Lecture Overview

- What is Artificial Intelligence?
- Agents acting in an environment

Artificial Intelligence

Changing random stuff
until your program works is
"hacky" and "bad coding practice".

But if you do it fast enough it is
"Artificial Intelligence"
and pays 4x the best SE salary.

What is Artificial Intelligence?

- Artificial Intelligence is the synthesis and analysis of computational agents that act intelligently.
- An agent is something that acts in an environment.
- An agent acts intelligently if:
  - its actions are *appropriate* for its goals and circumstances
  - it is *flexible* to changing environments and goals
  - it *learns* from experience
  - it makes appropriate choices given *perceptual and computational limitations*

Computational Agents

- A computational agent is an agent whose decisions about its actions can be explained in terms of computation.
- The decisions can be broken down into primitive *operations* that can be implemented in a physical device.

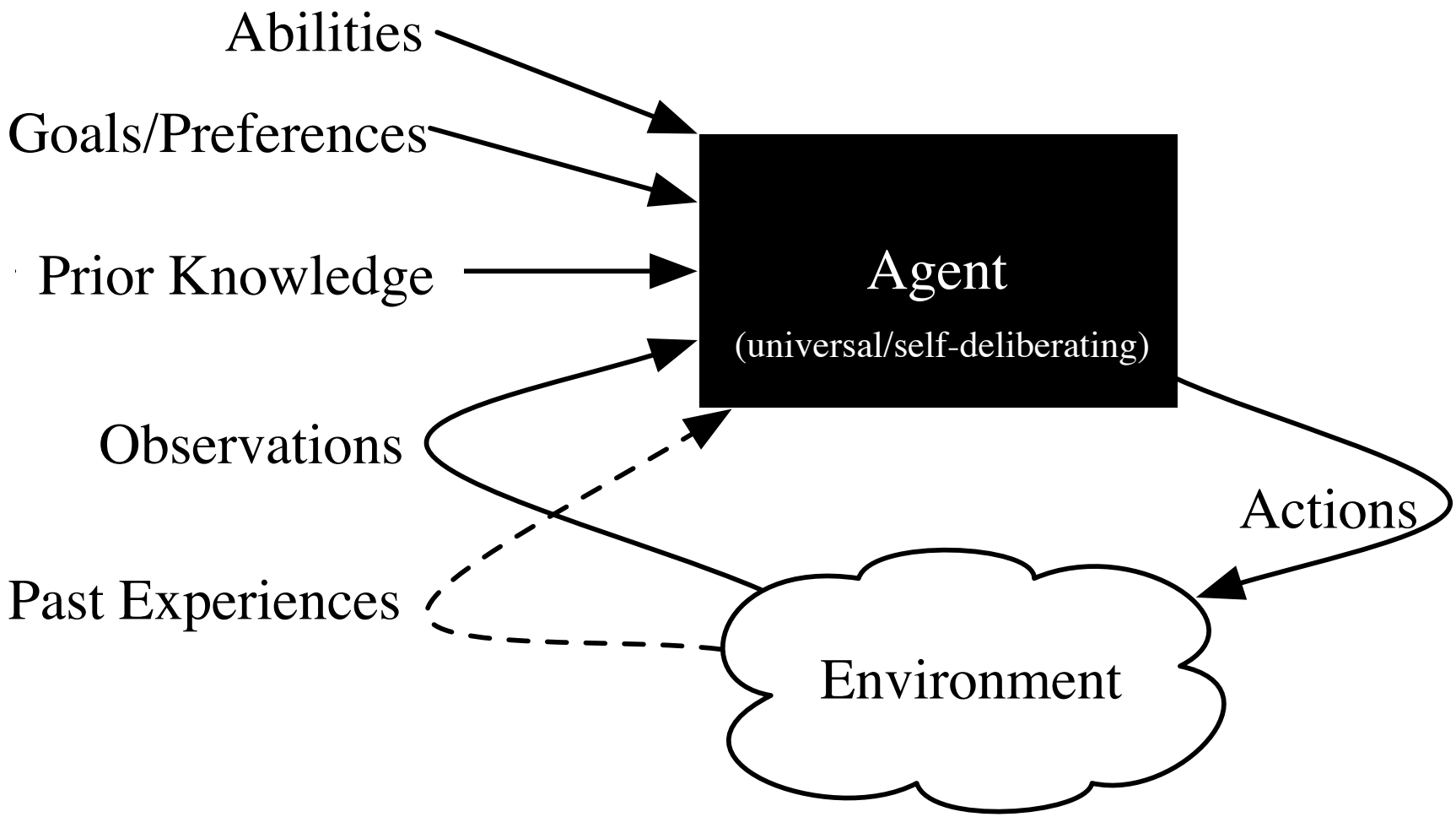
Artificial and Natural Intelligence

- For any phenomenon, you can distinguish real versus fake, where the fake is non-real.
- You can also distinguish natural versus artificial; natural means occurring in nature and artificial means made by people.
- Example:
  - A tsunami is a large wave in an ocean caused by an earthquake or a landslide. Natural tsunamis occur from time to time. You could imagine an artificial tsunami that was made by people, for example, by exploding a bomb in the ocean, yet which is still a real tsunami. One could also imagine fake tsunamis: either artificial, using computer graphics, or natural, for example, a mirage that looks like a tsunami but is not one.

Goals of Artificial Intelligence

- Scientific goal: to understand the principles that make intelligent behavior possible in natural or artificial systems.
  - analyze natural and artificial agents
  - formulate and test hypotheses about what it takes to construct intelligent agents
  - design, build, and experiment with computational systems that perform tasks that require intelligence
- Engineering goal: design useful, intelligent artifacts.
- Analogy between studying flying machines and thinking machines.

Agents acting in an environment: a function!



`Agent: (observation)* → action`

Example agent: robot

