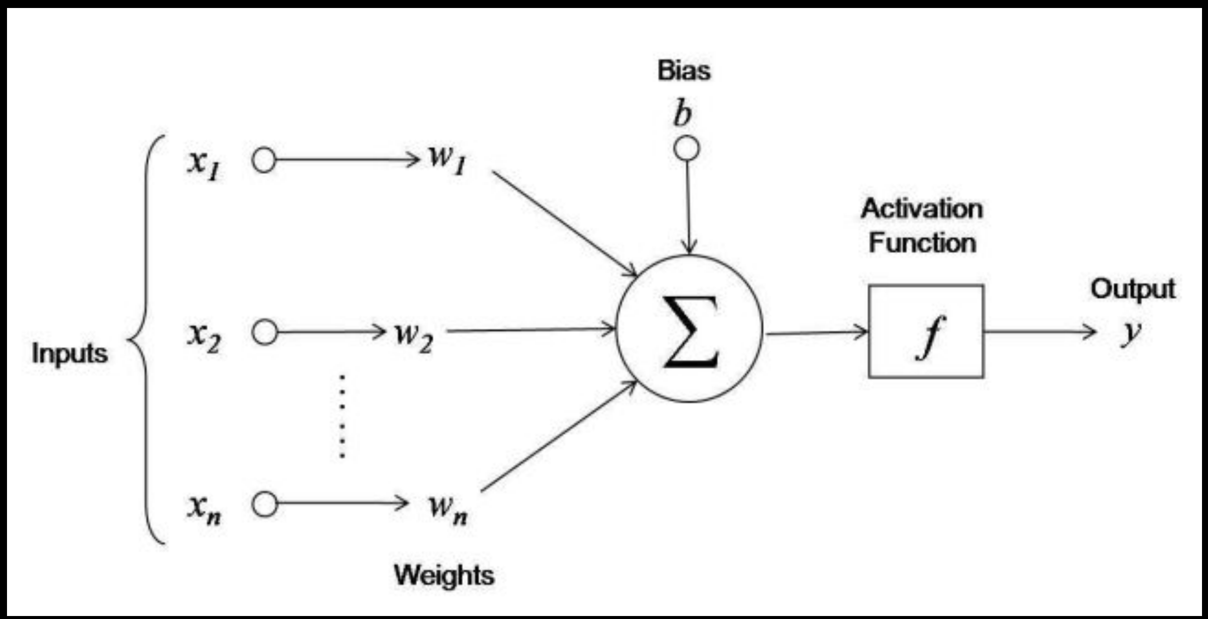


Backpropagation

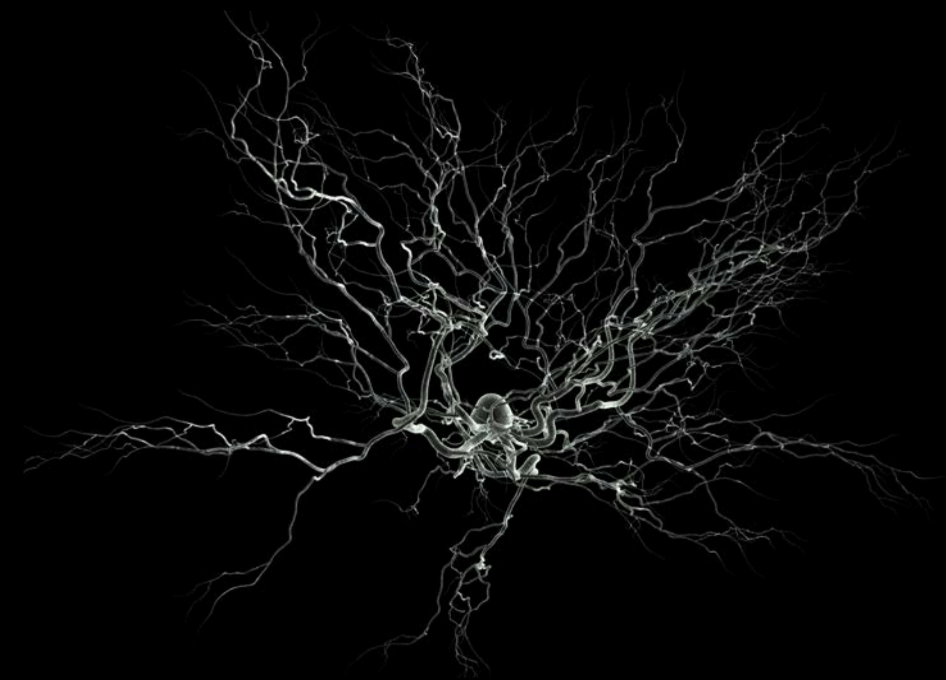


Biological Plausibility
&
Correspondence

Neurons



artificial neuron

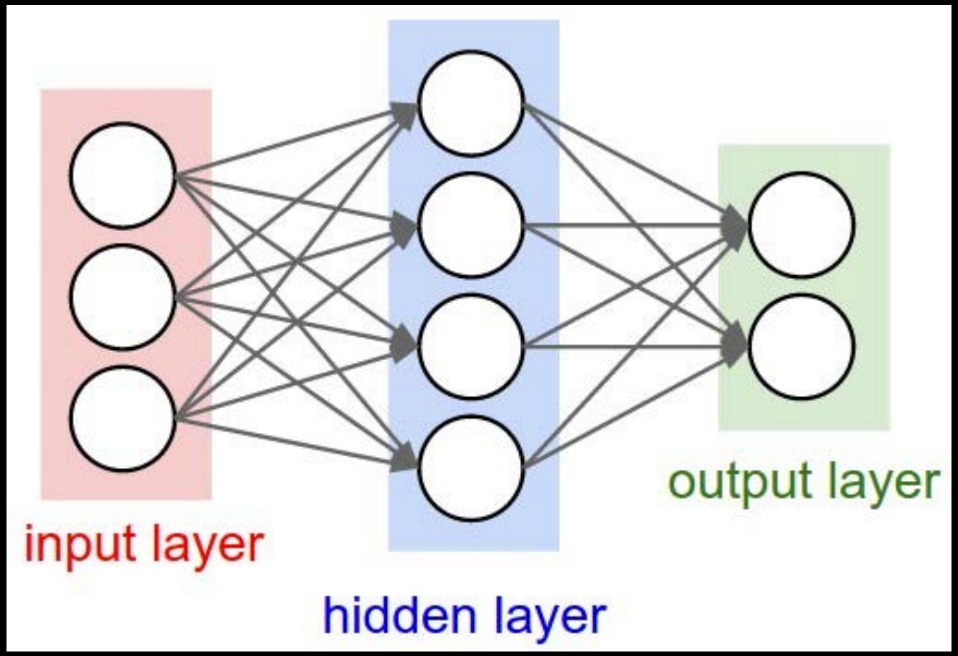


biological neuron

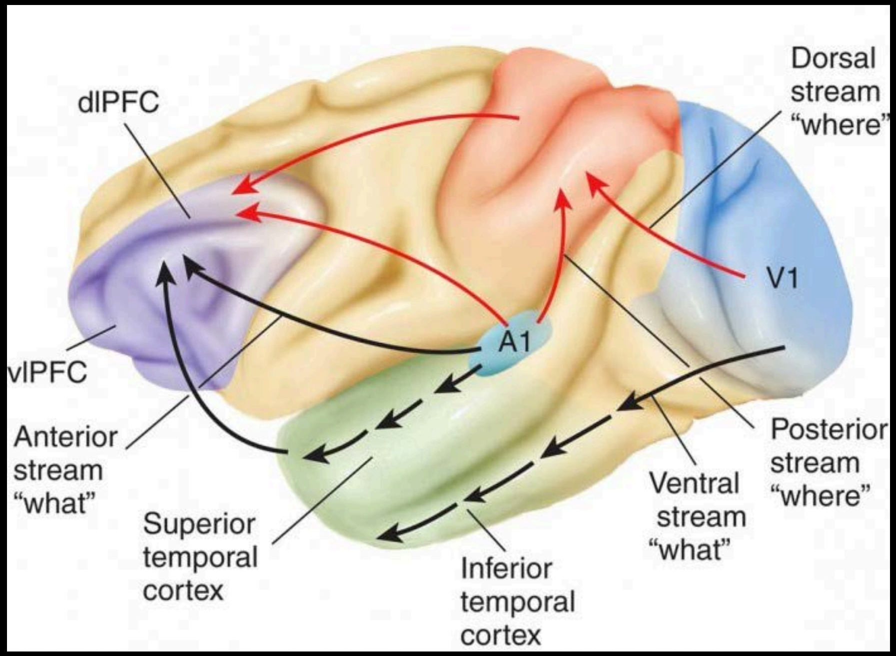
Similarities
inputs/weights (dendrites), bias/non-linearity (threshold voltage), output (axons)

Differences
simplistic dendrites, static (non-temporal), deterministic (?), continuous output, etc.

