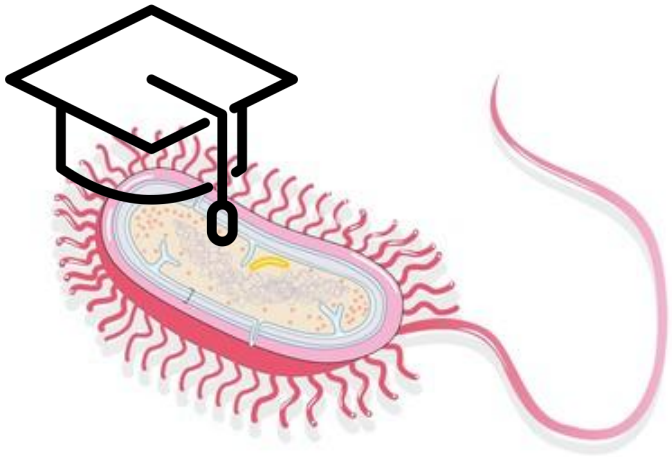
General Principles of Human and Machine Learning

Tutorial 1: Introduction

What is learning? Is it a monolithic concept or composed of different heterogeneous capabilities?

- Does a single cellular organism learn?
- Does a tomato plant learn?
- Does a meteorological system learn (e.g., to cope with climate change)

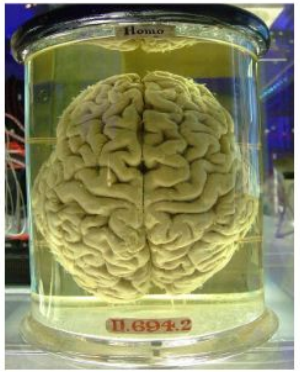


“Learning, in its most basic form can be seen as the process by which we become able to use past and current events to predict what the future holds” (Niv & Schoenbaum, 2008)

- How accurate is this statement?
- How would you amend this definition?

What aspects of learning do we expect to be the same across biological and artificial systems? What do we expect to be different?

- What are fundamental differences between how biological and artificial systems learn? How is the learning problem different?
- How is the learning problem the same?

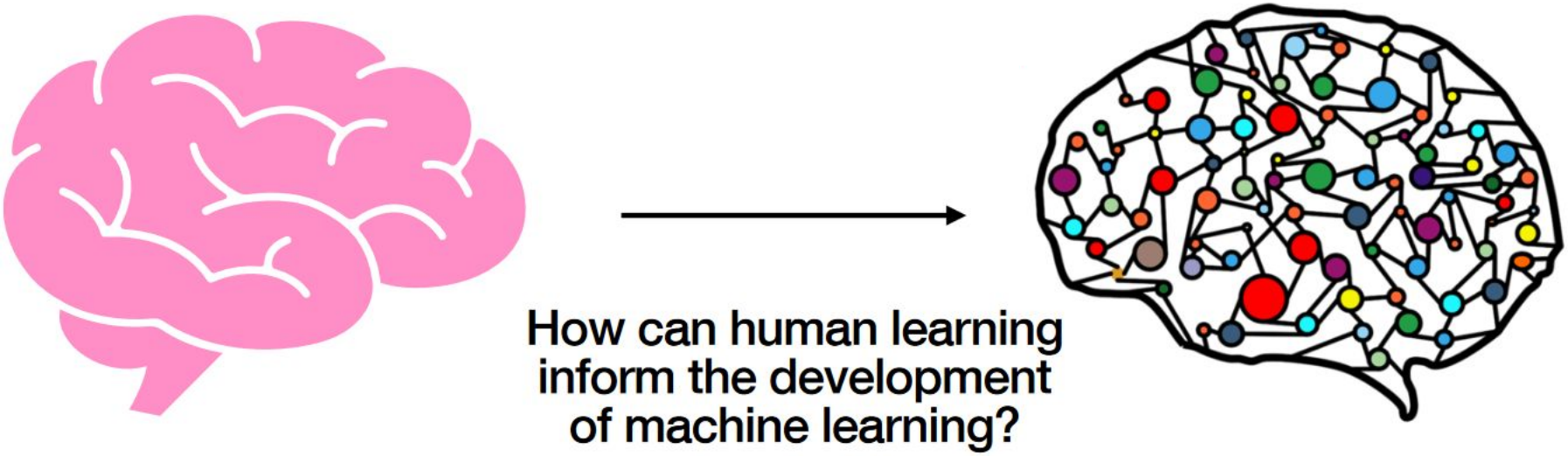


What aspects of learning do we expect to be the same across biological and artificial systems? What do we expect to be different?

