Module1

Source Book: Stuart J. Russell and Peter Norvig, Artificial Intelligence, 3rd Edition, Pearson,2015

Topics

- What is AI?
- Foundations of AI
- History of AI,
- Agents, The structure of agents
- Concept of Rationality,
- The nature of environment.

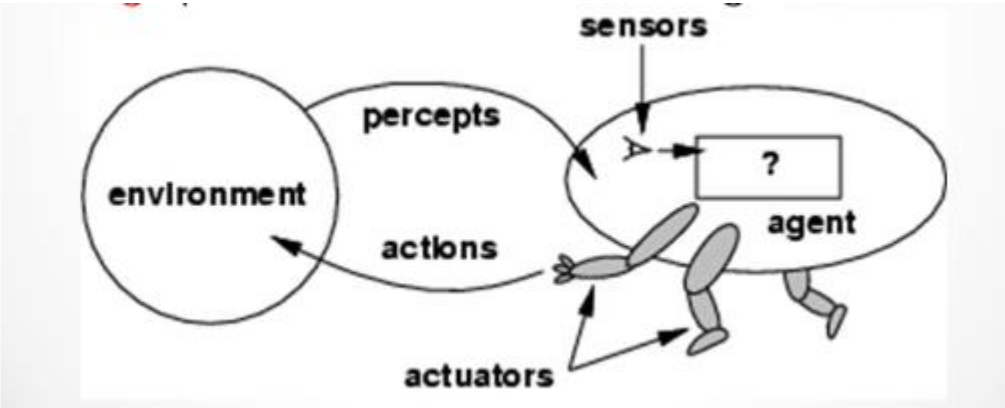
What is AI?

- Artificial Intelligence (AI) is a field of computer science dedicated to develop systems capable of performing tasks that would typically require human intelligence.
- These tasks include **learning, reasoning, problem-solving, perception, understanding natural language**, and even interacting with the environment.
- AI aims to create machines or software that can mimic or simulate human cognitive functions.

According to Russell and Norvig, AI can be defined as follows:

" *AI (Artificial Intelligence) is the study of agents that perceive their environment, reason about it, and take actions to achieve goals. Each such agent implements a function that maps percept sequences to actions, and the study of these functions is the subject of AI.*"

Agent

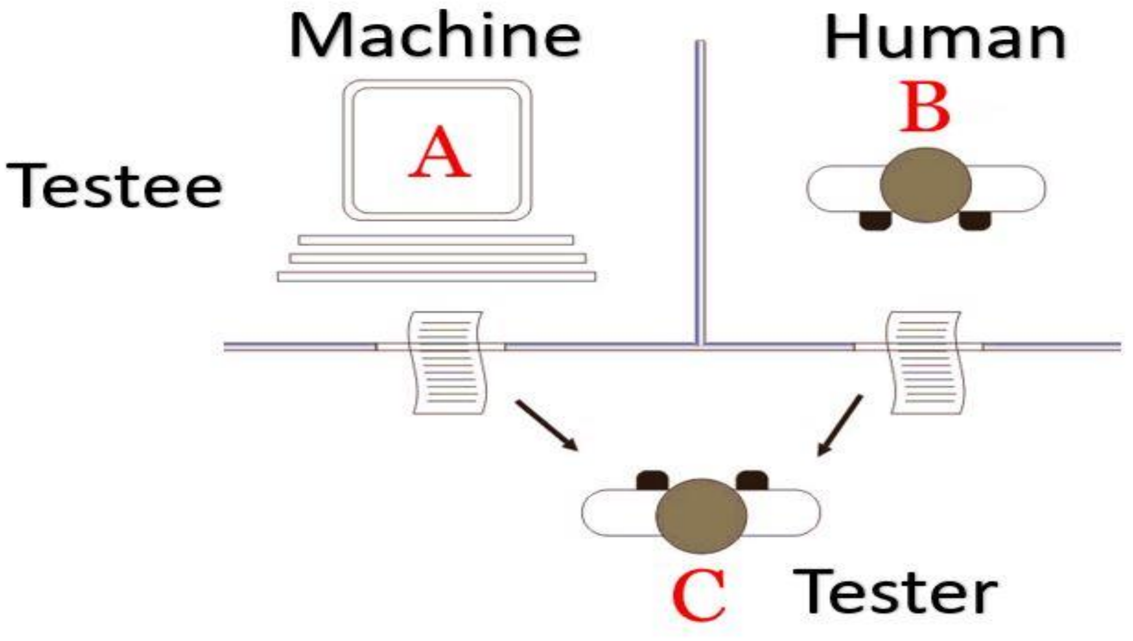


In AI, an agent is any entity that can perceive its environment and take actions to achieve its goals. These agents can be physical robots, software programs, or any system capable of interacting with the world. The agents sense the environment through sensors and act on their environment through actuators. An AI agent can have mental properties such as knowledge, belief, intention, etc.