Sensory Processing



artificial neural network

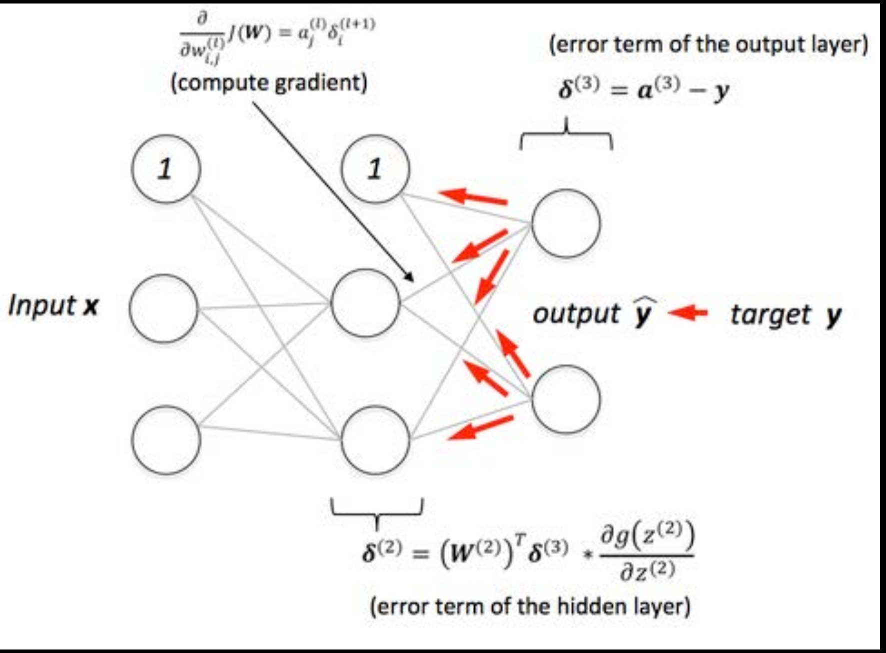


biological neural network

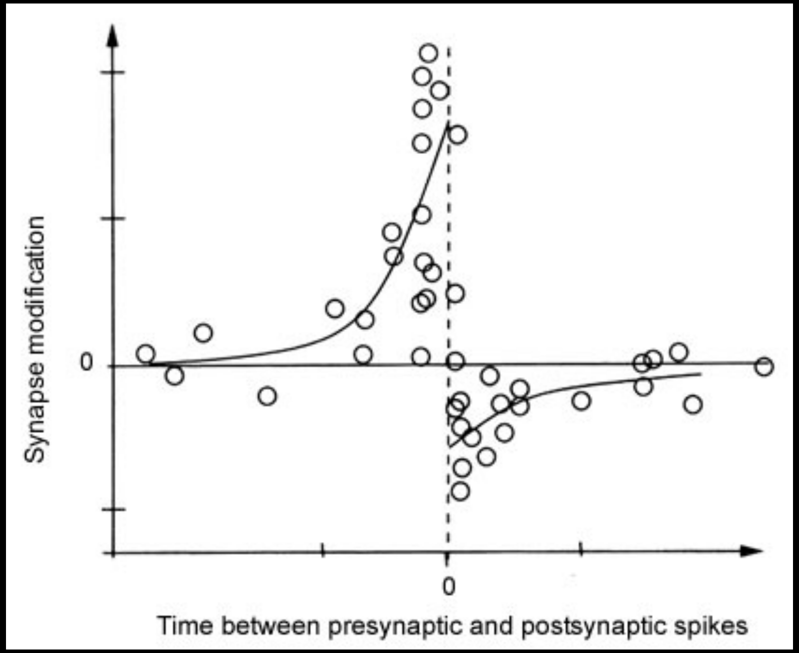
Similarities
repeated basic unit, hierarchical

Differences
typically no feedback, different basic units, different scales, different inputs/outputs, etc.

Learning



backpropagation



spike timing dependent plasticity (STDP)

Similarities

optimization of synaptic strengths

Differences

…

Bi & Poo, 1998

Is Backprop Biologically Plausible?

- backpropagation is purely linear, whereas neurons use linear and non-linear operations

- if feedback paths are used, they would need precise knowledge of the derivatives of the non-linearities at their operating point

