Where Neuroscience Meets AI

And what it means for the future

Jane Wang, Adam Marblestone, Kevin Miller
December 7

Bios

Jane Wang is a research scientist at DeepMind on the neuroscience team, working on meta-reinforcement learning and neuroscience-inspired artificial agents. Her background is in physics, complex systems, and computational and cognitive neuroscience.

Kevin Miller is a research scientist on the Neuroscience Team at DeepMind and a postdoc at University College London. He is currently working on understanding structured reinforcement learning in mice and machines.

Adam Marblestone is a Schmidt Futures innovation fellow, was previously a research scientist at DeepMind, and earlier did a PhD in BioPhysics and worked at a brain computer interface company.

Intended audience

- Those with some background in ML
- Those with interest in how insights from neuroscience can apply to AI research
- *No background in neuroscience required*

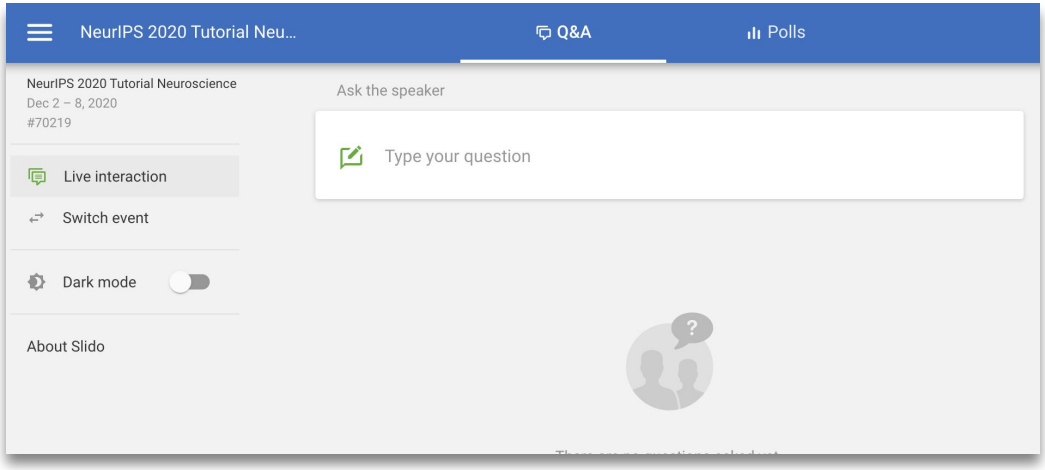
Schedule

1. Introduction / background (15 min)
2. Cognitive neuroscience (30 min)
    a. Q/A (10 min)
3. Learning circuits and mechanistic neuroscience (30 min)
    a. Q/A (10 min)
4. Recent advancements at the intersection (25 min)
5. General discussion (30 min), sli.do questions

Please submit questions at sli.do here

Be sure to address your question to one of the speakers if it's regarding a specific section of the tutorial

https://app.sli.do/event/92gy6nuo



Introduction

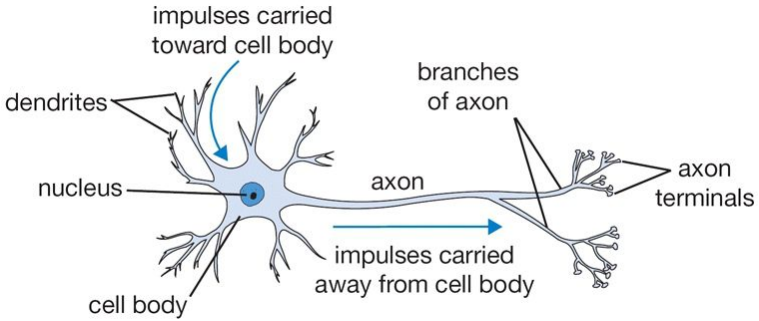
Jane Wang

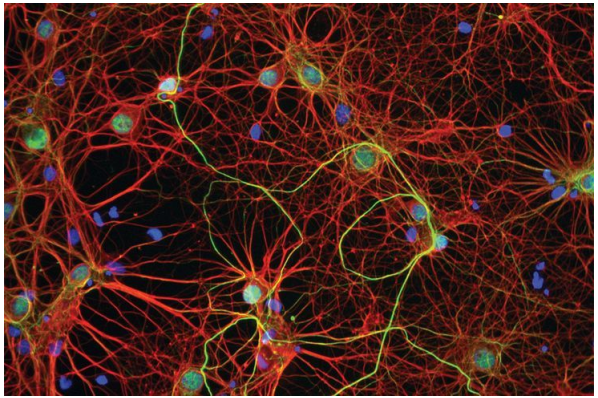
Why does neuroscience matter for AI?

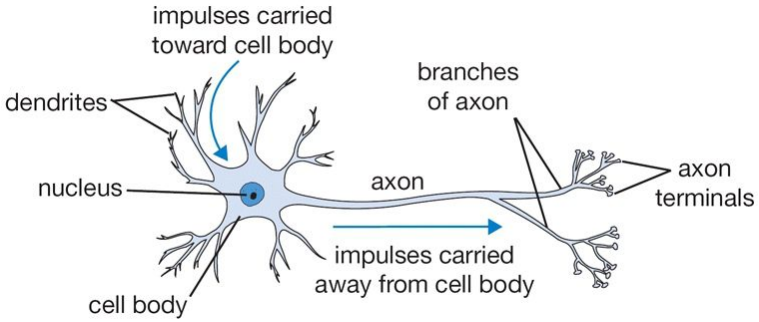
"The brain's the only known example of truly general intelligence."