Eight explanations of **AI shown in two groups**

Thinking Humanly	**Thinking Rationally**
"The exciting new effort to make computers think ... *machines with minds*, in the full and literal sense." (Haugeland, 1985) "[The automation of] activities that we associate with human thinking, activities such as decision-making, problem solving, learning ..." (Bellman, 1978)	"The study of mental faculties through the use of computational models." (Charniak and McDermott, 1985) "The study of the computations that make it possible to perceive, reason, and act." (Winston, 1992)
Acting Humanly	**Acting Rationally**
"The art of creating machines that perform functions that require intelligence when performed by people." (Kurzweil, 1990) "The study of how to make computers do things at which, at the moment, people are better." (Rich and Knight, 1991)	"Computational Intelligence is the study of the design of intelligent agents." (Poole *et al.*, 1998) "AI ... is concerned with intelligent behavior in artifacts." (Nilsson, 1998)

Turing Test



The Foundations of Artificial Intelligence

Philosophy: In the philosophical exploration of AI, following questions arises:

- Can formal rules be used to draw valid conclusions?
- How does the mind arise from a physical brain?
- Where does knowledge come from?
- How does knowledge lead to action

Mathematics: In the mathematical exploration of AI, following questions arises:

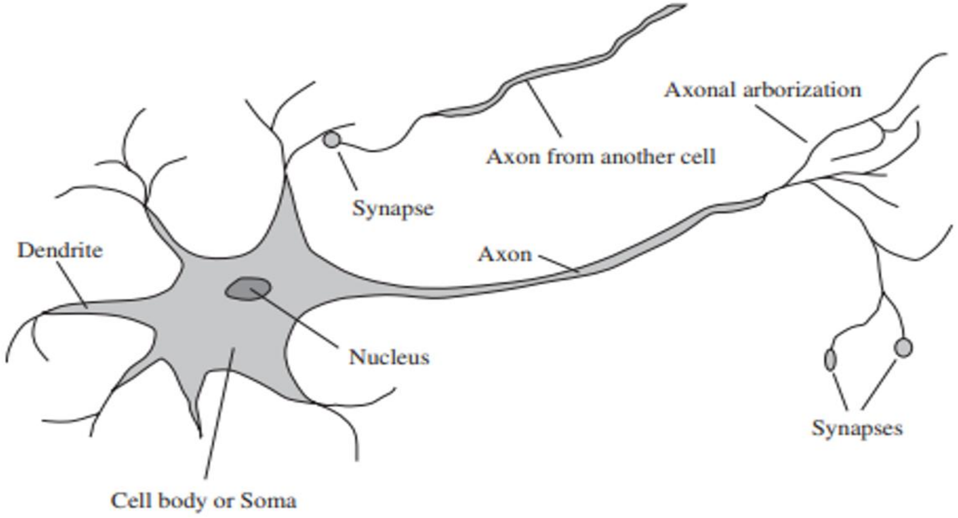
- What are the formal rules to draw valid conclusions?
- What can be computed?
- How do we reason with uncertain information?

Economics: In the Economics of AI, various attempts have been done to address the following questions:

- How should we make decisions so as to maximize payoff?
- How should we do this when others may not go along?
- How should we do this when the payoff may be far in the future

The Foundations of Artificial Intelligence

Neuroscience (How do brains process information?)



	Supercomputer	Personal Computer	Human Brain
Computational units	10^4 CPUs, 10^{12} transistors	4 CPUs, 10^9 transistors	10^{11} neurons
Storage units	10^{14} bits RAM	10^{11} bits RAM	10^{11} neurons
	10^{15} bits disk	10^{13} bits disk	10^{14} synapses
Cycle time	10^{-9} sec	10^{-9} sec	10^{-3} sec
Operations/sec	10^{15}	10^{10}	10^{17}
Memory updates/sec	10^{14}	10^{10}	10^{14}

Psychology: How do humans and animals think and act?

Computer Engineering (How can we build an efficient computer?)

For artificial intelligence (AI) to succeed, two key elements are essential: **intelligence and an artifact**, with the **computer** being the chosen artifact.

Control Theory and Cybernetics (How can artifacts operate under their own control?)

Linguistics (How does language relate to thought?)

History of AI

1.3.1 The Gestation of Artificial Intelligence (1943–1955)

1943: Warren McCulloch and Walter Pitts develop the first mathematical model of a neural network.

1949: Hebbian learning, formulated by Donald Hebb, becomes a lasting influence on neural network development.

1950: Alan Turing introduces the Turing Test and key AI principles.

1951: UNIVAC I, the first commercially produced computer, is used for statistical analysis, laying the groundwork for data processing.

1951: McCarthy earns his PhD and later plays a pivotal role in establishing AI at Dartmouth College.

1.3.2 The Birth of Artificial Intelligence (1956)

1956: John McCarthy organizes a landmark AI workshop at Dartmouth College, marking the official birth of artificial intelligence.

1956: The term "artificial intelligence" is coined at the Dartmouth Conference.

History of AI

1.3.3 Early Enthusiasm, Great Expectations (1952–1969)

Late 1950s: IBM produces early AI programs challenging predefined task limitations.