- abilities: movement, grippers, speech, facial expressions,...
- observations: vision, sonar, sound, speech recognition, gesture recognition,...
- goals: deliver food, rescue people, score goals, explore,...
- past experiences: effect of steering, slipperiness, how people move,...
- prior knowledge: what is important feature, categories of objects, what a sensor tell us,...

Example agent: teacher

- abilities: present new concept, drill, give test, explain concept,...
- observations: test results, facial expressions, errors, focus,...
- goals: particular knowledge, skills, inquisitiveness, social skills,...
- past experiences: prior test results, effects of teaching strategies, ...
- prior knowledge: subject material, teaching strategies,...

Example agent: medical doctor

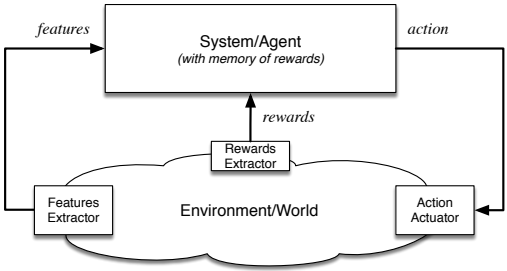
- abilities: operate, test, prescribe drugs, explain instructions,...
- observations: verbal symptoms, test results, visual appearance...
- goals: remove disease, relieve pain, increase life expectancy, reduce costs,...
- past experiences: treatment outcomes, effects of drugs, test results given symptoms...
- prior knowledge: possible diseases, symptoms, possible causal relationships...

Example agent: user interface

- abilities: present information, ask user, find another information source, filter information, interrupt,...
- observations: users request, information retrieved, user feedback, facial expressions...
- goals: present information, maximize useful information, minimize irrelevant information, privacy,...
- past experiences: effect of presentation modes, reliability of information sources,...
- prior knowledge: information sources, presentation modalities...

AI-based Self-Deliberating Dynamic Systems: Reinforcement Learning

$$(features)^* \rightarrow action$$



- **Features**: these are **possibly "low-level" properties** (**possibly far from human thinking**), in fact learning the features themself is a trendy area of research
- **Rewards**: form a sort of **implicit temporally extended task specification** (due to accumulation)
- **Self-deliberating**: automatically learn an agent dynamics/behavior "a policy" such that its execution on the environment maximize the expected rewards over time.
- **Not queryable**: if the features are not high-level, e.g. are learned themself from the data.