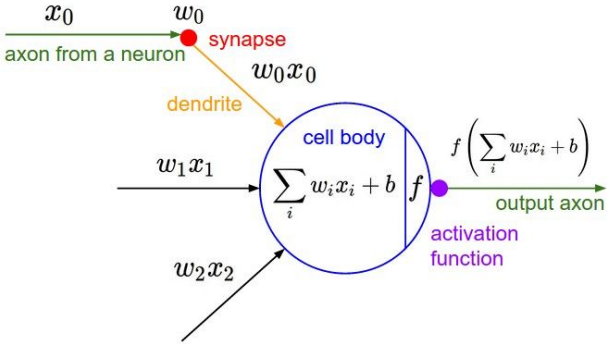
"Neuroscience discoveries can inspire new architectures, models, or cost functions."

"Humans have the amazing ability to ____, and we'd like our agents to have that ability as well."

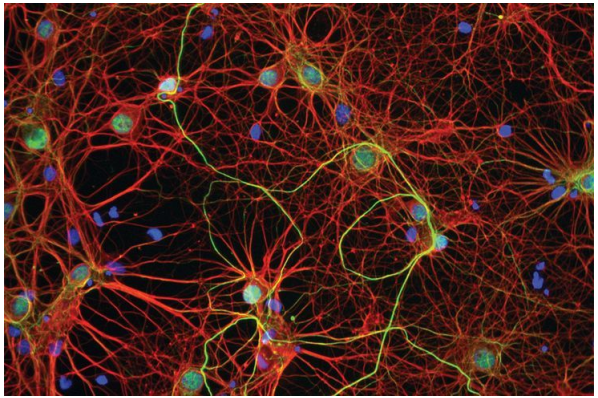

impulses carried
toward cell body
dendrites
branches
of axon
nucleus
axon
axon
terminals
impulses carried
away from cell body
cell body

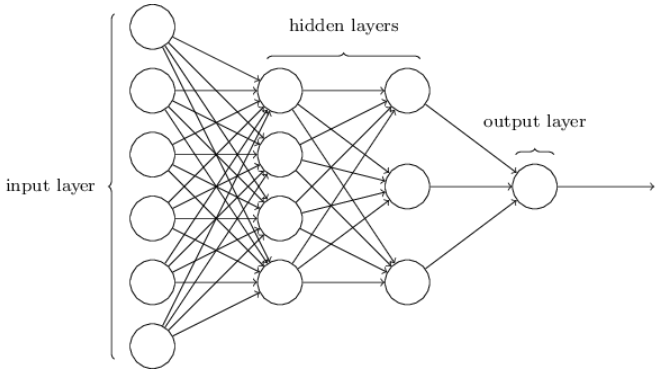






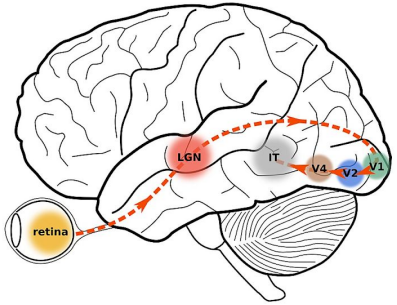
http://cs231n.github.io/neural-networks-1/





McCulloch and Pitts, 1943, Rosenblatt, 1958, Rumelhart et al., 1985, Werbos, 1974

CO



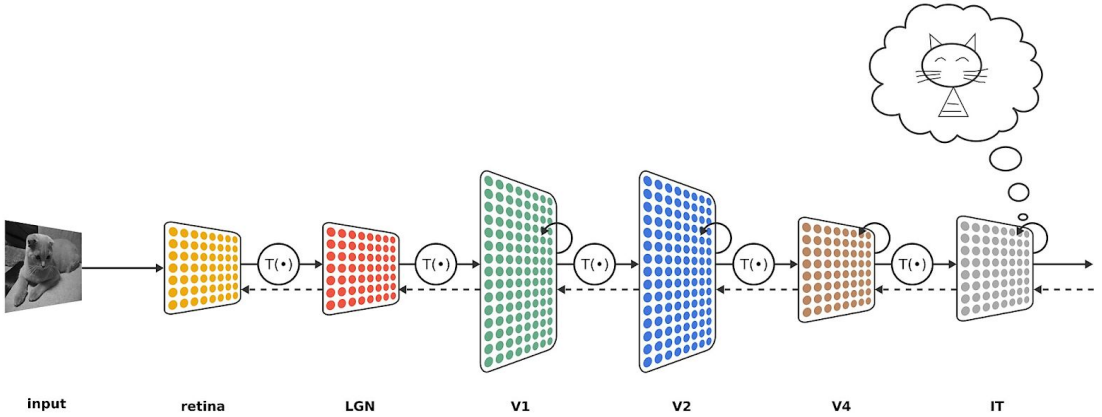
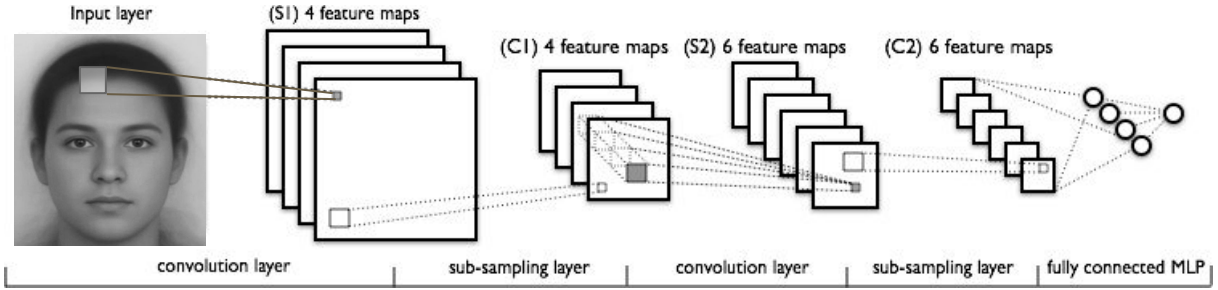
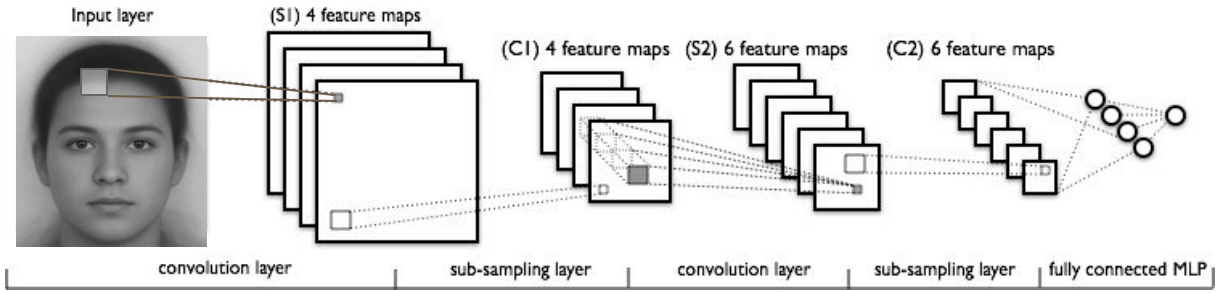