- feedback paths would need to be symmetric

- neurons communicate through binary, not continuous, signals

- computation would need to be precisely clocked

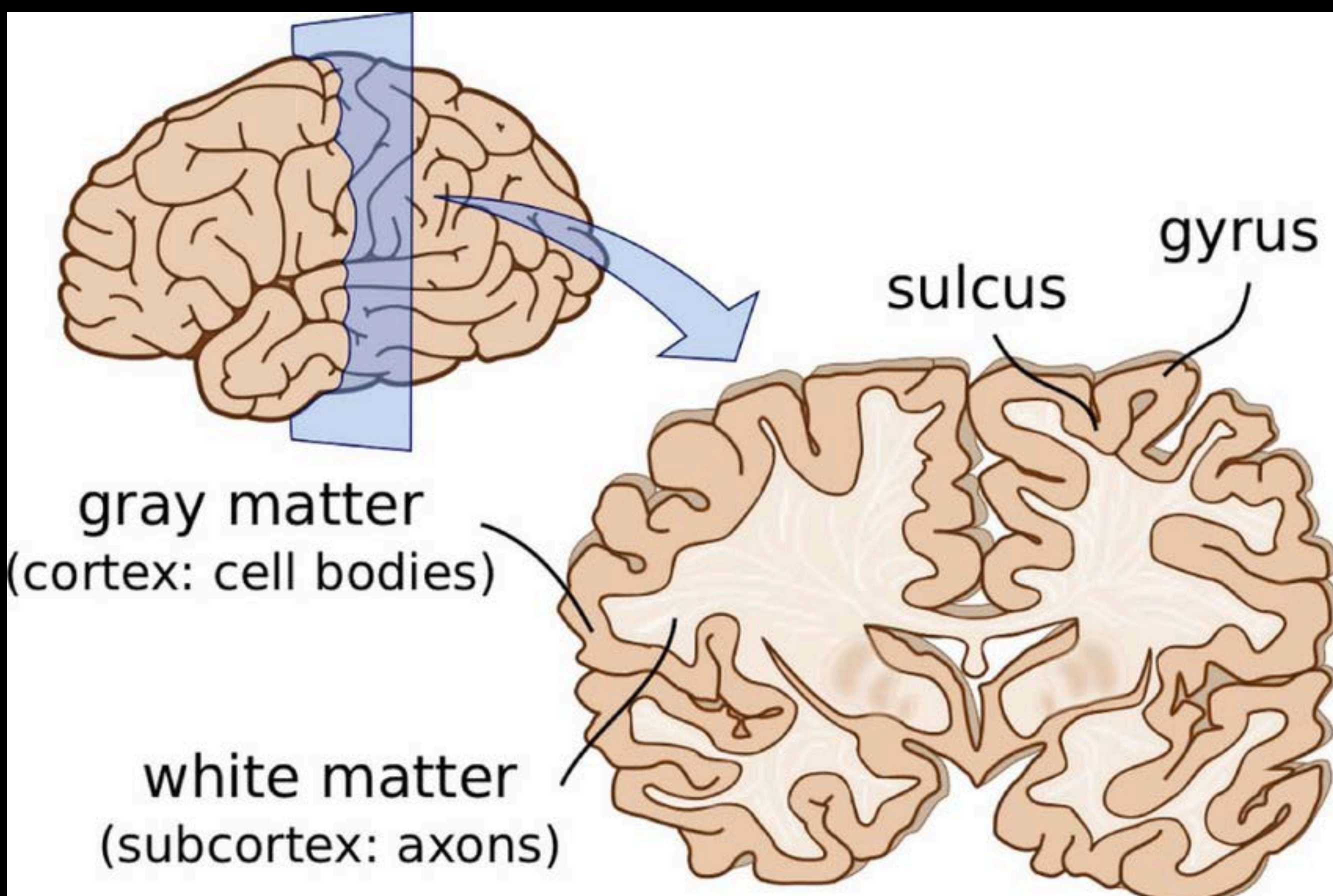
- not clear where outputs come from

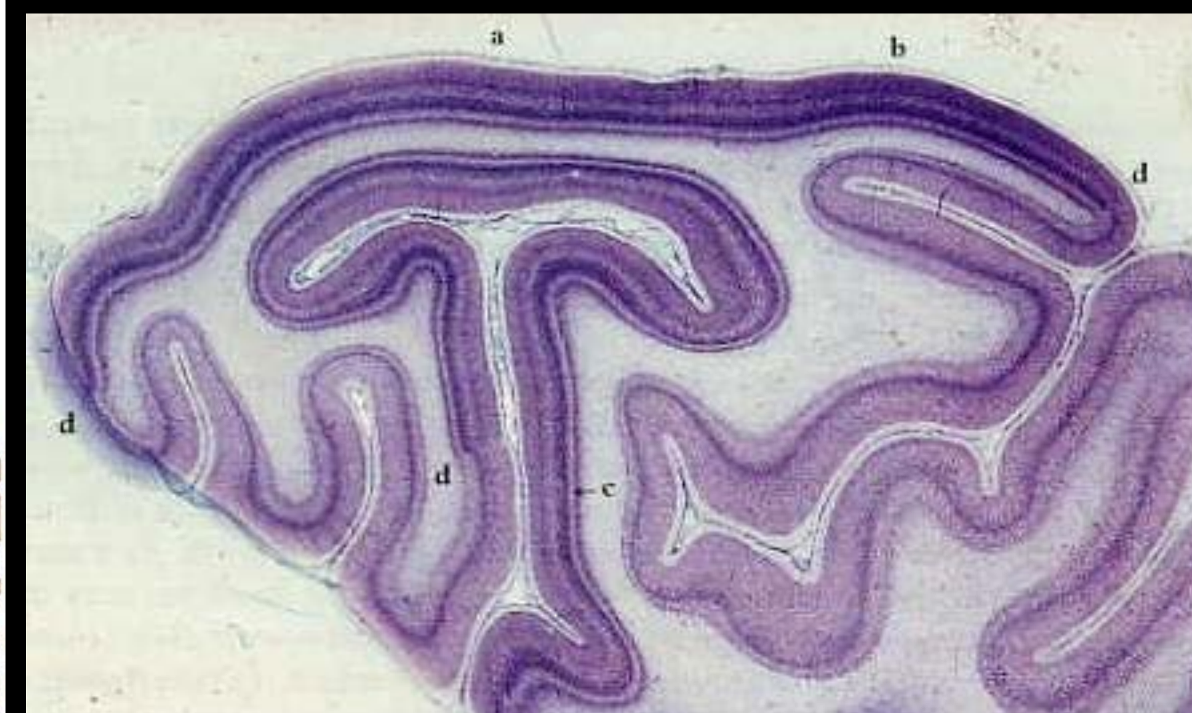
- not clear how to backpropagate through time

Towards Biologically Plausible Deep Learning, Bengio 2015

Sensory Cortex

Cortical Sheet



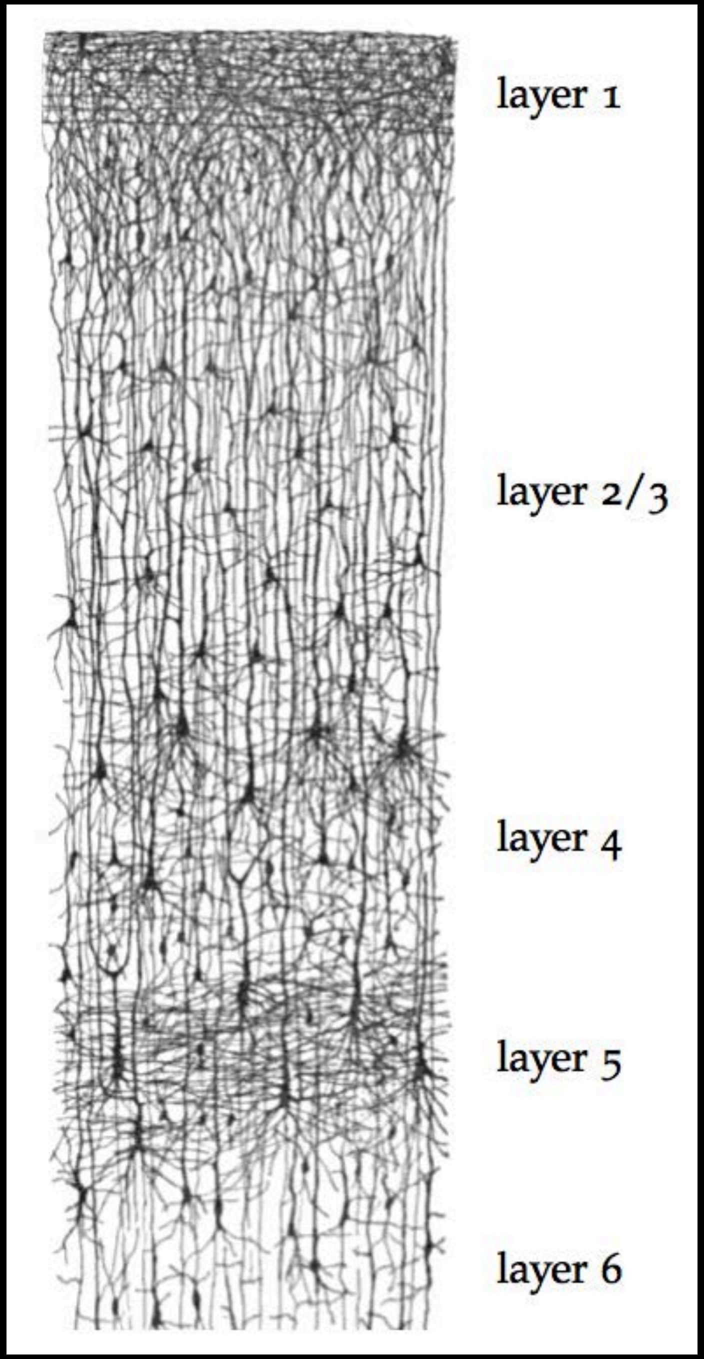


cell bodies form a sheet 2 to 3 mm thick, highly folded

roughly 100,000 neurons per sq. mm

van den Broeke 2016

Cortical Layers



Layer 1: primarily axons and dendrites

Layer 2/3: dense lateral connections in patchy patterns, sparse activations, sends outputs to higher cortical areas

