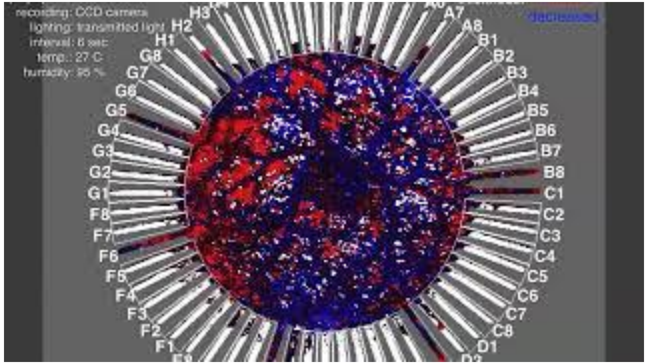
Tutorial 1

General Principles of Human & Machine Learning

TA: Mani (mani.hamidi@uni-tuebingen.de)

1. What is learning?

- Is it a monolithic concept or composed of different heterogeneous capabilities?
  a. Does a single cellular organism learn?
  b. Does a tomato plant learn?
  c. Does meteorological system learn (e.g., to cope with climate change)
  d. Does a “smart” mattress learn the shape of your hand?



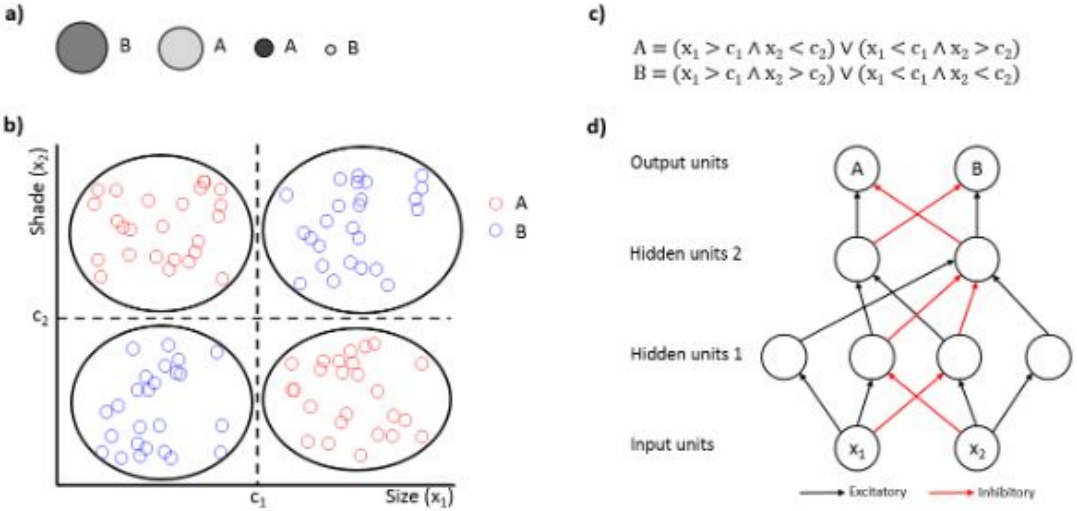
Learning as …

- Information processing
  - Does a calculator learn?
- Knowledge acquisition
  - Beware of homunculi.
- Search:
  - For questions or for answers?
  - Does a heat seeking missile learn the position of target?
  - Is evolution a learning algorithm?
- Inference/prediction
  - "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?

Inference

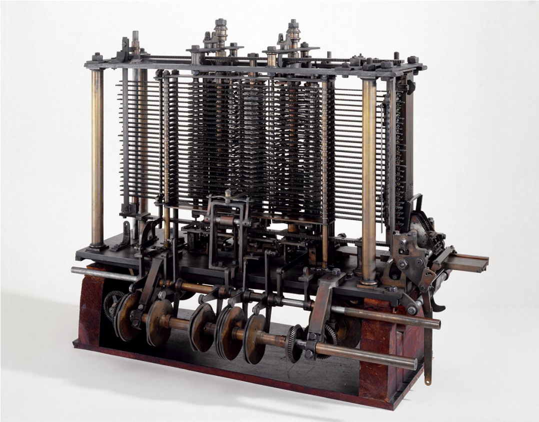
- Logical (symbolic)
- Probabilistic (Bayesian)