Image: Jonas Kubilias *https://figshare.com/articles/Ventral_visual_stream/106794*

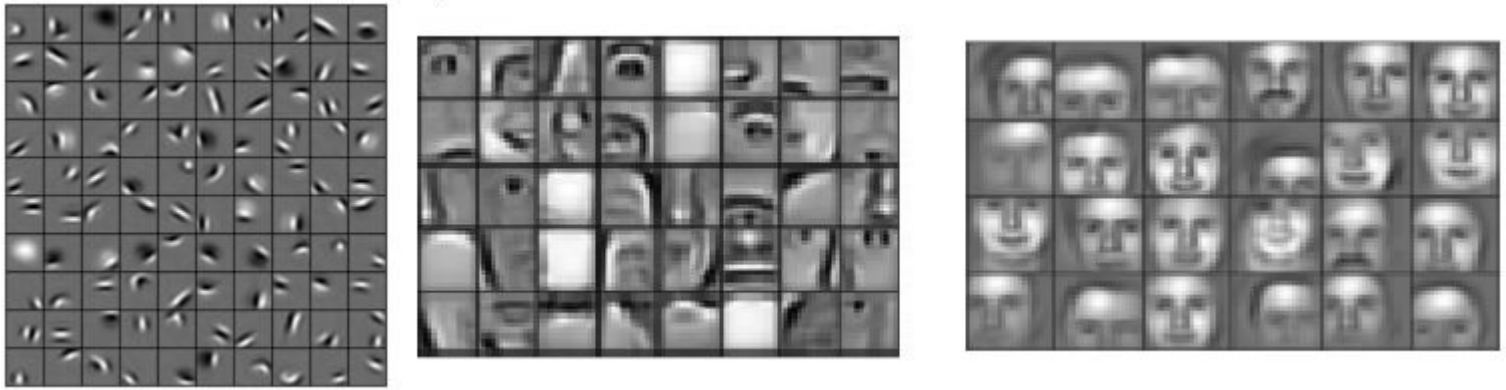
Convolutional neural network



https://deeplearning4j.org/convolutionalnetwork

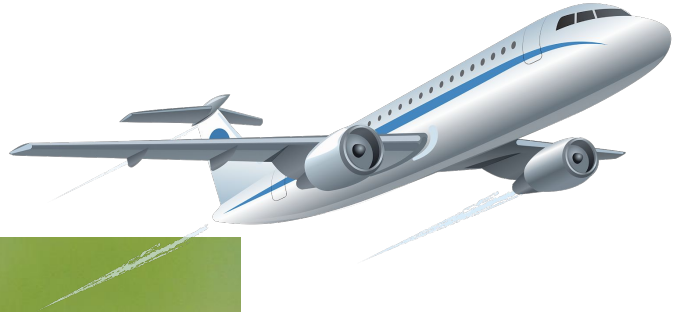


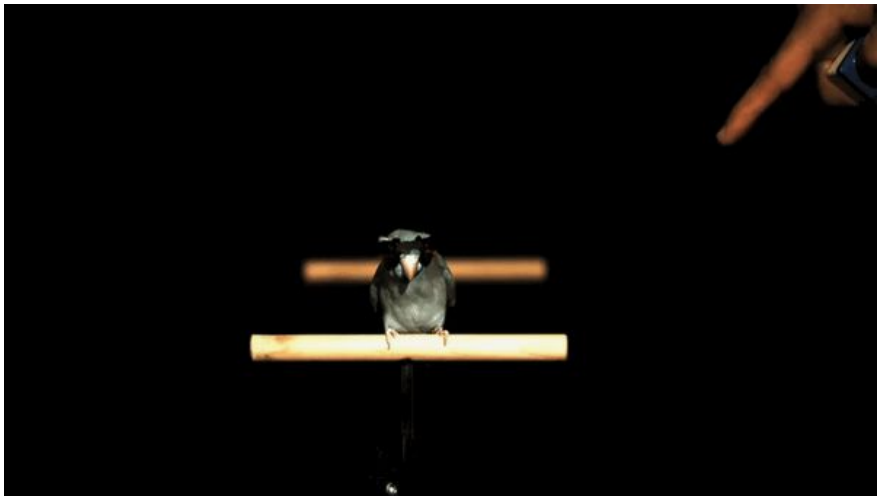
https://deeplearning4j.org/convolutionalnetwork



Lee et al. (2011). "Unsupervised Learning of Hierarchical Representations with Convolutional Deep Belief Networks." 54 (10), Comm. ACM