1958: McCarthy defines the Lisp language and introduces time-sharing.

1963: Marvin Minsky establishes Stanford's AI lab, emphasizing practical functionality.

1965: Joseph Weizenbaum creates ELIZA, an early natural language processing program.

1966–1973: Setbacks occur as early successes fail to scale, leading to reduced support for AI research.

1969: The Stanford Research Institute develops Shakey, the first mobile robot with reasoning abilities.

1.3.4 A Dose of Reality (1966–1973)

1969: Despite setbacks, the discovery of back-propagation learning algorithms for neural networks leads to a resurgence of interest.

1970s: Initial enthusiasm for AI fades, leading to the first "AI winter" as progress stalls

History of AI

1.3.5 Knowledge-Based Systems: The Key to Power? (1969–1979)

1969: DENDRAL exemplifies a shift towards domain-specific knowledge.

Late 1970s: The Heuristic Programming Project explores expert systems, emphasizing domain-specific knowledge.

1979: Marvin Minsky and Seymour Papert publish "Perceptrons," a book critical of certain AI approaches.

1.3.6 AI Becomes an Industry (1980–Present)

Early 1980s: The first successful commercial expert system, R1, is implemented at Digital Equipment Corporation.

1981: Japan's "Fifth Generation" project and the U.S.'s Microelectronics and Computer Technology Corporation respond to AI's growing influence.

1980s: Rapid growth followed by the "AI Winter," a period of decline in the AI industry.

1985: Expert systems, software that emulates decision-making of a human expert, gain popularity.

History of AI

1.3.7 The Return of Neural Networks (1986–Present)

Mid-1980s: Rediscovery of the back-propagation learning algorithm leads to the emergence of connectionist models.

Late 1980s: The AI industry experiences a decline known as the AI Winter.

1.3.8 AI Adopts the Scientific Method (1987–Present)

Late 1980s: AI shifts towards a more scientific and application-focused approach, experiencing a revival in the late 1990s.

1990-2005: Neural Networks Resurgence and Practical Applications

1997: IBM's Deep Blue defeats chess champion Garry Kasparov.

1999: Rodney Brooks introduces the concept of "embodied intelligence" with Cog, a humanoid robot.

History of AI

1.3.9 The Emergence of Intelligent Agents (1995–Present)

Late 1980s: The SOAR architecture addresses the "whole agent" problem.

Late 1990s–2000s: AI technologies underlie Internet tools, contributing to search engines, recommender systems, and website aggregators.

1.3.10 The Availability of Very Large Data Sets (2001–Present)

Late 1990s: A revival in AI with a shift towards a more scientific approach.

2000s: Emphasis on the importance of large datasets in AI research, leading to significant advancements.

2001: The DARPA Grand Challenge initiates research in autonomous vehicles.

2005: Stanford's Stanley wins the DARPA Grand Challenge, showcasing advances in self-driving technology.

2006: Geoffrey Hinton and colleagues publish a paper on deep learning, reigniting interest in neural networks.

2011: IBM's Watson wins Jeopardy!, demonstrating the power of natural language processing.

2012: AlexNet, a deep convolutional neural network, achieves a breakthrough in image recognition at the ImageNet competition.

2012-Present: AI in the Mainstream and Ethical Concerns

2014: Facebook's AI lab introduces DeepFace for facial recognition, reaching human-level accuracy.

2016: AlphaGo, an AI developed by DeepMind, defeats world champion Go player Lee Sedol.

2018: OpenAI releases GPT-2, a large-scale language model.

2020s: AI applications become integral in various industries, raising concerns about ethics, bias, and job displacement.

Agent

An agent is defined as anything capable of perceiving its environment through sensors and acting upon that environment through actuators. This basic concept is depicted in Figure 2.1.