- **What are fundamental differences between how biological and artificial systems learn? How is the learning problem different?**
  - differences in access to data, different computational constraints, costs of errors

- **How is the learning problem the same?**
  - Stochasticity, partial observability, the need for generalization

How can the study of biological intelligence inform us about artificial systems?



How can the study of biological intelligence inform us about artificial systems?

- Design principles
- Resource rationality
- Heuristics
- Also, recently, applying cognitive science methods trying to understand what GPT is up to

What can artificial intelligence teach us about biological intelligence?

How can machines inform our understanding of human learning?

What can artificial intelligence teach us about biological intelligence?

- Rational model
- Instantiation of a theory
- Inspiration for investigations

Spatial vs. logical vs. ANN methods of representation learning

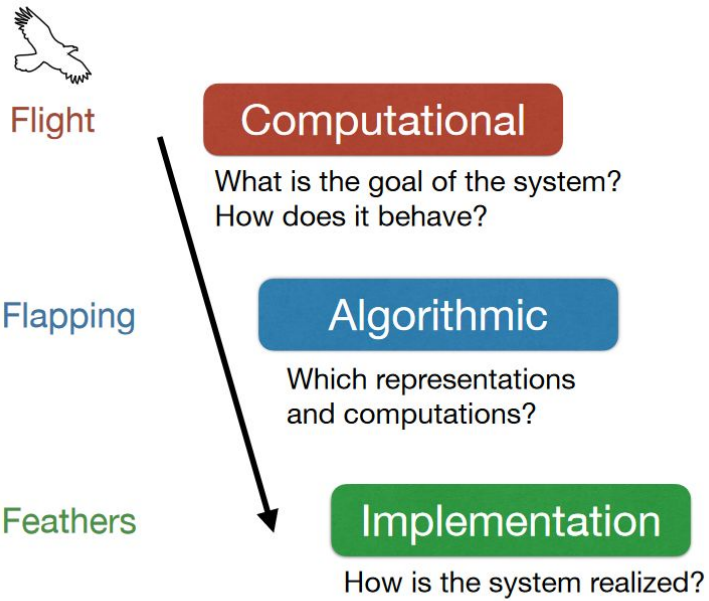
- What are the advantages of spatial vs. logical methods?
- Are neural networks really a third category? How are they similar to spatial representation learning? How are they similar to logical methods? How are they dissimilar?
- Are all three approaches helpful for understanding biological intelligence? In what ways are the underlying assumptions broken when applied to human learning?

Spatial vs. logical vs. ANN methods of representation learning

- What are the advantages of spatial vs. logical methods?
  - Spatial: flexible, easily generalizes
  - Logical: interpretable, compositional
- Are neural networks really a third category? How are they similar to spatial representation learning? How are they similar to logical methods? How are they dissimilar?
  - ANNs are a different Marr's level – they can behave like either method
- Are all three approaches helpful for understanding biological intelligence? In what ways are the underlying assumptions broken when applied to human learning?
  - Yes! We can use different models for different phenomena, and depending on what we find interesting, different models can be more or less useful for our purpose. Humans tend to behave more flexibly than one set model (you can do both spatial and logical representation!

