2. Intelligent Agents

2.1 Agents and Environments

- Definition of Agent: **Agent** is anything that can be viewed as perceiving the **environment** through **sensors** and acting upon the environment through **actuators**.

Diagram showing how an agent interacts with the environment through sensors and actuators.

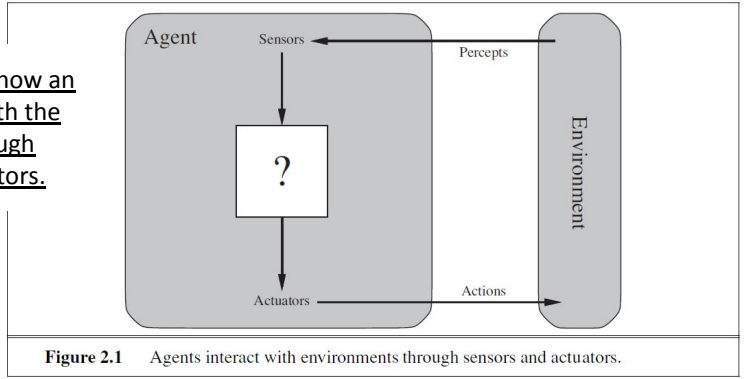


Figure 2.1 Agents interact with environments through sensors and actuators.

2.1 Agents and Environments

- Definition of Agent: **Agent** is anything that can be viewed as perceiving the **environment** through **sensors** and acting upon the environment through **actuators**.
- The human agent has
  - eyes, ears, and other organs for sensors
  - hands, legs, vocal tract for actuators

Diagram showing how an agent interacts with the environment through sensors and actuators.

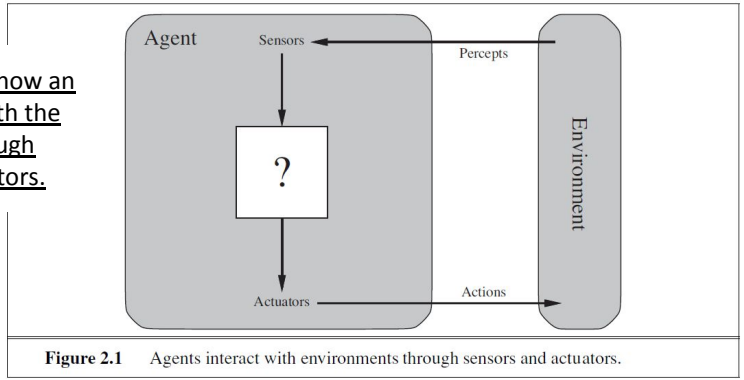


Figure 2.1 Agents interact with environments through sensors and actuators.

2.1 Agents and Environments

- Definition of Agent: **Agent** is anything that can be viewed as perceiving the **environment** through **sensors** and acting upon the environment through **actuators**.
- The human agent has
  - eyes, ears, and other organs for sensors
  - hands, legs, vocal tract for actuators
- A robotic agent has
  - cameras and infrared range finders for sensors
  - various motors for actuators

Diagram showing how an agent interacts with the environment through sensors and actuators.

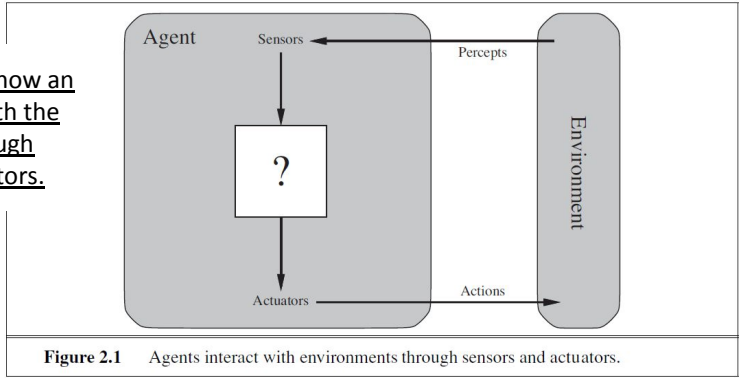


Figure 2.1 Agents interact with environments through sensors and actuators.