Layer 4: receives input from thalamus or lower cortical area, outputs to layer 3

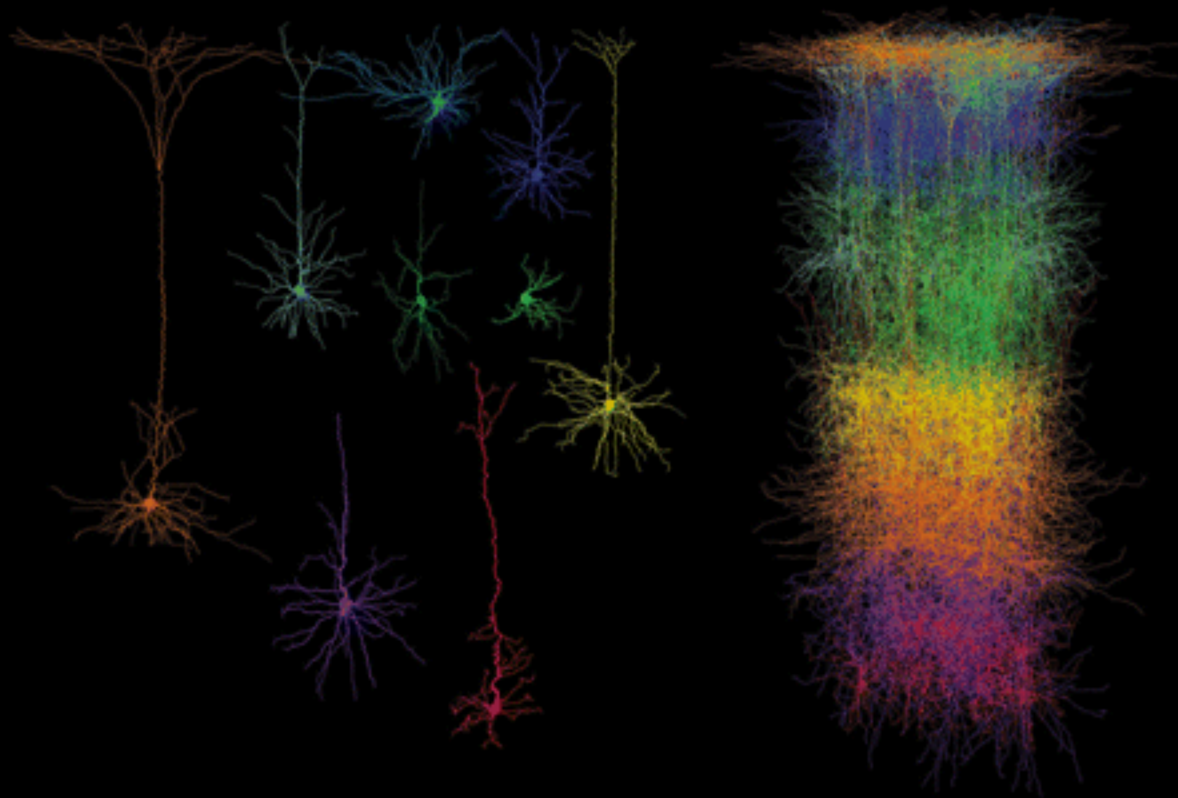
Layer 5: sends outputs to spinal cord and thalamus

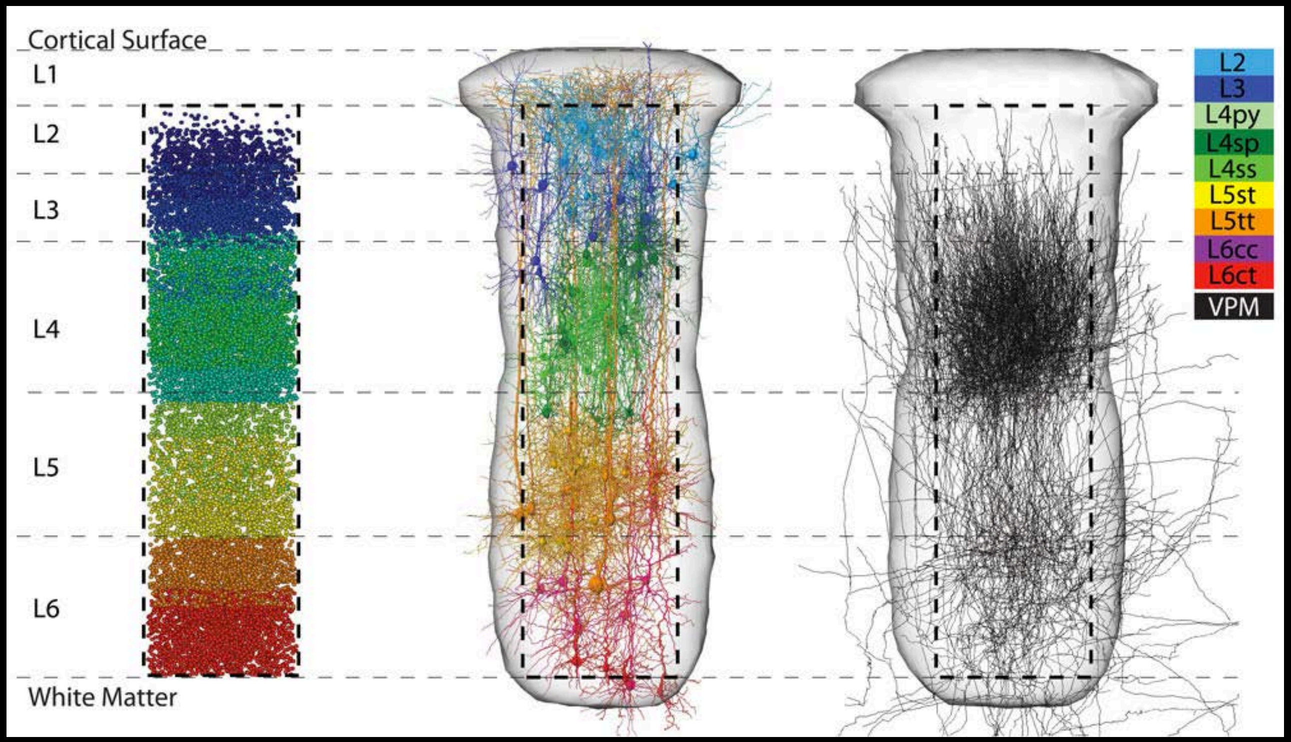
Layer 6: connected to other cortical areas, forms loops with thalamus