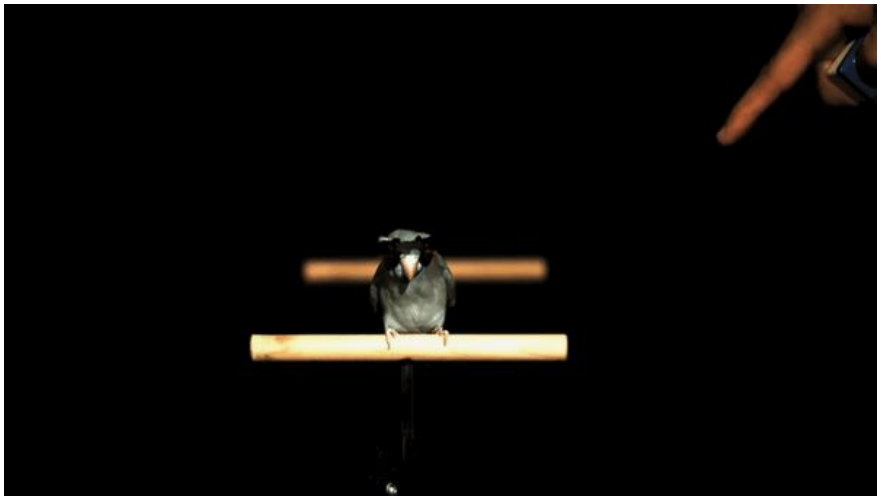
Airplanes don’t fly the way birds do, so what can neuroscience possibly contribute to AI?





From Stanford, https://news.stanford.edu/2016/12/05/birds-flying-laser-light-reveal-faults-flight-research/

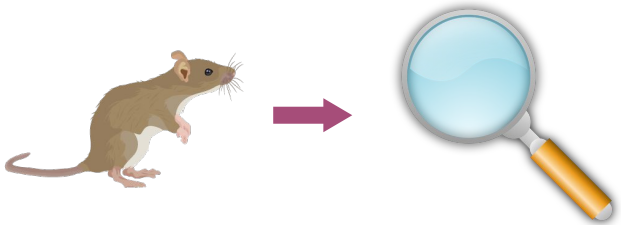
Gutierrez et al. Bioinspiration & Biomemetics. 2016.



From Stanford, https://news.stanford.edu/2016/12/05/birds-flying-laser-light-reveal-faults-flight-research/

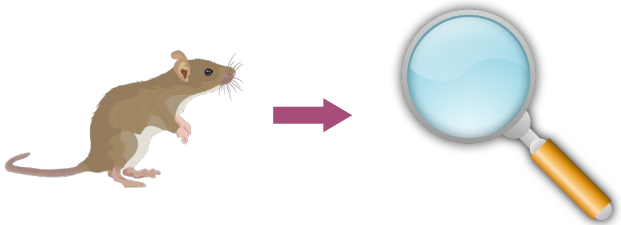
Gutierrez et al. Bioinspiration & Biomemetics. 2016.

https://www.purdue.edu/newsroom/releases/2019/Q2/hummingbird-robot-uses-ai-to-soon-go-where-drones-cant.html



The agent is already “fully formed”, both through previous experience, and evolutionarily programmed instinctual responses.

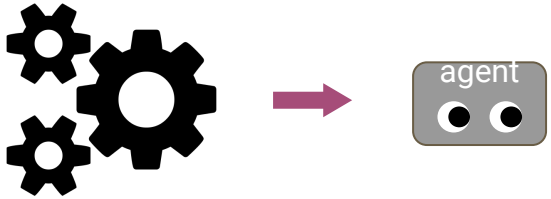
Goal is then to try to infer already existing behavioral patterns, and to match these to neural signatures.



The agent is already “fully formed”, both through previous experience, and evolutionarily programmed instinctual responses.

Goal is then to try to infer already existing behavioral patterns, and to match these to neural signatures.

Architecture, learning rule, objective function, tasks

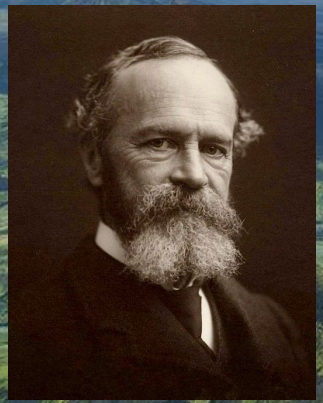


The training process is specified first, with the goal of creating an agent

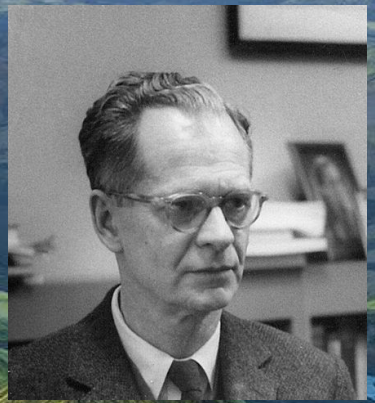
- Architecture, learning rule, objective function *(Richards et al 2019, Nat Neuro Rev; Marblestone et al 2016, Front. Comp. Neurosci.)*
- Task/environmental demands, which constrains behavior and determines the formation of priors

A vast landscape of neuroscience

History
- Psychology (1800s)