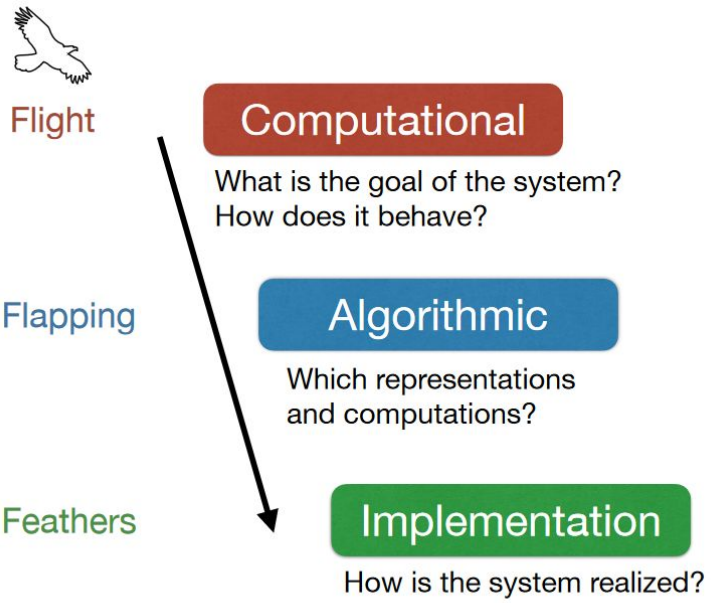
Discuss Marr's Levels

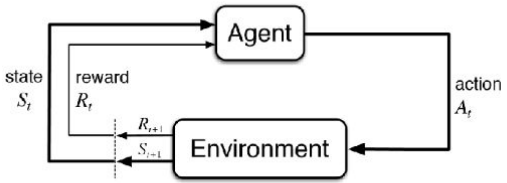
- How can we describe learning at these different levels of description?



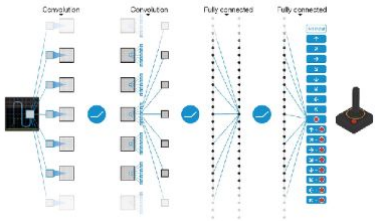
Discuss Marr's Levels

- How can we describe learning at these different levels of description?

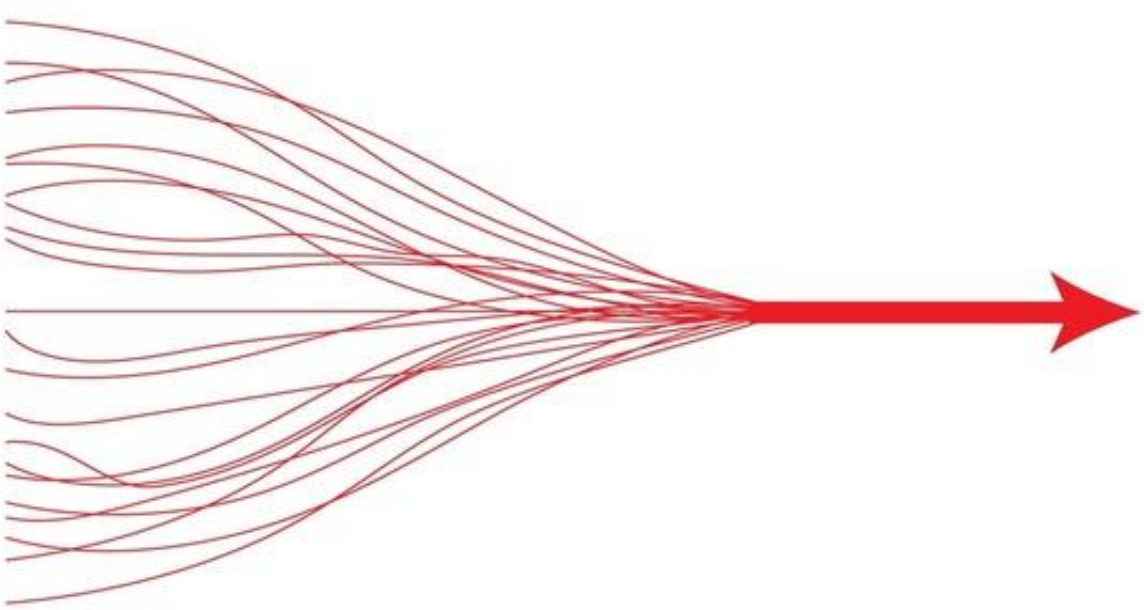




Initialize $Q(s,a)$ arbitrarily
Repeat (for each episode):
  Initialize s
  Repeat (for each step of episode):
    Choose a from s using policy derived from Q (e.g., ε-greedy)
    Take action a, observe r, s'
    $Q(s,a) \leftarrow Q(s,a) + \alpha\left[r + \gamma \max_{a'} Q(s',a') - Q(s,a)\right]$
    $s \leftarrow s'$;
  until s is terminal



Which principles of cognition have we seen converge between human and machine learning implementations?



Which principles of cognition have we seen converge between human and machine learning implementations?

- Learning from prediction error
- Building models of the environment
- The need for lossy compression and throwing away data
- Building representations that help solve problems rather than capturing maximum fidelity
- Explore-exploit trade-off in active learning