van den Broeke 2016

Ramon y Cajal 1899

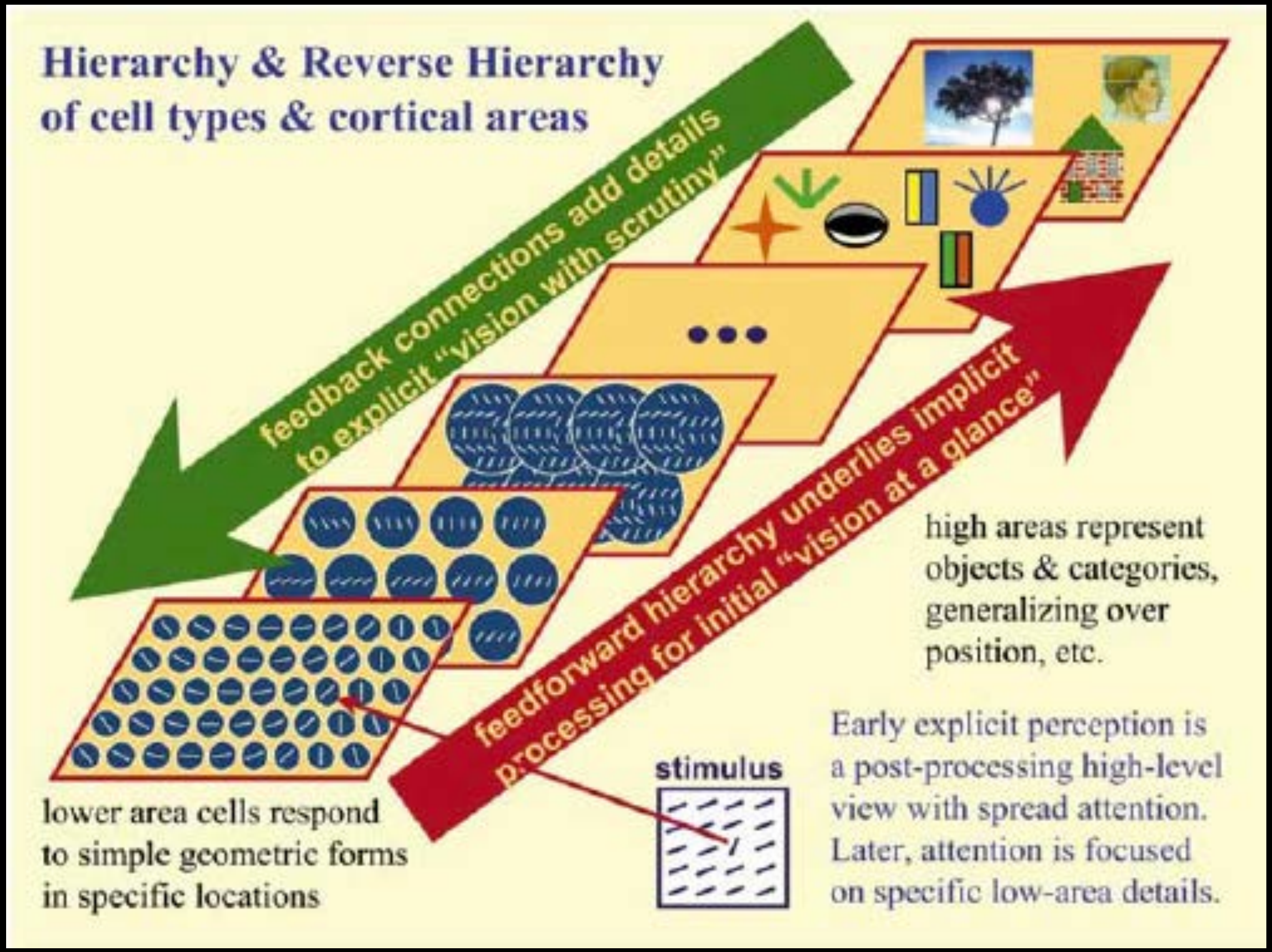
Cortical Columns





neurons within a vertical column have closely related functions, considered to be the basic computational circuit/unit of cortex

Cortical Hierarchy



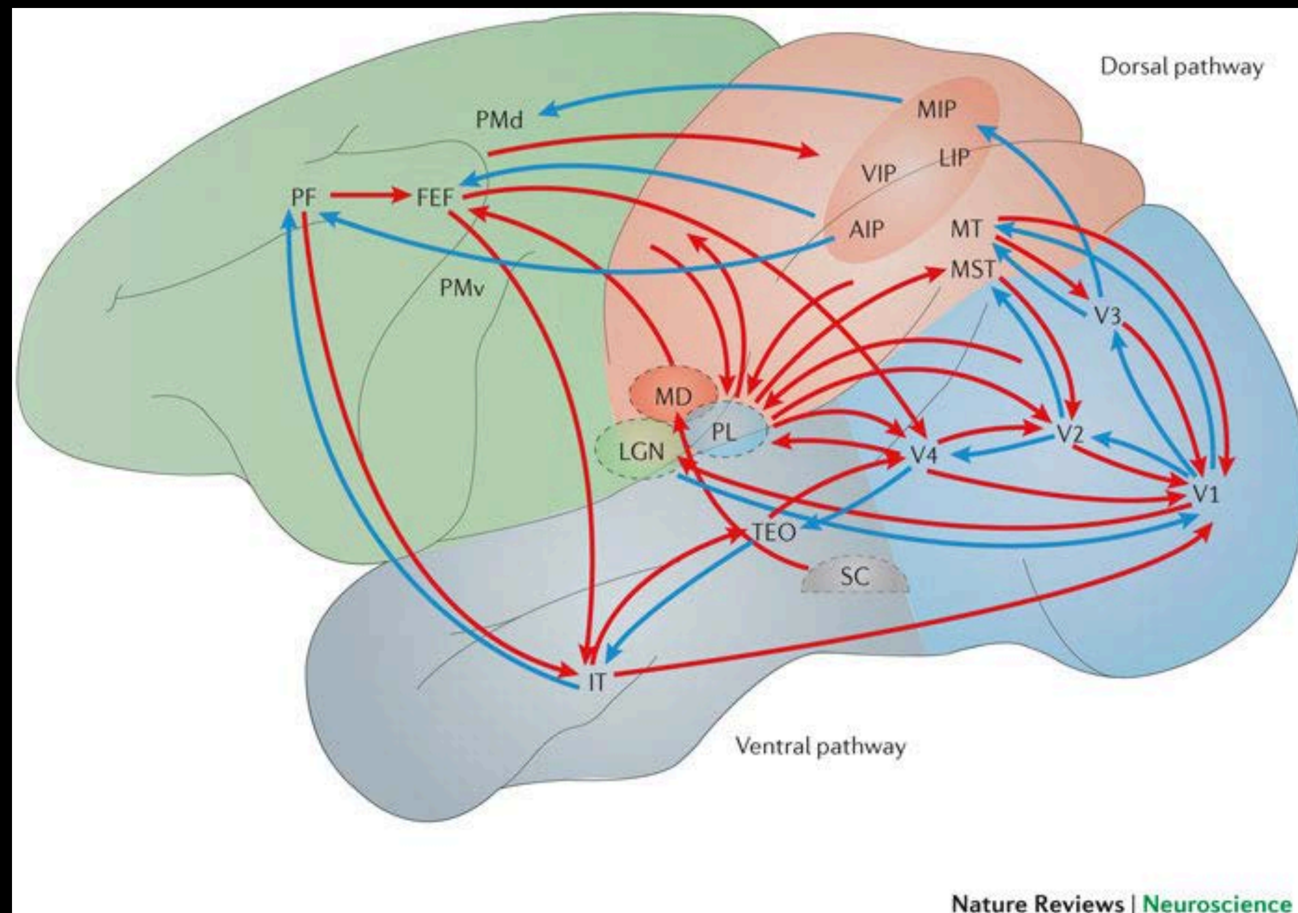
classical view of hierarchical processing in sensory cortex

hierarchical processing allows the compositional structure of natural stimuli to be broken down