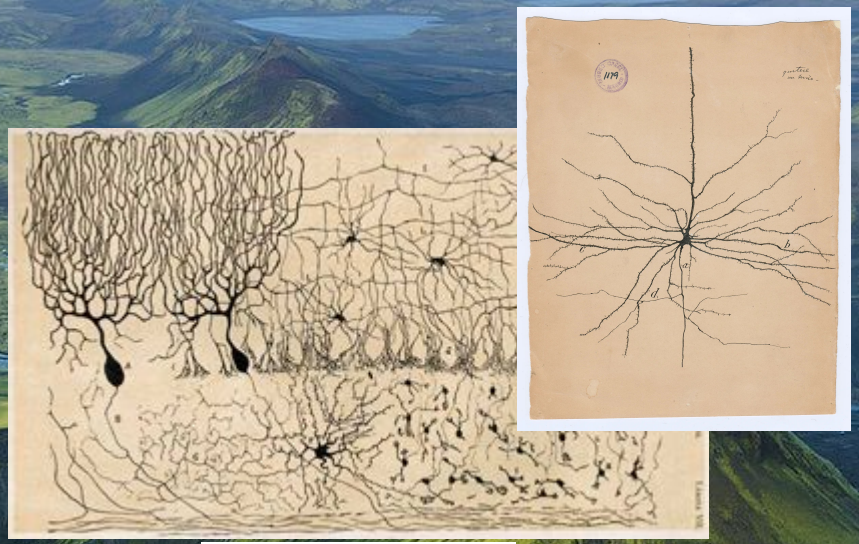
William James
1842-1910



B. F. Skinner
1904-1990

A vast landscape of neuroscience

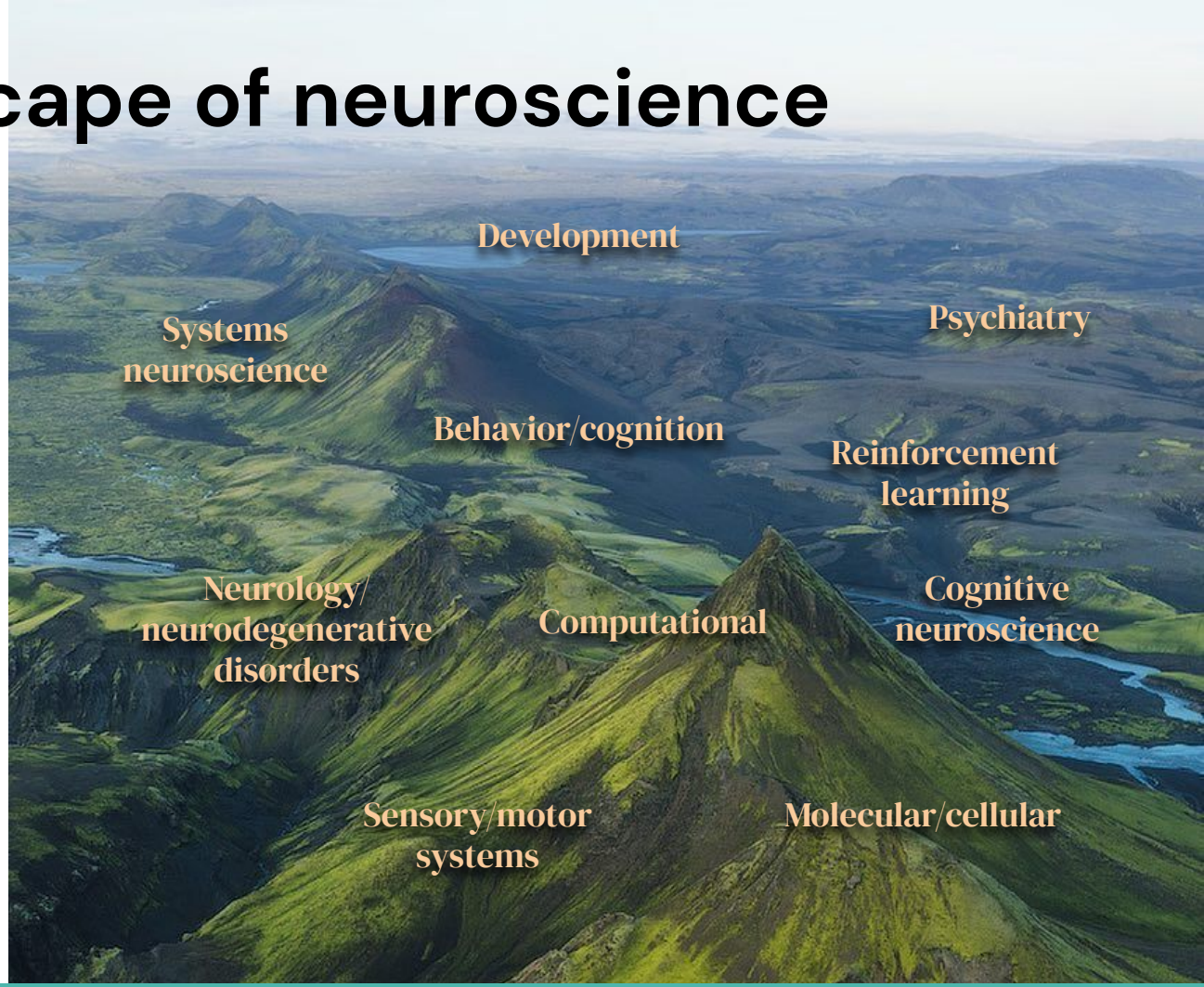
History
- Psychology (1800s)
- Neuroscience (1900s)