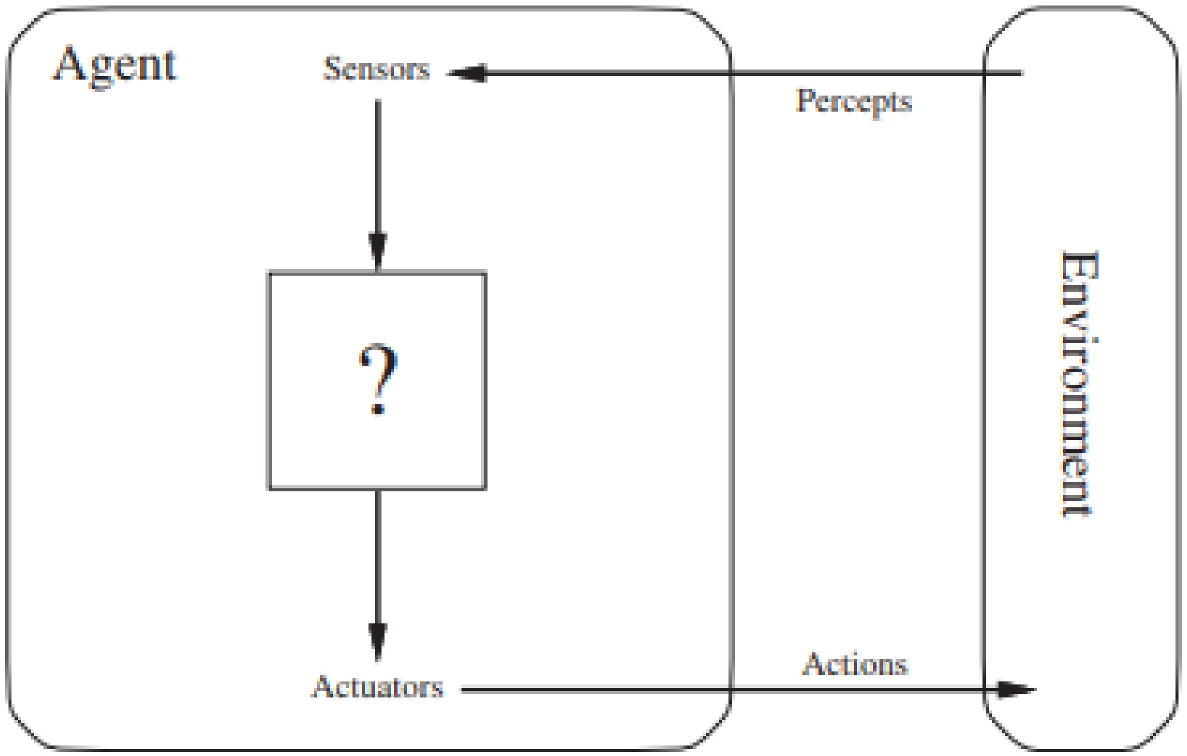


Figure 2.1 Agents interact with environments through sensors and actuators.

For instance, a **human agent** employs eyes, ears, and other sensory organs while utilizing hands, legs, vocal tract, etc., as actuators.

In contrast, a **robotic agent** might use cameras and infrared range finders as sensors and various motors as actuators.

A **software agent** takes in sensory inputs such as keystrokes, file contents, and network packets and responds by displaying on the screen, writing files, or sending network packets.

Percept: The term **"percept"** refers to an agent's perceptual inputs at any given moment, and a percept sequence encompasses the complete history of everything the agent has perceived

Table: A table serves as an external representation of an agent's behaviour, specifically presented in tabular form that outlines the agent's actions for every possible percept sequences.

Agent Program: An agent program represents the internal implementation of an agent's behavior. It is a concrete realization of the abstract agent function, running within a physical system. The agent program executes the specified actions determined by the agent function in response to the sensory inputs received by the agent.

Vacuum Cleaner World

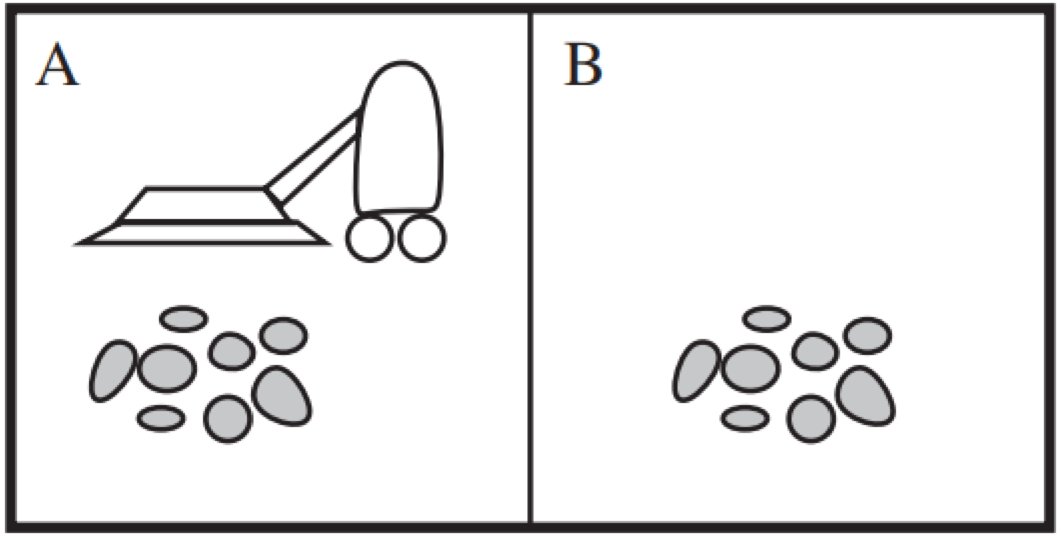


Table [Agent function]

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

Agent Program

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

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

Characteristics of Intelligent agents

In artificial intelligence (AI), intelligent agents are entities that perceive their environment and take actions to achieve **specific goals**. These agents are designed to operate autonomously, making decisions based on their perception of the world and their programming or learning mechanisms.

Characteristics

Perception: Agents have the ability to perceive or sense their environment. This involves gathering information from the surroundings through sensors or other means.

Action: Intelligent agents can take actions in response to their perceptions. These actions are chosen to influence the state of the environment or achieve specific goals.