2.1 Agents and Environments

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

Diagram showing how an agent interacts with the environment through sensors and actuators.

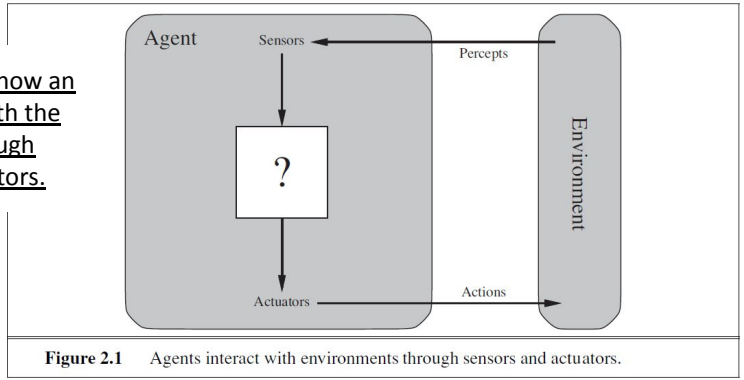


Figure 2.1 Agents interact with environments through sensors and actuators.

2.1 Agents and Environments

- Definition of Agent: **Agent** is anything that can be viewed as perceiving the **environment** through **sensors** and acting upon the environment through **actuators**.
- The human agent has
  - eyes, ears, and other organs for sensors
  - hands, legs, vocal tract for actuators
- A robotic agent has
  - cameras and infrared range finders for sensors
  - various motors for actuators
- A software agent
  - receives keystrokes, file contents, and network packets as sensory inputs
  - acts on the environment by displaying on the screen, writing files, and sending network packets
- Percept vs Percept Sequence: **Percept** is the agent's perceptual input at any given instant, and **percept sequence** is the complete history of everything the agent has ever perceived.

Diagram showing how an agent interacts with the environment through sensors and actuators.

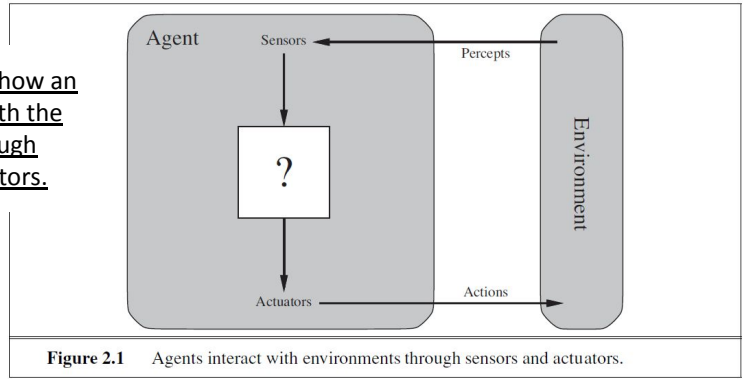
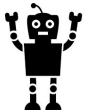


Figure 2.1 Agents interact with environments through sensors and actuators.

2.1 Agent Function vs Agent Program

What does this agent do?



A new agent!

- **Agent function** - describes agent's behavior by mapping any given percept sequence to an action
  - To describe any given agent, we have to tabulate the agent function - and this will typically be a very large table (potentially infinitely large table)
  - We can, in principle, construct this table by trying out all possible percept sequences and recording which actions the agent does in response
  - This table is external characterization of the agent
  - Agent function is abstract mathematical description
- **Agent program** - is an internal implementation of the agent function for an artificial agent
  - It is a concrete implementation, running within some physical system

2.1 Agent Function vs Agent Program (An Example)

- The vacuum cleaner world
  - has just two locations: squares A and B.
- The vacuum agent
  - perceives which square it is in and whether there is dirt in the square
  - can choose to move left, move right, such up the dirt or do nothing
- Agent function
  - If the current square is dirty, then suck; otherwise move to the other square

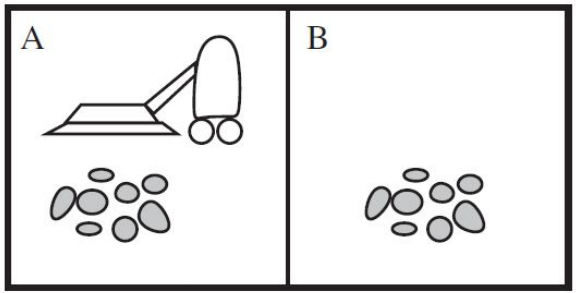


Figure 2.2 A vacuum-cleaner world with just two locations.

