Artificial Intelligence

What is Artificial Intelligence

- Basically, there are four school of thoughts
  - Systems that think like humans
  - Systems that act like humans
  - Systems that think rationally
  - Systems that act rationally
- We will follow “act rationally” approach

Turing Test

- The Turing Test, proposed by Alan Turing (1950), was designed to provide a satisfactory operational definition of intelligence
- A computer passes the test if a human interrogator, after posing some written questions, cannot tell whether the written responses come from a person or from a computer
- Yet AI researchers have devoted little effort to passing the Turing Test, believing that it is more important to study the underlying principles of intelligence than to duplicate an exemplar
- The quest for “artificial flight” succeeded when the Wright brothers and others stopped imitating birds and started using wind tunnels and learning about aerodynamics [Russell and Norvig]

Acting rationally: The rational agent approach

- An agent is just something that acts
- Agents operate autonomously, perceive their environment, persist over a prolonged time period, adapt to change, and create and pursue goals
- A rational agent is one that acts so as to achieve the best outcome or, when there is uncertainty, the best expected outcome

- The rational-agent approach has two advantages over the other approaches
- First, it is more general than the “laws of thought” approach because correct inference is just one of several possible mechanisms for achieving rationality (Think about reflex actions)
- Second, it is more amenable to scientific development than are approaches based on human behavior or human thought
- Therefore, the concept of rational agents is central to our approach to artificial intelligence