lateral inhibition at each processing stage

Hochstein & Ahissar, 2002

Cortical Feedback



feedback connections outnumber feedforward connections

typically thought to send prediction or attention signals

cortex is generative, can imagine low level details from high level cue

Top down influences on visual processing, Gilbert & Li 2013

Latent Variable Models

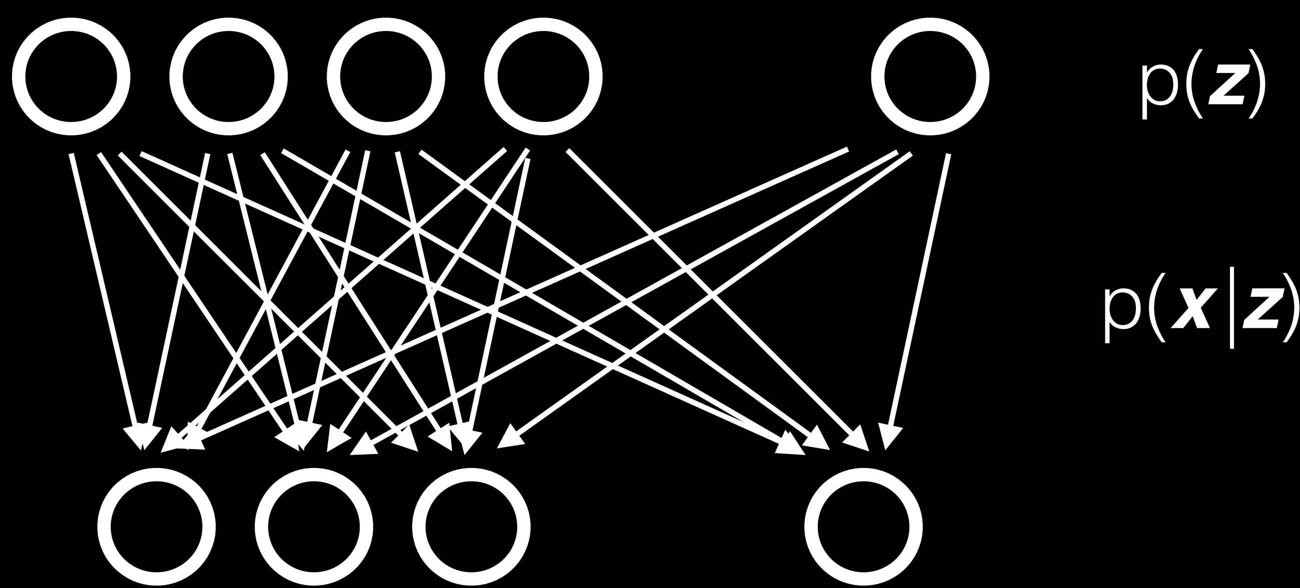
Latent Variable Models

z Latent Variables

hidden structure/causes of the environment

x Observed Variables

environmental stimuli



model the observed data as resulting from a set of latent variables
—> generative model of the data p(***x***|***z***)

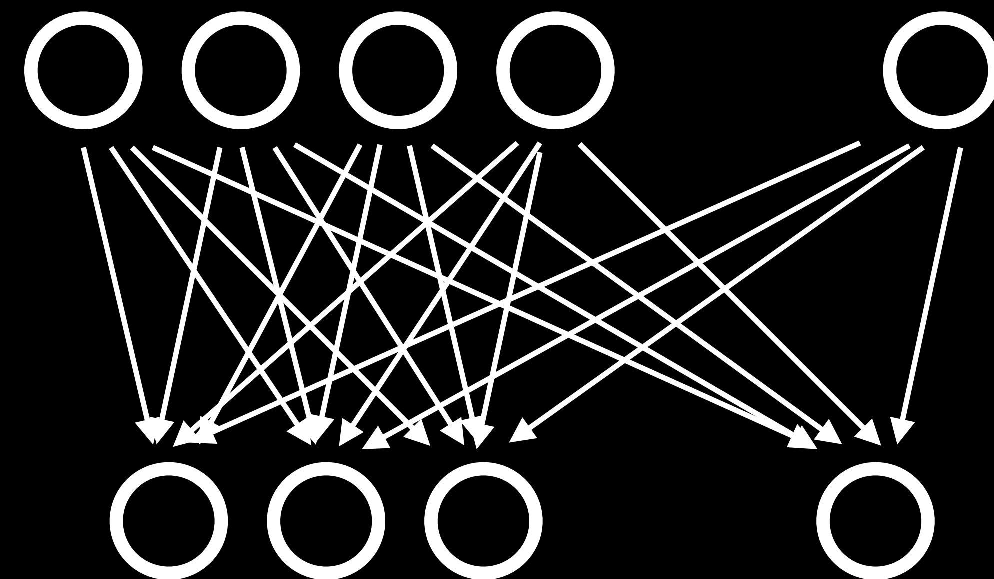
place some prior activation or structural constraint p(***z***) on the latent variables in order to learn some underlying structure in the data

train by maximizing the marginal likelihood of the data p(***x***)

Latent Variable Models

z Latent Variables

hidden structure/causes of the environment



p(***z***)

p(***x***|***z***)

x Observed Variables

environmental stimuli

causal inference: infer latent variables from observed, p(***z***|***x***)

evidential inference: infer observed variables from latent variables, p(***x***|***z***)

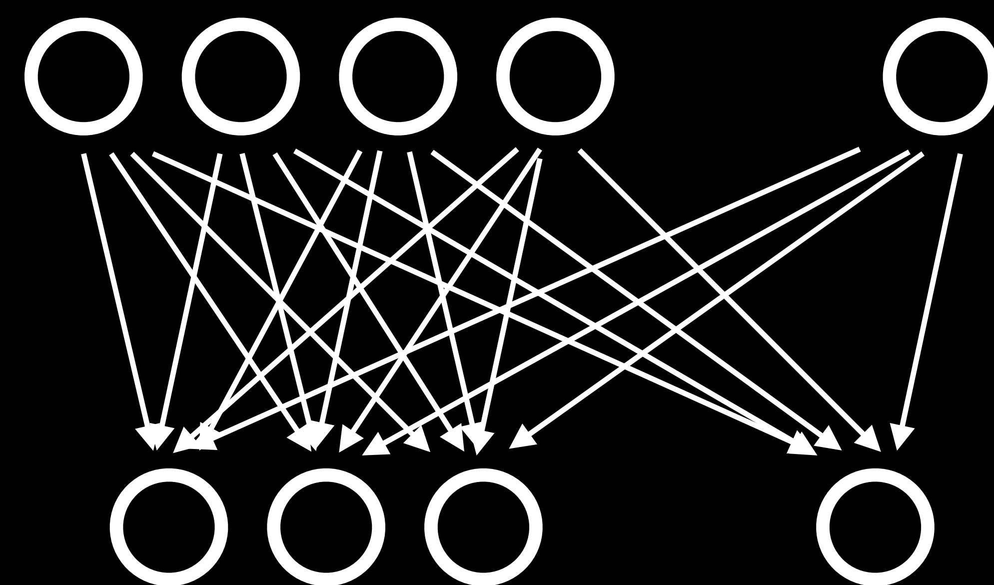
inter-causal inference: infer latent variables from other latent variables, p(***z***|***z***)

Koller & Friedman, 2009
Murphy, 2012

Latent Variable Models

z Latent Variables

hidden structure/causes of the environment



p($\boldsymbol{z}$)

p($\boldsymbol{x}|\boldsymbol{z}$)

x Observed Variables

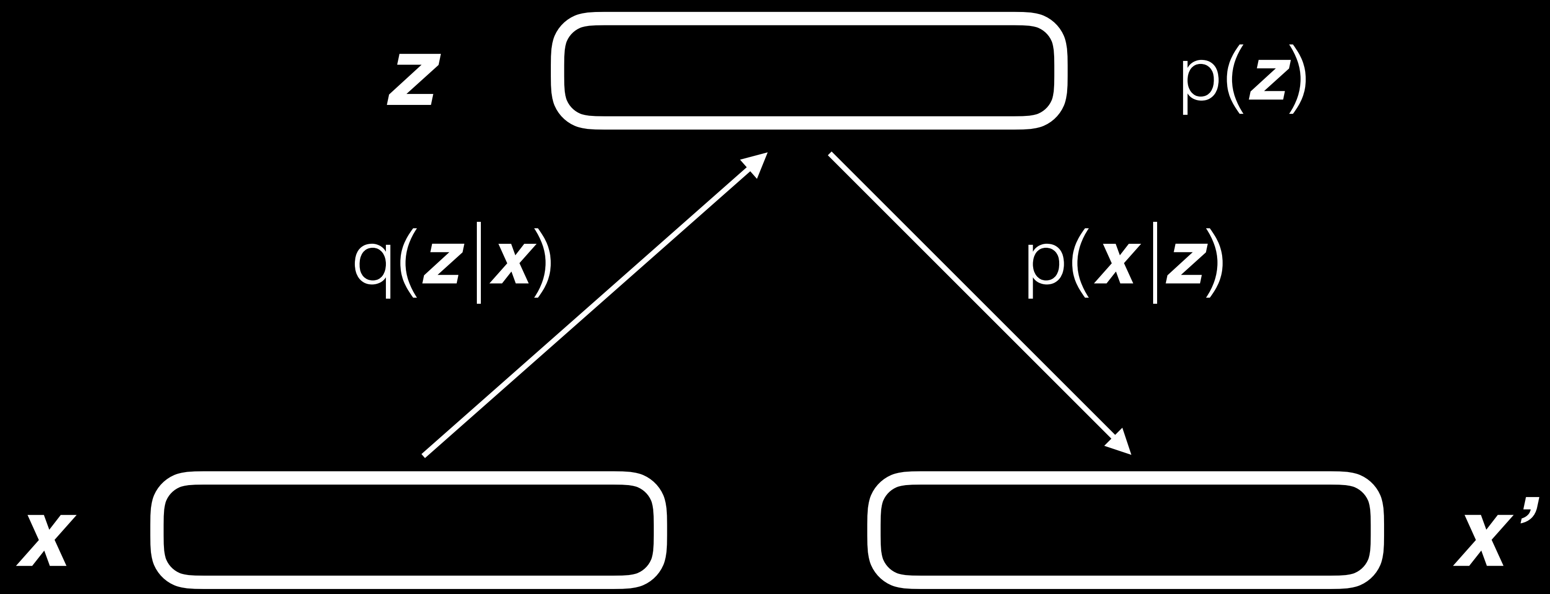
environmental stimuli

inferring the latent variables p($\boldsymbol{z}|\boldsymbol{x}$) directly is often **intractable** in practice since it involves marginalizing over all possible latent states

we need to resort to **approximate inference**:

- sampling-based methods
- variational methods

Variational Inference



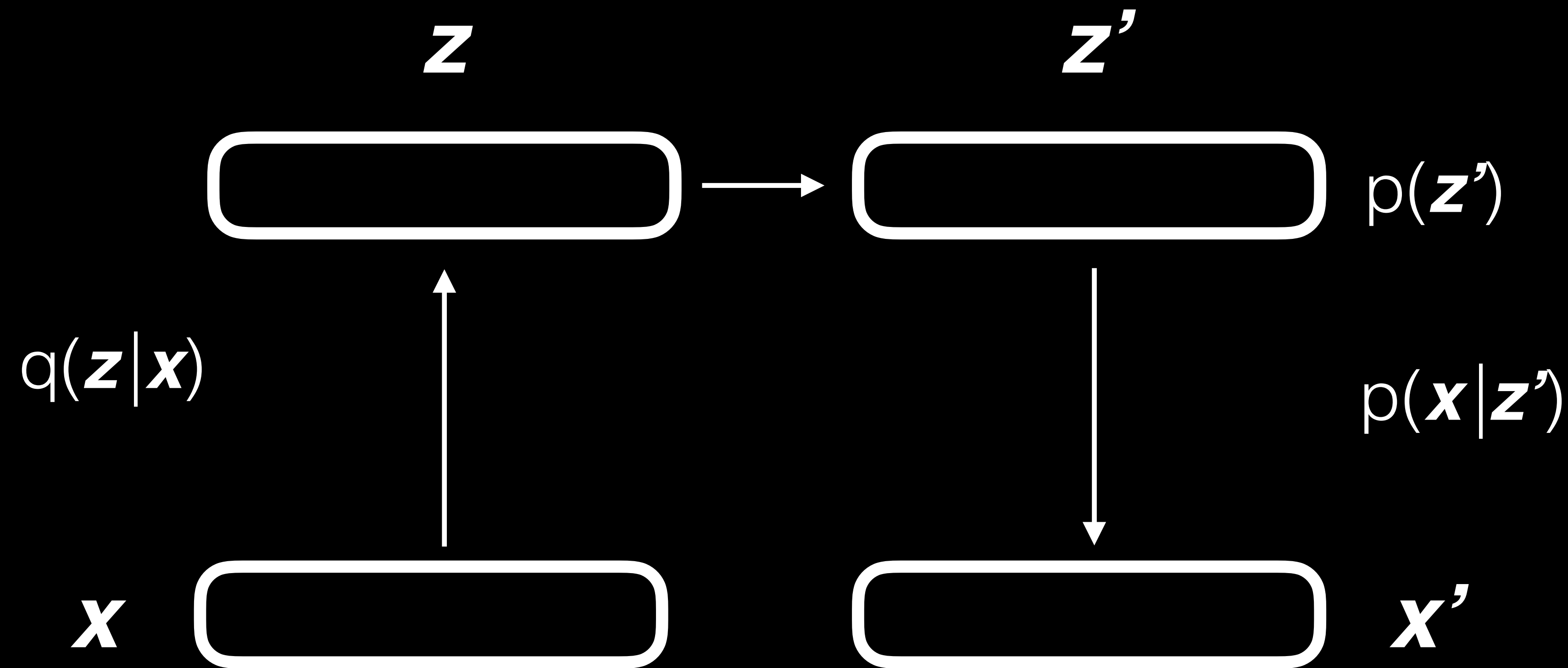
variational methods learn a separate model $q(\boldsymbol{z}|\boldsymbol{x})$ that approximates $p(\boldsymbol{z}|\boldsymbol{x})$

amortized inference: share parameters in approximate inference model across all data points

VAE: use neural networks, learn $q(\boldsymbol{z}|\boldsymbol{x})$ and $p(\boldsymbol{x}|\boldsymbol{z})$ jointly by maximizing lower bound on marginal likelihood, also called *variational free energy*

Kingma & Welling, 2013
Rezende et al., 2014

Normalizing Flows



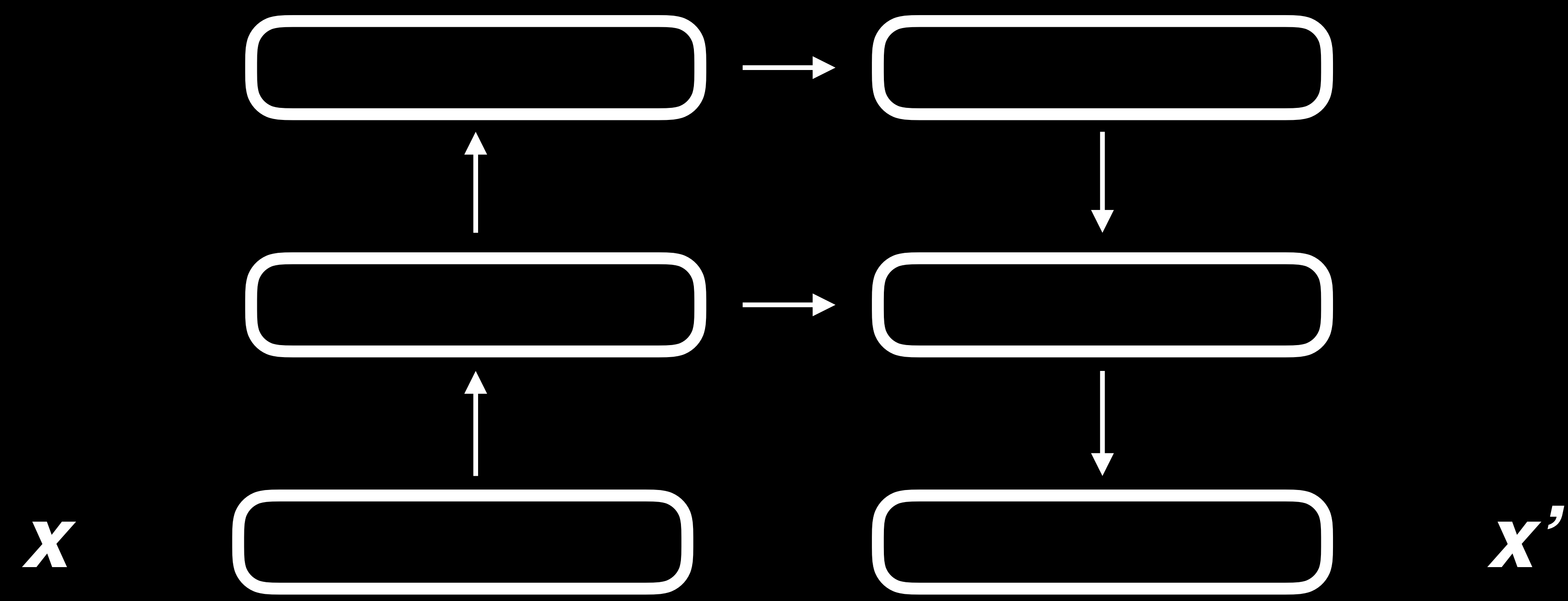
use a set of invertible transformations to sharpen the variational approximation to the posterior distribution