2.1 Agent Function vs Agent Program

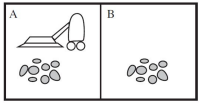


Figure 2.2 A vacuum-cleaner world with just two locations.

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*
⋮	⋮

Figure 2.3 Partial tabulation of a simple agent function for the vacuum-cleaner world shown in Figure 2.2.

Agent Function

function REFLEX-VACUUM-AGENT([*location*,*status*]) **returns** an action

if *status* = *Dirty* **then return** ?
else if *location* = *A* **then return** *Right*
else if *location* = *B* **then return** ?

Figure 2.8 The agent program for a simple reflex agent in the two-state vacuum environment. This program implements the agent function tabulated in Figure 2.3.

Agent Program

2.2 Good Behavior: The Concept of Rationality

We learnt that we should design agents that “act rational”

- How do we define ‘acting rationally’ so we can write programs?
- Should we consider the environment where the agent will be deployed?

2.2 Good Behavior: The Concept of Rationality

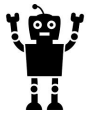
- A rational agent is one that does the right thing (not a definition of rational agent)
  - In other words, every entry in the table for the agent function is filled out correctly

2.2 Good Behavior: The Concept of Rationality

- A rational agent is one that does the right thing (not a definition of rational agent)
  - In other words, every entry in the table for the agent function is filled out correctly
- Doing the right thing is better than doing the wrong thing, but what does it mean to do the right thing?
  - We consider the consequences of the agent's behavior
  - When an agent is plunked down in an environment, it generates a sequence of actions according to the percepts it receives
  - The sequence of actions causes the environment to go through a sequence of states
  - If this sequence of environment is desirable, the agent has performed well

2.2 Good Behavior: The Concept of Rationality

- A rational agent is one that does the right thing (not a definition of rational agent)
  - In other words, every entry in the table for the agent function is filled out correctly
- What does it mean to do the right thing?
  - We consider the consequences of the agent's behavior
  - When an agent is plunked down in an environment, it generates a sequence of actions according to the percepts it receives
  - The sequence of actions causes the environment to go through a sequence of states
  - If this sequence of environment is desirable, the agent has performed well



A new agent!

2.2 Good Behavior: The Concept of Rationality

- We are interested in **environment states** not **agent states**
  - We should not define success in terms of the agent's opinion
  - The agent could achieve a perfect rationality simply by deluding itself that its performance was perfect
  - Human agents, for example, are notorious for "sour grapes" - believing that they did not really want something after not getting it

2.2 Design Performance Measure with Environment Focus

- 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.

- For example, for the vacuum agent, we propose to measure the performance by the **amount of dirt cleaned up in a single eight-hour shift**.

  - What is wrong with this performance measure? 

2.2.1 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.

2.2.1 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.

Definition of a **rational agent**:

- 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**.

Neither save humanity nor harm!

2.2.1 Rationality

- Let's assume the following for the vacuum-cleaner example:
  - Performance measure - award one point for each clean square at each time step, over a lifetime of 1000 time steps
  - Prior knowledge - the agent knows that there are two squares but it does not know the dirt distribution
  - Agent actions - available actions are left, right and suck
  - Percept sequence - agent correctly perceives its location and whether that location contains dirt
- Here is what the vacuum-cleaner agent does:
  - Cleans a square if it is dirty and moves to the other square if not.

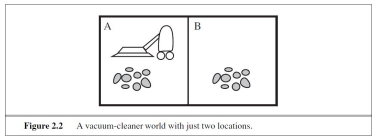


Figure 2.2 A vacuum-cleaner world with just two locations.