Autonomy: Agents operate autonomously, making decisions and taking actions without direct human intervention. Autonomy allows them to adapt to changing conditions in their environment.

Goal-Directed Behaviour: Intelligent agents are typically designed to achieve specific goals. These goals can be predefined by a programmer or learned by the agent through experience.

Learning and Adaptation: Many intelligent agents have the capability to learn from experience and adapt their behaviour over time. This learning process may involve acquiring new knowledge or improving decision-making strategies.

Concept of Rationality

A rational agent is defined as one that consistently does right thing and makes decisions leading to favourable outcomes. This is reflected in the **correct completion of every entry in the agent function table**. However, determining what constitutes the "**right thing**" involves the following:

- **Assessing Consequences**
- **Performance Measure**:
  - **Challenges and Considerations in Designing Performance Measures**
  - **Alignment with Environmental Goals**
- **Philosophical Considerations**

What is Rationality?

Rationality encompasses the state of being **reasonable, sensible, and possessing a sound judgment**. It pertains to anticipated actions and outcomes based on the agent's perceptions. Engaging in actions with the objective of acquiring valuable information is a crucial aspect of rationality.

Rationality at Any Given Time depends on the following four things:

1. **Performance Measure**: The criterion defining success.
2. **Agent’s Prior Knowledge**: Understanding of the environment.
3. **Available Actions**: The actions the agent can perform.
4. **Percept Sequence**: The agent’s historical sensory input.

Omniscience, Learning, and Autonomy in Rational Agents

Omniscience and Rationality: Distinguishing between rationality and omniscience is crucial. An omniscient agent knows the actual outcome of its actions and can act accordingly; but omniscience is impossible in reality. While rationality maximizes expected performance.

Information Gathering : Rational agents should perform actions to modify future percepts, known as information gathering. This concept is vital for maximizing expected performance.

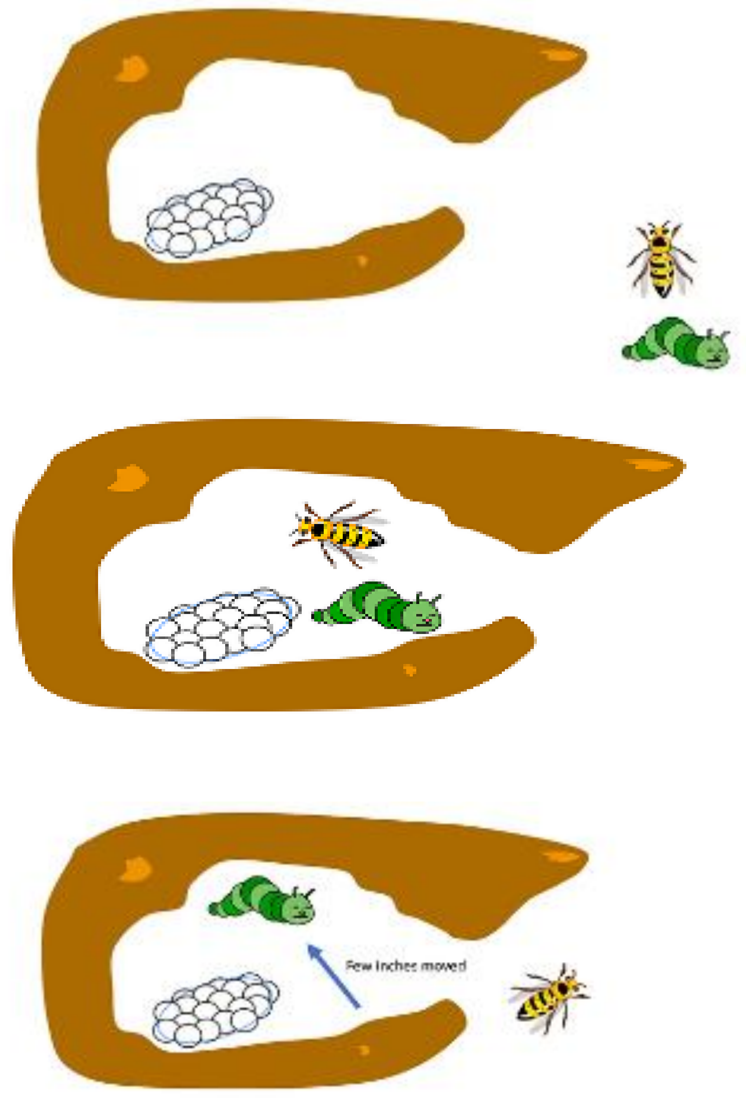
Learning in Rational Agents: Rational agents are not only expected to gather information but also to learn from their perceptions. An agent's initial configuration may reflect prior knowledge, but as it gains experience, learning becomes imperative.