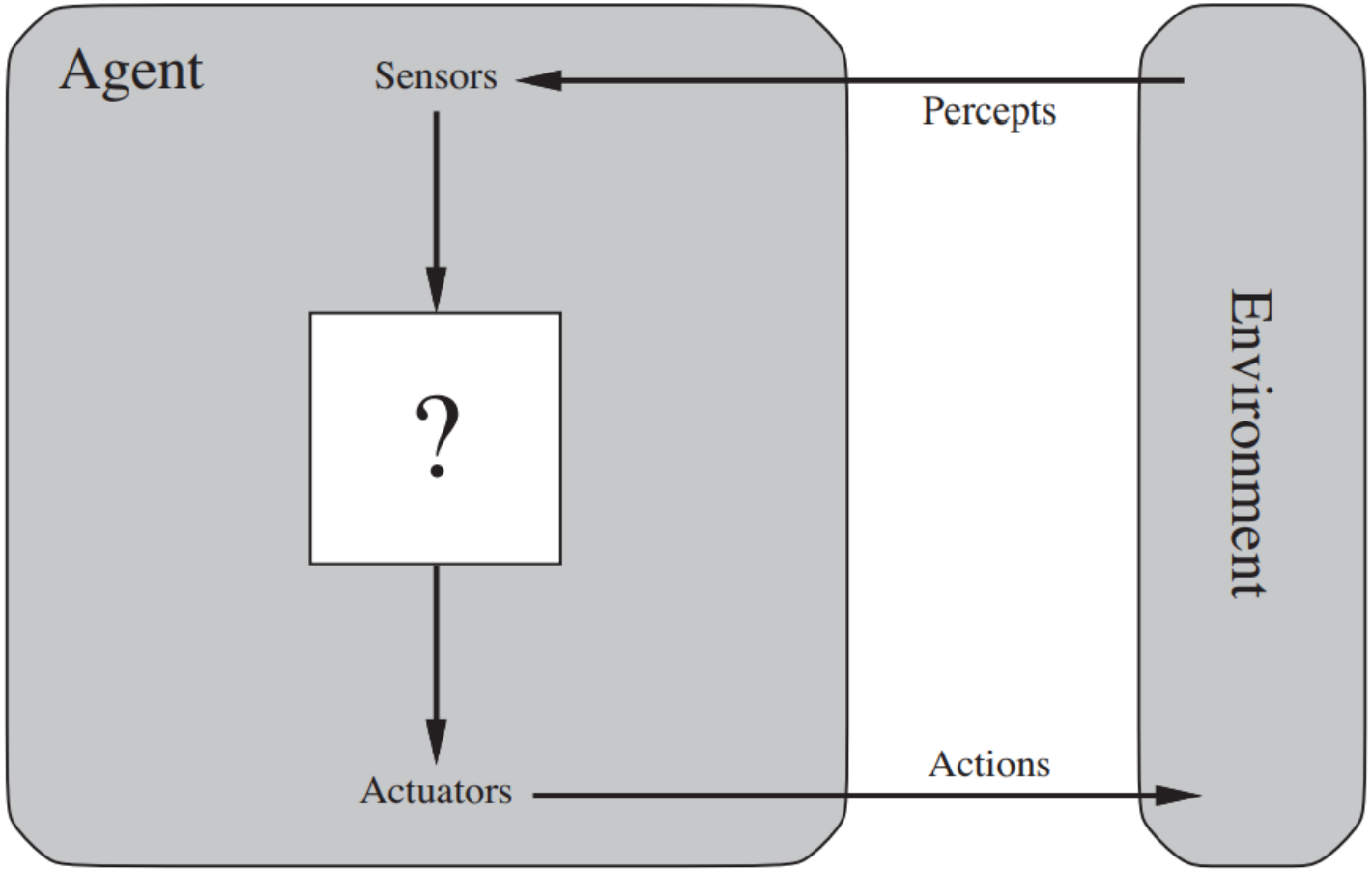
Agents

- An agent is anything that can be viewed as perceiving its environment through sensors and acting upon that environment through actuators
- A human agent has eyes, ears, and other organs for sensors and hands, legs, vocal tract, and so on for actuators
- A robotic agent might have cameras and infrared range finders for sensors and various motors for actuators
- A software agent receives keystrokes, file contents, and network packets as sensory inputs and acts on the environment by displaying on the screen, writing files, and sending network packets

Percept

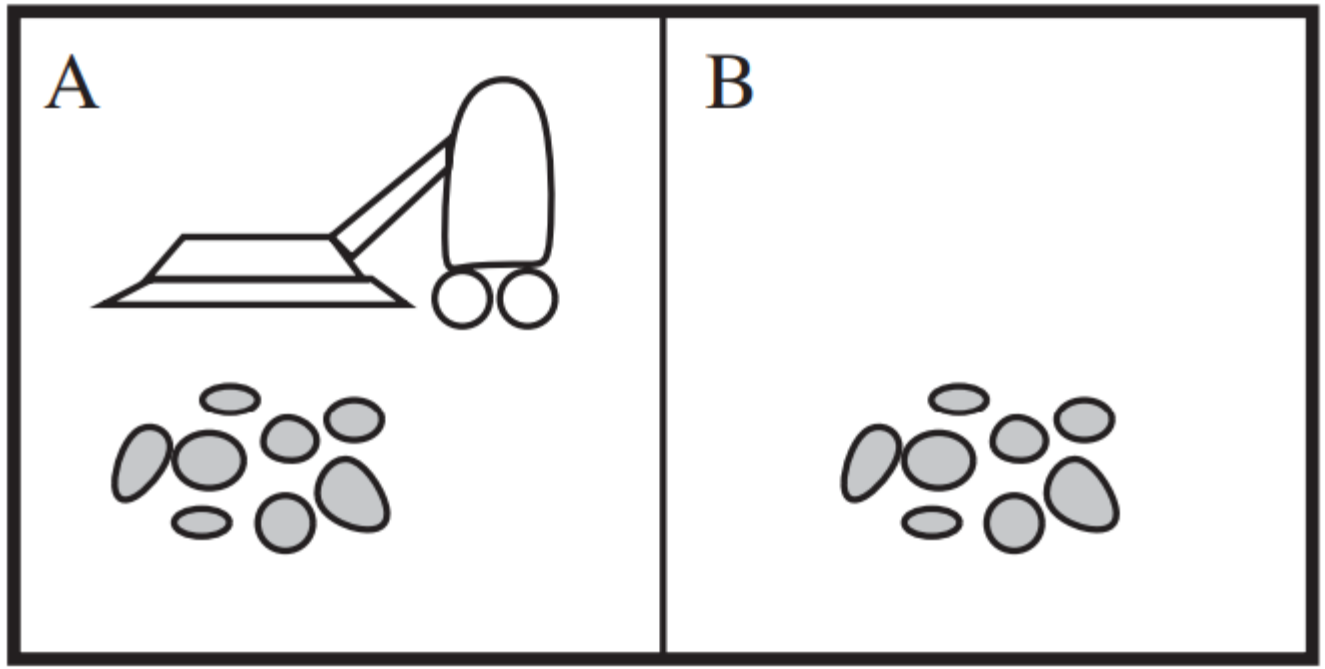
- The term percept refer to the agent’s perceptual inputs at any given instant
- An agent’s percept sequence is the complete history of everything the agent has ever perceived
- An agent’s choice of action at any given instant can depend on the entire percept sequence observed to date, but not on anything it hasn’t perceived
- Mathematically speaking, we say that an agent’s behavior is described by the agent function that maps any given percept sequence to an action