Deduction:	All the beans from this bag are white.	(rule)
	These beans are from this bag.	(case)
	⇒ These beans are white.	(result)
Induction:	These beans are white.	(result)
	These beans are from this bag.	(case)
	⇒All the beans from this bag are white.	(rule)
Abduction:	All the beans from this bag are white.	(rule)
	These beans are white.	(result)
	⇒These beans are from this bag.	(case)

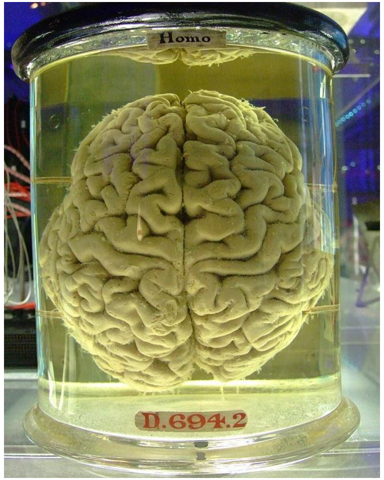


2. Biology vs. Machine

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

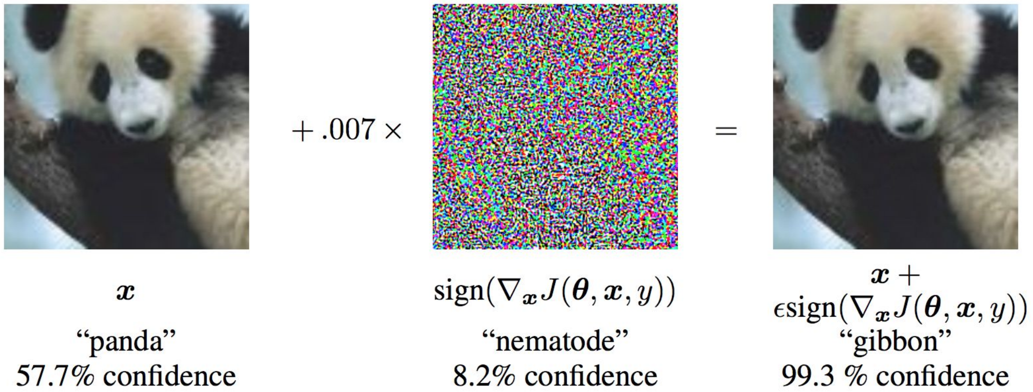


Babbage's Turing-complete "analytical engine"



2. Biology vs. Machine

- What are fundamental differences between how biological and artificial systems learn? How is the learning problem different?
  - E.g., differences in access to data, different computational constraints, costs of errors, etc…
- How is the learning problem the same?
  - Stochasticity
  - Partial observability
  - The need for generalization



2. Biology vs. Machine

1. What can the study of biological intelligence inform us about artificial systems?
    - E.g., design principles, resource rationality, heuristics, embodiment

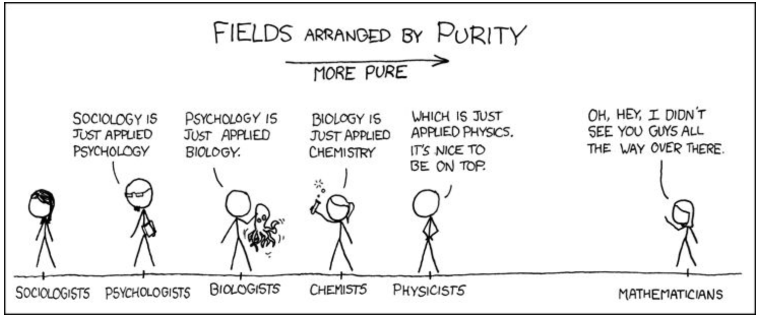
2. What can artificial intelligence teach us about biological intelligence?
    - E.g., as a comparison to a rational model, as an instantiation of a theory (components XY are necessary and sufficient to generate behavior Z)