Dung Beetle

*Think about the **dung beetle**, a small insect that carefully constructs its nest, lays eggs, and uses a ball of dung to seal the entrance. **If the dung ball is taken away** while the beetle is carrying it, the beetle continues its job as if the dung ball is still there, not realizing its missing.*

Sphex Wasp

*The **sphex wasp** is a bit smarter. It digs a hole, stings a caterpillar, drags it into the hole, checks everything is fine, and then lays eggs. The caterpillar becomes food for the hatching eggs. However, if someone moves the caterpillar a little while the wasp is checking, it goes back to dragging the caterpillar, not realizing the change. Even after many attempts to move the caterpillar, the wasp doesn't learn that its plan isn't working and keeps doing it the same way.*



Autonomy in Rational Agents

Autonomy is emphasized as a crucial aspect of rationality. An autonomous agent learns to compensate for **partial or incorrect prior knowledge**, ensuring effective adaptation to the environment.

In essence, the combination of **rationality, information gathering, and learning enables the design of autonomous agents** capable of success across diverse environments.

Nature of Environments

- The task environment, refers to the **external system or surroundings** within which an **agent** operates and performs tasks.
- **Task environment is specified in terms of PEAS (Performance, Environment, Actuators, Sensors**).

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

Agent Type	Performance Measure	Environment	Actuators	Sensors
Medical diagnosis system	Healthy patient, reduced costs	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	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	Set of students, testing agency	Display of exercises, suggestions, corrections	Keyboard entry

Properties of Task Environments

1.Fully Observable vs. Partially Observable:

Fully Observable: The agent can see the complete environment all the time.

Partially Observable: The agent's sensors may not capture all relevant aspects due to noise or missing information.

2. Single Agent vs. Multiagent:

Single Agent: An agent operates independently, like solving a crossword puzzle.

Multiagent: Agents interact with each other, introducing complexities in decision-making.

3. Deterministic vs. Stochastic:

Deterministic: The environment's next state is entirely determined by the current state and the agent's action.

Stochastic: Some uncertainty exists, making predictions challenging. For example, traffic behavior in taxi driving is stochastic.

4. **Episodic vs. Sequential:**

Episodic: Each decision is independent, not influenced by past decisions.

Sequential: Current decisions affect future ones, as seen in chess or taxi driving.

5. **Static vs. Dynamic**:

Static: The environment remains unchanged during decision-making.

Dynamic: The environment evolves continuously, demanding constant

6. **Discrete vs. Continuous**:

Discrete: Finite states, actions, or percepts.

Continuous: Values or actions vary smoothly.

7. **Known vs. Unknown:**

Known: The agent knows outcomes or probabilities.

Unknown: The agent needs to learn about the environment.

Task Environment	Observable	Agents	Deterministic	Episodic	Static	Discrete
Crossword puzzle	Fully	Single	Deterministic	Sequential	Static	Discrete
Chess with a clock	Fully	Multi	Deterministic	Sequential	Semi	Discrete
Poker	Partially	Multi	Stochastic	Sequential	Static	Discrete
Backgammon	Fully	Multi	Stochastic	Sequential	Static	Discrete
Taxi driving	Partially	Multi	Stochastic	Sequential	Dynamic	Continuous
Medical diagnosis	Partially	Single	Stochastic	Sequential	Dynamic	Continuous
Image analysis	Fully	Single	Deterministic	Episodic	Semi	Continuous
Part-picking robot	Partially	Single	Stochastic	Episodic	Dynamic	Continuous
Refinery controller	Partially	Single	Stochastic	Sequential	Dynamic	Continuous
Interactive English tutor	Partially	Multi	Stochastic	Sequential	Dynamic	Discrete

Structure of Intelligent Agents

Agent's structure can be viewed as –

- **Agent = Architecture + Agent Program**
- **Architecture = the machinery that an agent executes on.**
- **Agent Program = an implementation of an agent function.** **All agent programs share a common structure**: they take the current percept as input from the **sensors** and produce an action for the **actuators**.

Agent program

- The **agent program**, operates based on the current percept,
- The **agent programs** are presented in a simple pseudocode language

```
function TABLE-DRIVEN-AGENT(percept) returns an action
   persistent: percepts, a sequence, initially empty
               table, a table of actions, indexed by percept sequences, initially fully specified

   append percept to the end of percepts
   action ← LOOKUP(percepts, table)
   return action
```

Agent Function

- The **agent function**, considers the entire percept history.
- Agent Function is represented using Table.