Agents interact with environments through sensors and actuators



- The agent function for an artificial agent will be implemented by an agent program
- The agent function is an abstract mathematical description; the agent program is a concrete implementation, running within some physical system

A vacuum-cleaner world with just two locations



Partial tabulation of a simple agent function for the vacuum-cleaner world

Percept sequence	Action
$[A, Clean]$	$Right$
$[A, Dirty]$	$Suck$
$[B, Clean]$	$Left$
$[B, Dirty]$	$Suck$
$[A, Clean], [A, Clean]$	$Right$
$[A, Clean], [A, Dirty]$	$Suck$
⋮	⋮
$[A, Clean], [A, Clean], [A, Clean]$	$Right$
$[A, Clean], [A, Clean], [A, Dirty]$	$Suck$
⋮	⋮

The concept of rationality

- A rational agent is one that does the right thing—conceptually speaking, every entry in the table for the agent function is filled out correctly
- What does it mean to do the right thing?
- Can it be decided by considering the consequences of the agent's behavior? (yes exactly)

- Agents generates a sequence of actions according to the percepts it receives
- This sequence of actions causes the environment to go through a sequence of states
- If the sequence is desirable, then the agent has performed well
- This notion of desirability is captured by a performance measure that evaluates any given sequence of environment states

- Notice that we said environment states, not agent states
- If we define success in terms of agent's opinion of its own performance, an agent could achieve perfect rationality simply by deluding itself that its performance was perfect
- Human agents in particular are notorious- believing they did not really want something (e.g., a Nobel Prize) after not getting it
- Obviously, there is not one fixed performance measure for all tasks and agents; typically, a designer will devise one appropriate to the circumstances
- As a general rule, it is better to design performance measures according to what one actually wants in the environment, rather than according to how one thinks the agent should behave

- Issue- suppose the notion of “clean floor” is based on average cleanliness over time
- Yet the same average cleanliness can be achieved by two different agents
- One of which does a mediocre job all the time while the other cleans energetically but takes long breaks
- It is a deep philosophical question with far-reaching implications
- Which is better—an economy where everyone lives in moderate poverty, or one in which some live in plenty while others are very poor?

Rationality

- What is rational at any given time depends on four things
  - -The performance measure that defines the criterion of success
  - -The agent's prior knowledge of the environment
  - -The actions that the agent can perform
  - -The agent's percept sequence to date
- For each possible percept sequence, a rational agent should select an action that is expected to maximize its performance measure, given the evidence provided by the percept sequence and whatever built-in knowledge the agent has

Case study

- Consider the simple vacuum-cleaner agent that cleans a square if it is dirty and moves to the other square if not
- Is this a rational agent? That depends! First, we need to say what the performance measure is, what is known about the environment, and what sensors and actuators the agent has

  -The performance measure awards one point for each clean square at each time step, over a "lifetime" of 1000 time steps

  - The "geography" of the environment is known a priori but the dirt distribution and the initial location of the agent are not. Clean squares stay clean and sucking cleans the current square. The Left and Right actions move the agent left and right except when this would take the agent outside the environment, in which case the agent remains where it is

  - The only available actions are Left, Right, and Suck

  - The agent correctly perceives its location and whether that location contains dirt
- We claim that under these circumstances the agent is indeed rational

- One can see easily that the same agent would be irrational under different circumstances
- For example, once all the dirt is cleaned up, the agent will oscillate needlessly back and forth; if the performance measure includes a penalty of one point for each movement left or right, the agent will fare poorly
- A better agent for this case would do nothing once it is sure that all the squares are clean
- If clean squares can become dirty again, the agent should occasionally check and re-clean them if needed
- If the geography of the environment is unknown, the agent will need to explore it rather than stick to squares A and B