often, this normalizes or *whitens* the latent variables

z' is a whitened version of ***z***

Rezende & Mohamed 2015
Kingma et al. 2016

Hierarchical Latent Variable Models



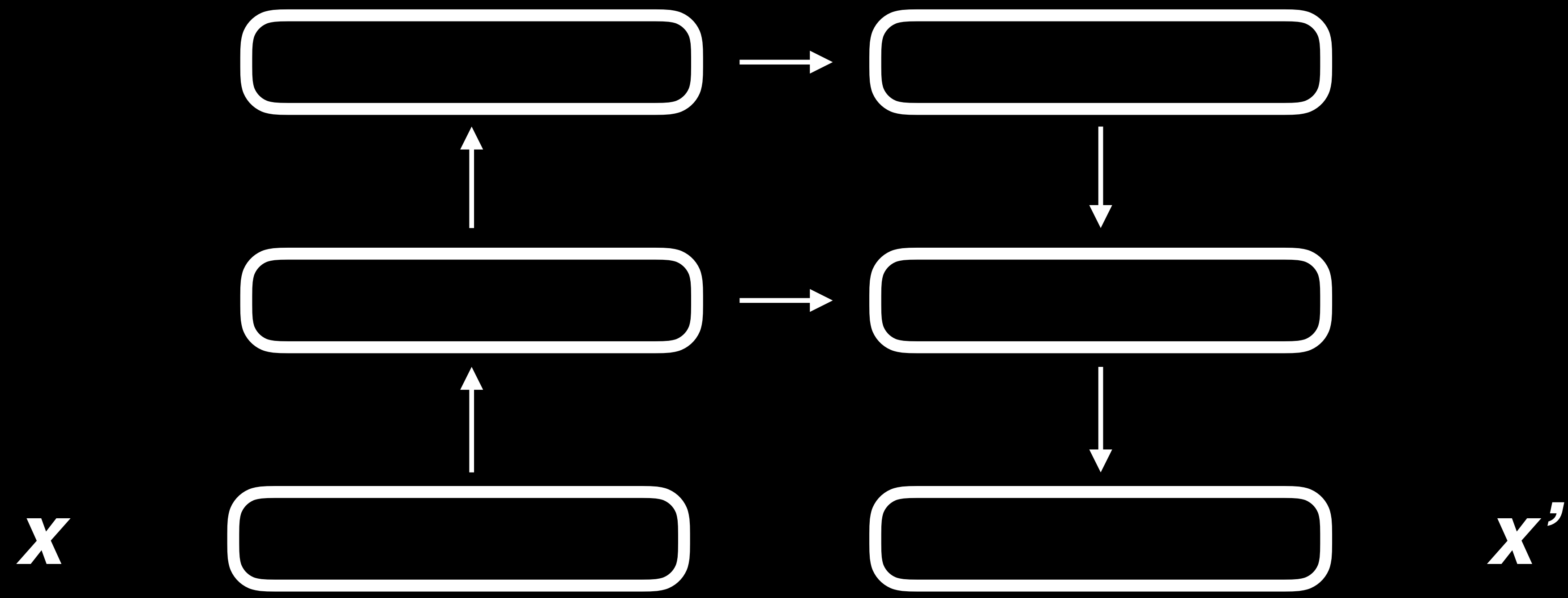
factorize ***z*** over multiple levels, learn multiple levels of hidden structure

reconstruct observed variables as well as lower latent variables

TargetProp: use local learning rules at each level of latent variables

Salimans 2016
Towards Biological Plausible Deep Learning, Bengio et al. 2015

Predictive Coding



send predictions down, send errors up

try to minimize errors/surprise: mismatch between prediction and observation

Rao & Ballard 1999
Friston 2007

Connections to Sensory Cortex

Hierarchical architecture

Probabilistic, Stochastic (?)

Bottom up and top down information

Local learning rules

Normalization at each processing step

Unsupervised/semi-supervised

Can enforce sparsity constraint, use convolutional connectivity

Different Interpretation of Sensory Cortex

Sensory cortex is not an input-output mapping

Not simply extracting patterns

Making predictions about the underlying causes of the sensory inputs

Check against input

Learn from the input rather than the output

Summary, Closing Points, Future Directions

Intelligent Systems

to construct an intelligent system, we need

- **priors** on the system's parameters
- **data** to learn the system's parameters

without appropriate priors, any system would likely be too complicated to learn or would require too much data

Biological Inspiration

- we can try to mimic evolution's priors on computational architecture, units, etc. (Marr's algorithmic level) to develop machine intelligence
- hopefully shorten the evolutionary learning process significantly, fewer experiments
- will likely develop a better understanding of biological intelligence
- need interdisciplinary insights

Biological Inspiration

- much of deep learning has focused on trying to mimic sensory processing in cortex
- need work on
  - motor planning and control (motor cortex, cerebellum)
  - low shot, novelty learning (hippocampus)
  - sensorimotor integration
  - etc.

Human Civilization

- we have developed better communication skills
  - language
  - writing
  - art
  - recordings
  - internet

- as a result, we can capture and transmit knowledge more easily between each other and to the next generation

Human Civilization

- human civilization has become an additional optimization loop:

evolution

    human civilization

        individual

Human Civilization

- we have also extended our abilities through external memory and computational resources
  - written language
  - other media
  - computers
  - internet

- human knowledge is far too vast to fit in any single person's brain

Machine Intelligence

- machine intelligence is starting as specialized applications
  - speech recognition
  - self-driving cars
- as we develop a better understanding, we will build more general machines
  - true personal assistants
  - machine scientists, agents
- general machine intelligence will be able to directly integrate with external computational and memory resources
- will likely eventually be a distributed network of systems, working, communicating, and learning together
  - direct communication, faster copying/transfer of information

References

Marr, David. "Vision: A computational investigation into the human representation and processing of visual information." (1982).

Bengio, Yoshua, et al. "Towards biologically plausible deep learning." arXiv preprint arXiv:1502.04156 (2015).

Bi, Guo-qiang, and Mu-ming Poo. "Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type." The Journal of neuroscience 18.24 (1998): 10464-10472.

van den Broeke, Gerben. "What auto-encoders could learn from brains." (2016).

Marblestone, Adam H., Greg Wayne, and Konrad P. Kording. "Toward an integration of deep learning and neuroscience." Frontiers in Computational Neuroscience 10 (2016).

Hochstein, Shaul, and Merav Ahissar. "View from the top: Hierarchies and reverse hierarchies in the visual system." Neuron 36.5 (2002): 791-804.

Gilbert, Charles D., and Wu Li. "Top-down influences on visual processing." Nature Reviews Neuroscience 14.5 (2013): 350-363.

Murphy, Kevin P. Machine learning: a probabilistic perspective. MIT press, 2012.

Koller, Daphne, and Nir Friedman. Probabilistic graphical models: principles and techniques. MIT press, 2009.

Kingma, Diederik P., and Max Welling. "Auto-encoding variational bayes." arXiv preprint arXiv:1312.6114 (2013).

Rezende, Danilo Jimenez, Shakir Mohamed, and Daan Wierstra. "Stochastic backpropagation and approximate inference in deep generative models." arXiv preprint arXiv:1401.4082 (2014).

Rezende, Danilo Jimenez, and Shakir Mohamed. "Variational inference with normalizing flows." arXiv preprint arXiv:1505.05770 (2015).

Kingma, Diederik P., Tim Salimans, and Max Welling. "Improving variational inference with inverse autoregressive flow." arXiv preprint arXiv:1606.04934 (2016).

Friston, Karl J., and Klaas E. Stephan. "Free-energy and the brain." Synthese 159.3 (2007): 417-458.

Rao, Rajesh PN, and Dana H. Ballard. "Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects." Nature neuroscience 2.1 (1999): 79-87.