Is the vacuum cleaner agent rational (is performance measure robust)?

2.2.1 Rationality

- Let's assume the following for the vacuum-cleaner example:
  - Performance measure - award one point for each clean square at each time step, over a lifetime of 1000 time steps
  - Prior knowledge - the agent knows that there are two squares but it does not know the dirt distribution
  - Agent actions - available actions are left, right and suck
  - Percept sequence - agent correctly perceives its location and whether that location contains dirt
- Here is what the vacuum-cleaner agent does:
  - Cleans a square if it is dirty and moves to the other square if not.

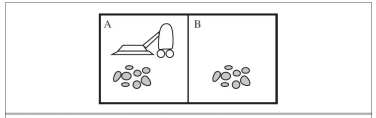


Figure 2.2 A vacuum-cleaner world with just two locations.

Is the vacuum cleaner agent rational (is performance measure robust)?

THINK PAIR SHARE

- What will the agent do after all dirt is cleaned up? Oscillate back and forth.

Definition of a rational agent - 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.

2.3.1 Specifying the Task Environment

- We would like to design and implement an agent
  - What is the first thing we need to do?

2.3.1 Specifying the Task Environment

- In designing an agent, the first step must always be to specify the task environment as fully as possible.
- Task environment is the description of **Performance, Environment, Actuators, and Sensors (PEAS)**

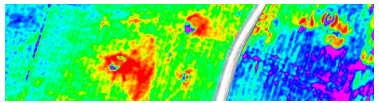
2.3.1 Specifying the Task Environment

- In designing an agent, the first step must always be to specify the task environment as fully as possible.
- Task environment is the description of **Performance, Environment, Actuators, and Sensors (PEAS)**
- A fully automated taxi currently is somewhat beyond the capabilities of existing technology because the full driving task is extremely open-ended.

Agent Type	Performance Measure	Environment	Actuators	Sensors
Taxi driver	Safe, fast, legal, comfortable trip, maximize profits works in all weather conditions	Roads, other traffic, pedestrians, customers	Steering, accelerator, brake, signal, horn, display	Cameras, sonar, speedometer, GPS, odometer, accelerometer, engine sensors, keyboard

Figure 2.4 PEAS description of the task environment for an automated taxi.

2.3.1 Specifying the Task Environment



Agent Type	Performance Measure	Environment	Actuators	Sensors
Medical diagnosis system	Healthy patient, reduced costs Not correct diagnosis!	Patient, hospital, staff	Display of questions, tests, diagnoses, treatments, referrals	Keyboard entry of symptoms, findings, patient's answers
Satellite image analysis system	Correct image categorization	Downlink from orbiting satellite	Display of scene categorization	Color pixel arrays
Part-picking robot	Percentage of parts in correct bins Not % of parts picked correctly!	Conveyor belt with parts; bins	Jointed arm and hand	Camera, joint angle sensors
Refinery controller	Purity, yield, safety	Refinery, operators	Valves, pumps, heaters, displays	Temperature, pressure, chemical sensors
Interactive English tutor	Student's score on test Not how well she/he teaches!	Set of students, testing agency	Display of exercises, suggestions, corrections	Keyboard entry

Figure 2.5 Examples of agent types and their PEAS descriptions.

Design a Rational Self-driving Vehicle

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.

Definition of a **rational agent**:

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**.

Define what your self-driving vehicle does:

What should be the performance measure?

What prior knowledge will it have?

What are the actions that the vehicle can do?

What will be the vehicle's percept sequence?

Summary

- An **agent** is something that perceives and acts in an environment.
- The **agent function** for an agent specifies the action taken by the agent in response to any percept sequence.
- The **performance measure** evaluates the behavior of the agent in an environment.
- A **rational agent** acts so as to maximize the expected value of the performance measure, given the percept sequence it has seen so far.
- A **task environment** specification includes the performance measure, the external environment, the actuators, and the sensors (PEAS).
- In designing an agent, the first step must always be to specify the task environment as fully as possible.