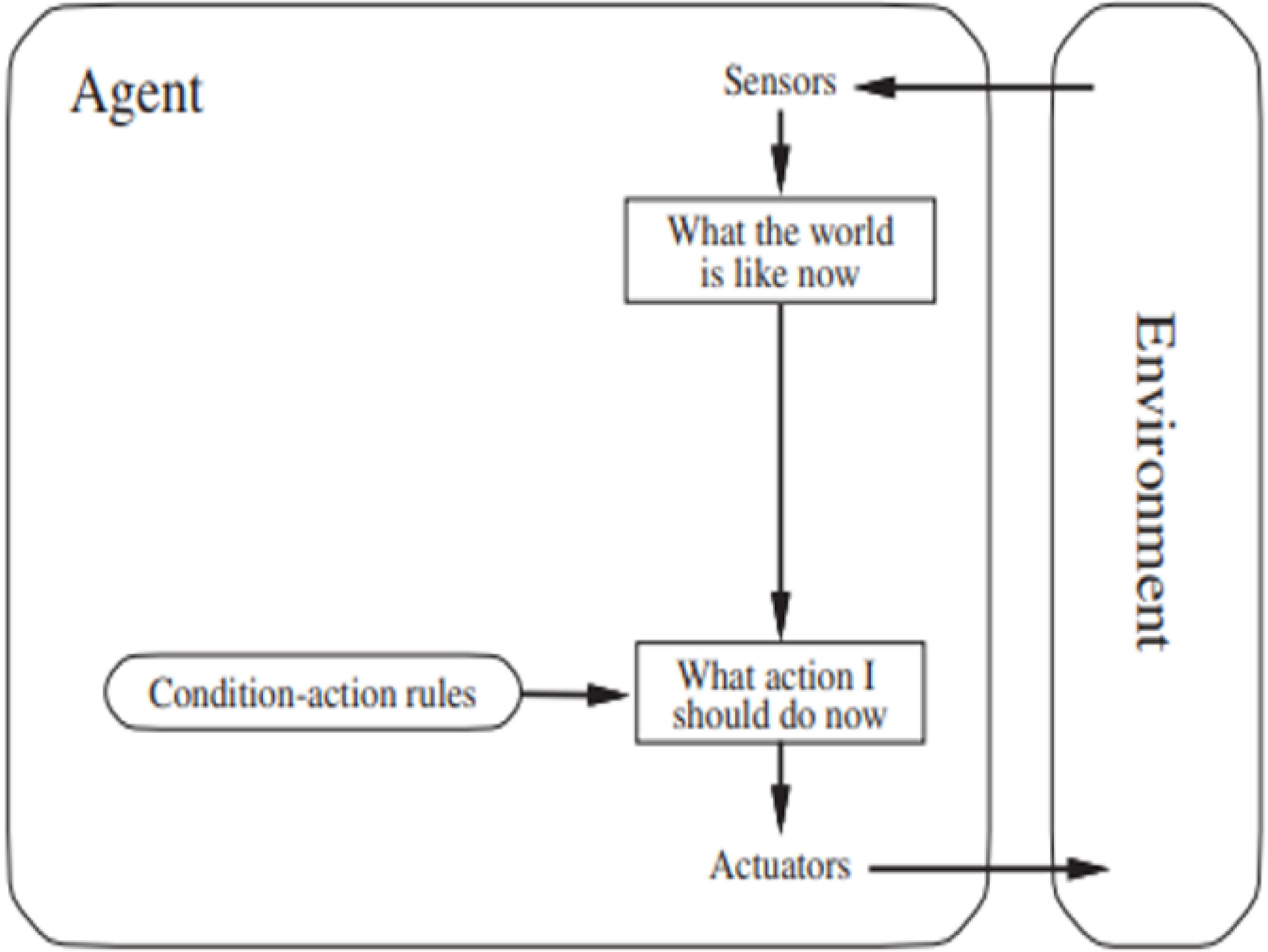
Percept sequence	Action
[*A*, *Clean*]	*Right*
[*A*, *Dirty*]	*Suck*
[*B*, *Clean*]	*Left*
[*B*, *Dirty*]	*Suck*
[*A*, *Clean*], [*A*, *Clean*]	*Right*
[*A*, *Clean*], [*A*, *Dirty*]	*Suck*
⋮	⋮
[*A*, *Clean*], [*A*, *Clean*], [*A*, *Clean*]	*Right*
[*A*, *Clean*], [*A*, *Clean*], [*A*, *Dirty*]	*Suck*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

Types of Agent Programs

There are, five basic types of agent programs as given below:

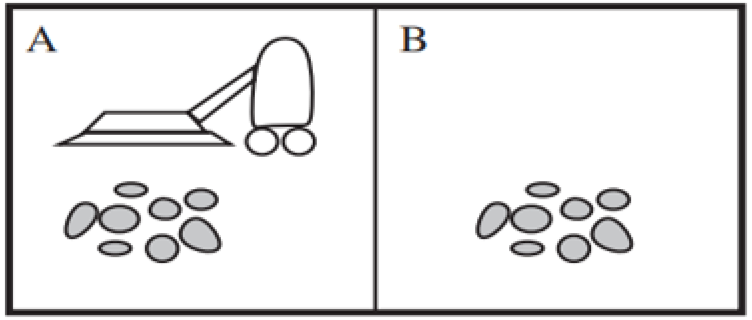
1. Simple reflex agents
2. Model-based reflex agents
3. Goal-based agents
4. Utility-based agents
5. Learning Agents