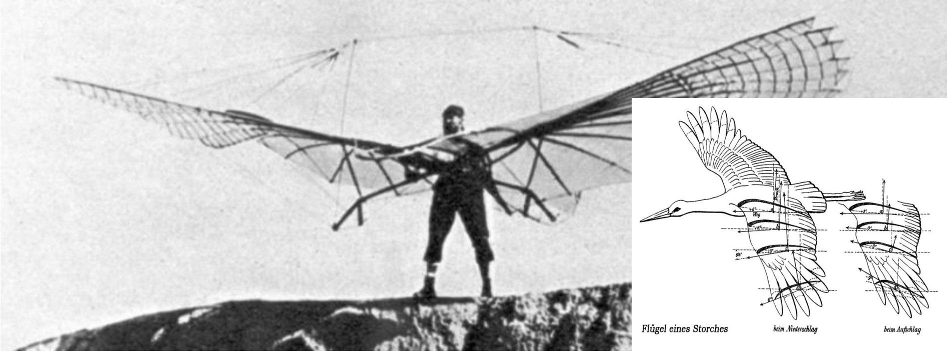
2. Biology vs. Machine

[What have we learned about artificial intelligence from studying the brain?](#) Gershman 2023

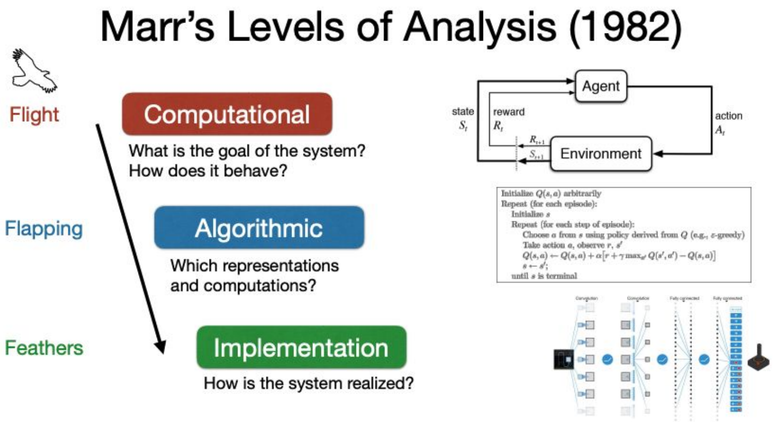
Convergence Heuristic:

“If engineers propose an algorithm and neuroscientists find evidence for it in the brain, it is a pretty good clue that the algorithm is on the right track...This **convergence heuristic** is consistent with **current computational neuroscience practice**, where AI has historically provided a fund of ideas for biological theories. It is also consistent with current **AI practice**, where researchers are primarily looking for directional signals from neuroscience...rather than specific algorithms.”





3. Marr’s Levels



Key Idea:

- Multiple realizability at each level

Another Example: Emotions

[From: Algorithms for survival: a comparative perspective on emotions](#) (Bach & Dayan 2017)

Box 1 | Levels of theoretical analysis

In computational neuroscience, it is common to distinguish different levels of analysis that go back to Marr[21].

Computational level

At the computational level[21], theoretical analysis focuses on formalizing the problem that the nervous system has to solve and on finding an appropriate, often optimal or normative, solution. One optimal solution to any decision-making problem is given by Bayesian decision theory (BDT)[112]. According to this theory, agents should create and maintain a so-called belief state that summarizes the whole history of their past observations. To do so, they must use what is known as a generative model of the possible trajectories of environmental states and how those states generate sensory data (note that the ‘environment’ in this case encompasses the body of the agent). Agents should then make the choices that maximize average long-run benefit by computing an expectation over all possible present and future states along such trajectories. The long-run benefit is typically a weighted sum of the utilities of each possible outcome in the future, with more weight given to outcomes that occur sooner (temporal discounting). Specifying these outcome values is therefore a key ingredient of BDT. The BDT solution is a benchmark that no natural or artificial agent can surpass.

Algorithmic level

The algorithmic level of analysis concerns how a given problem is solved. Various fields have suggested exact and approximate algorithmic approaches to BDT. These have been given names such as optimal control theory, dynamic programming and reinforcement learning[113–121]. Approximations are necessary because normative solutions are often analytically intractable and cannot even be computed numerically offline in an exact manner. Many neuroscientists use reinforcement learning theory as a formal framework for stating and solving the decision-making problems that they pose to their subjects.

Implementational level

The implementational level of analysis considers the ways in which algorithms are realized in neural circuits. This spans descriptions on a macroscopic level (brain areas and large populations of neurons), on a mesoscopic level (modestly sized circuits of neurons subject to neuromodulatory influences) and on a microscopic level (within-neuron computations).

- Link to digital brainstorming whiteboard