Ramon y Cajal, 1890s

A vast landscape of neuroscience

History
- Psychology
- Neuroscience
- Cognitive science



A vast landscape of neuroscience

History
- Psychology
- Neuroscience
- Cognitive science

Society for Neuroscience
- ~30K-35K attendees/yr
- 14K posters
- 

Development

Systems neuroscience

Psychiatry

Behavior/cognition

Reinforcement learning

Neurology/ neurodegenerative disorders

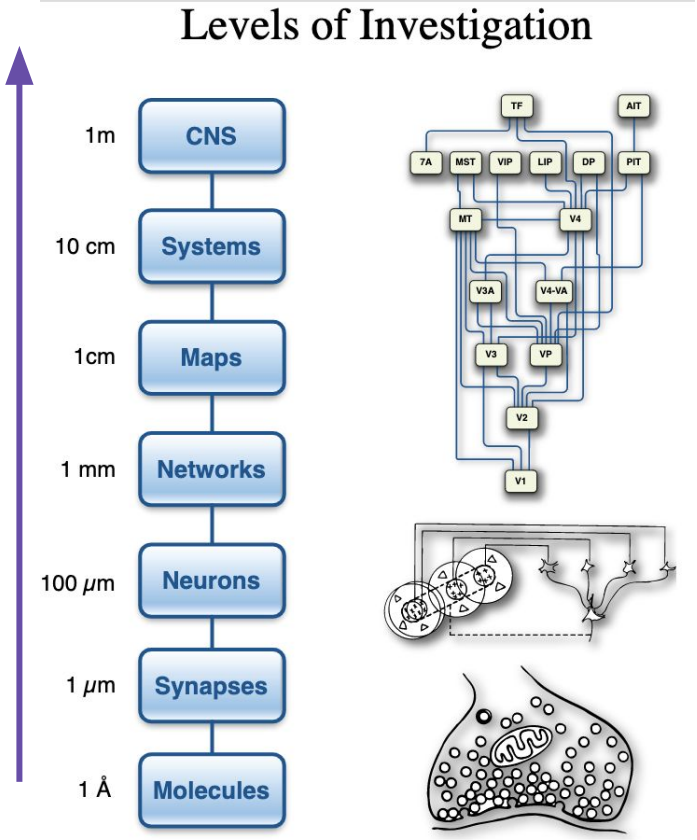
Computational

Cognitive neuroscience

Sensory/motor systems

Molecular/cellular

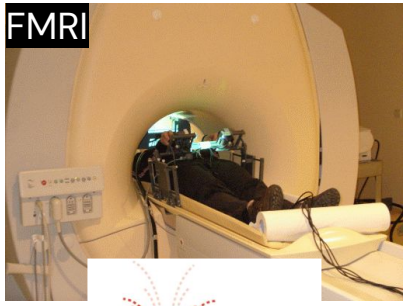
Another perspective: neuroscience by scale

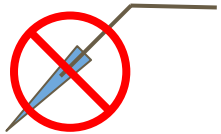


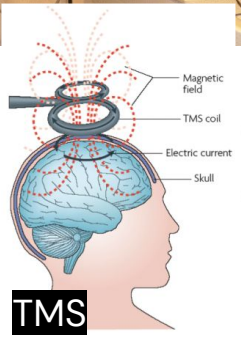
Churchland, Sejnowski. (1998) *Science*

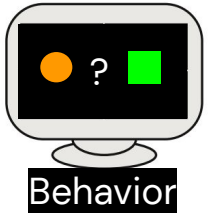
- Interpersonal interactions / sociology
- Psychology
- Cognition
- Higher order brain regions / complex behavior
- Brain regions / beginnings of behavior
  - (sensing, motor outputs)
- Groups of neurons (populations)
- Single neuron
- Cellular/molecular

Humans

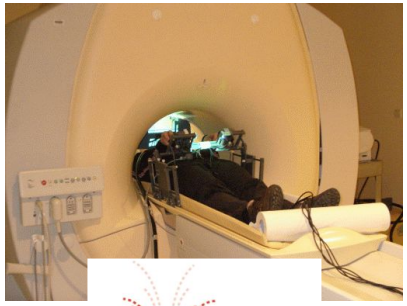

FMRI

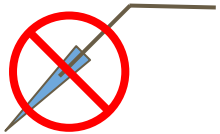


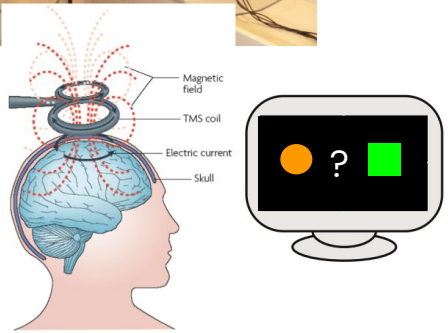

Magnetic field
TMS coil
Electric current
Skull
TMS


?
Behavior

Humans



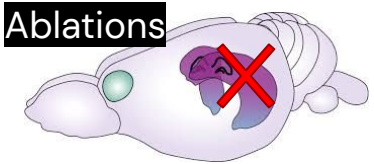




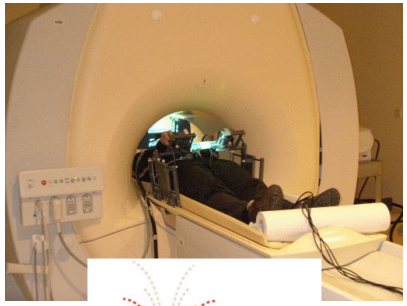
Animals

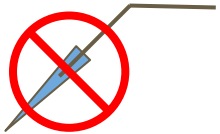
Electrophysiology

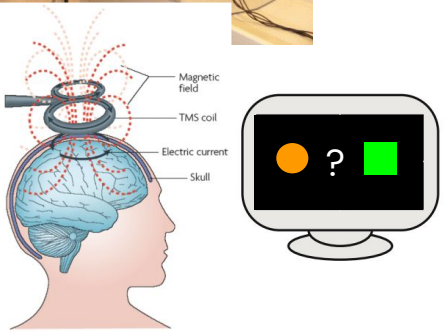
Ablations



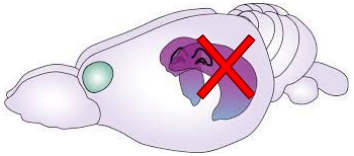
Humans



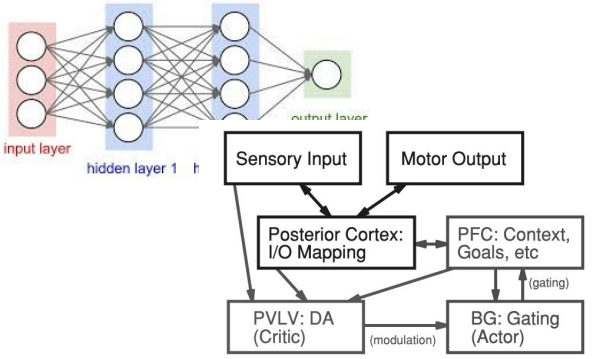




Animals



Cognitive / computational modeling

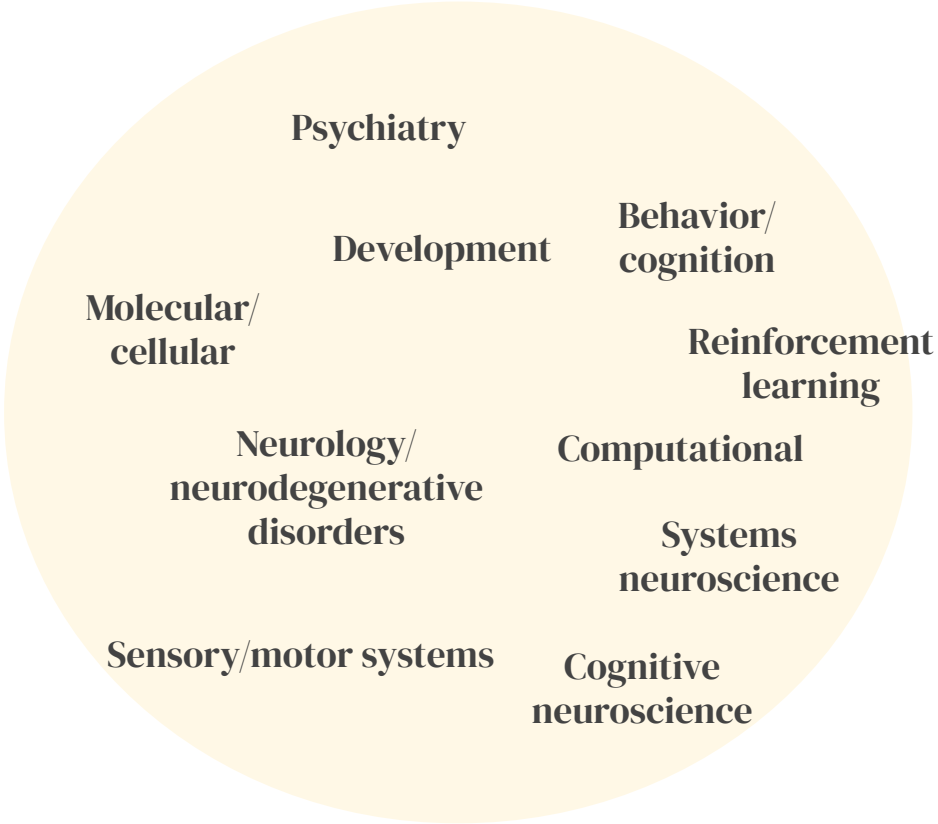


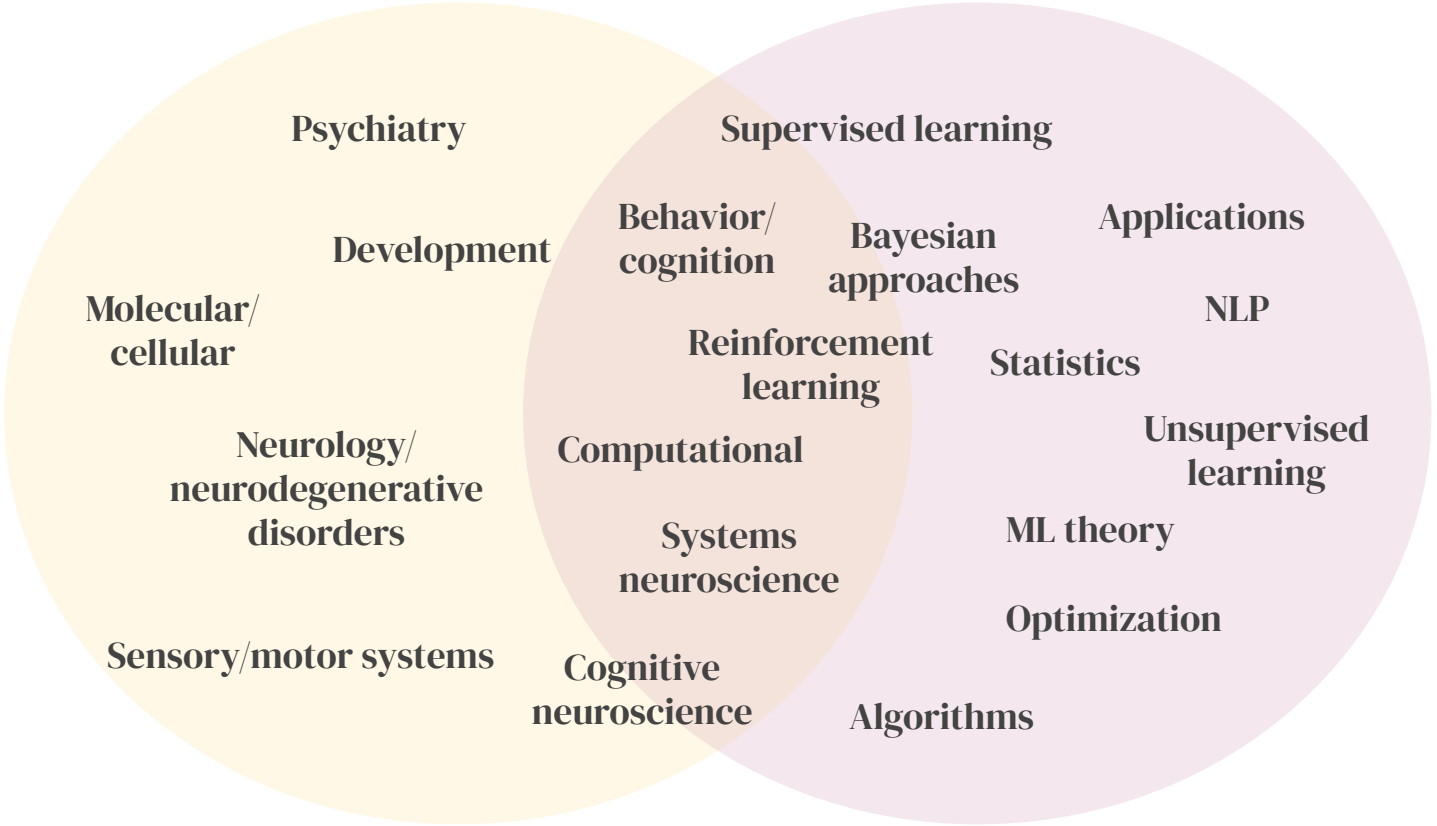
O'Reilly & Frank. 2006 Neural Comput

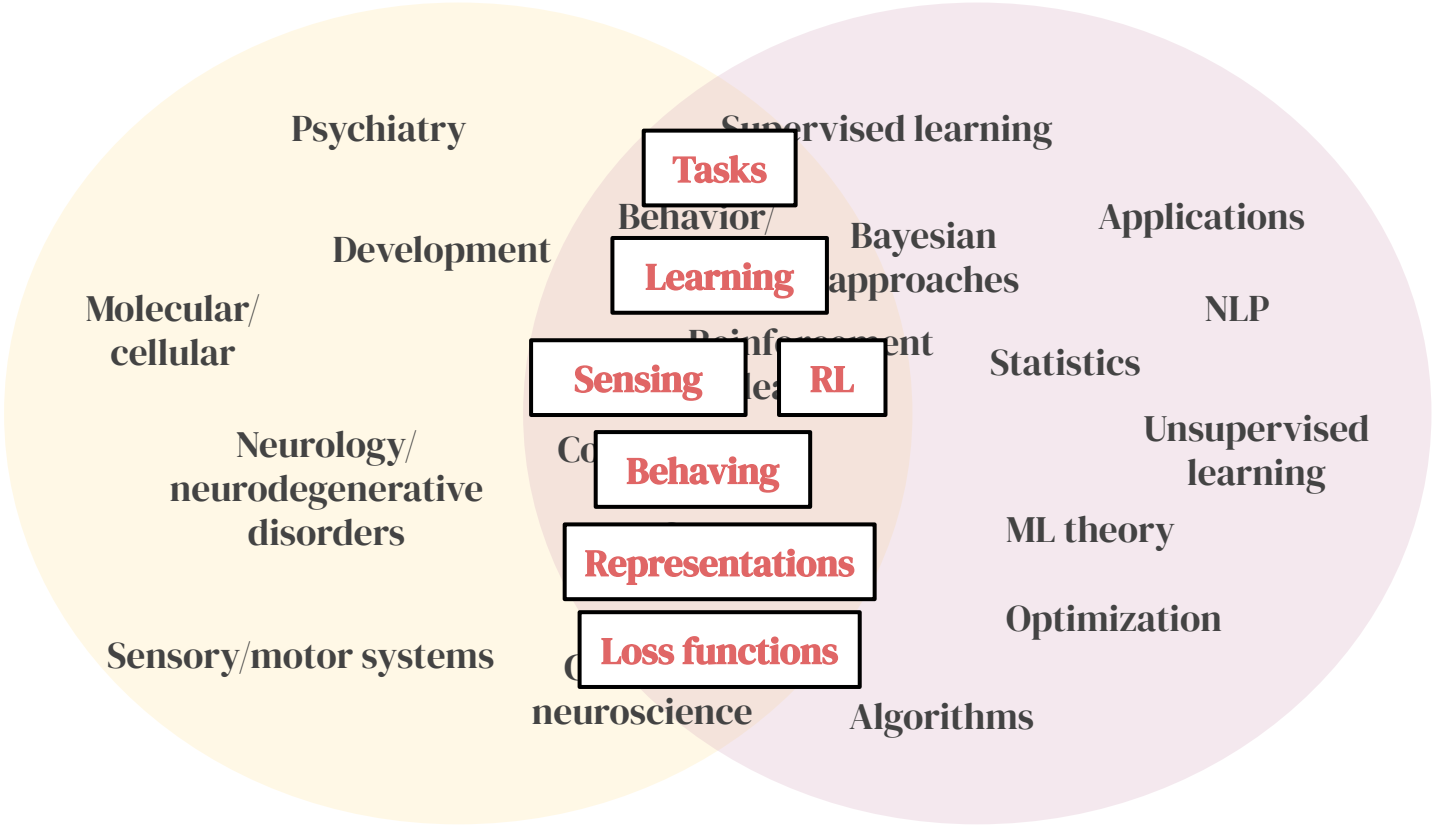
$$Q_{st,at} = Q_{st,at} + \alpha * \left(r_t + \gamma * \max Q(st+1, a) - Q_{st,at}\right)$$

Learning rate · Reward · Discount factor

New value · Current value · Future value estimate


Psychiatry
Development
Behavior/
cognition
Molecular/
cellular
Reinforcement
learning
Neurology/
neurodegenerative
disorders
Computational
Systems
neuroscience
Sensory/motor systems
Cognitive
neuroscience


Psychiatry
Development
Molecular/
cellular
Neurology/
neurodegenerative
disorders
Sensory/motor systems
Cognitive
neuroscience
Behavior/
cognition
Computational
Systems
neuroscience
Reinforcement
learning
Supervised learning
Bayesian
approaches
Applications
NLP
Statistics
Unsupervised
learning
ML theory
Optimization
Algorithms


Psychiatry
Tasks
Behavior/
Development
Learning
Molecular/
cellular
Sensing
RL
Neurology/
neurodegenerative
disorders
Behaving
Representations
Sensory/motor systems
Loss functions
neuroscience
Supervised learning
Applications
Bayesian
approaches
NLP
Statistics
Unsupervised
learning
ML theory
Optimization
Algorithms

Outline

1. Cognitive neuroscience

2. Learning circuits and mechanistic neuroscience

3. Recent advancements at the intersection