1.Simple reflex agents



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

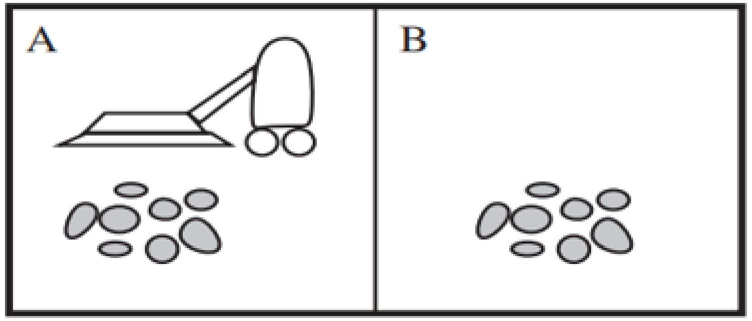
if *status* = *Dirty* **then return** *Suck*
else if *location* = *A* **then return** *Right*
else if *location* = *B* **then return** *Left*



Agent Program Example 1

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

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



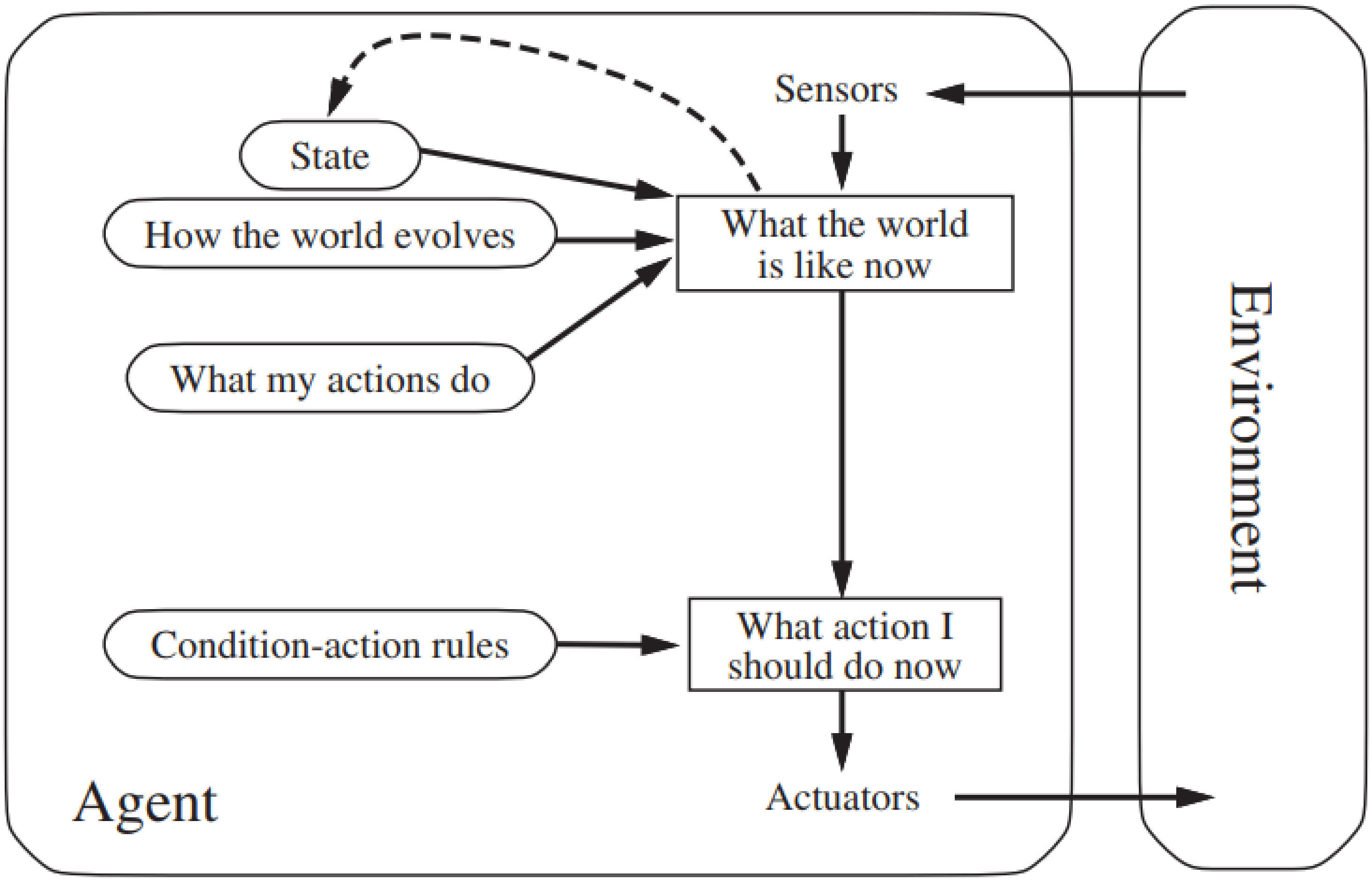
Agent Program Example 2

function SIMPLE-REFLEX-AGENT(*percept*) **returns** an action
  persistent: *rules*, a set of condition–action rules

  state ← INTERPRET-INPUT(*percept*)
  rule ← RULE-MATCH(*state*, *rules*)
  action ← *rule*.ACTION
  return *action*

if car-in-front-is-braking then initiate-braking

2. Model-based reflex agents

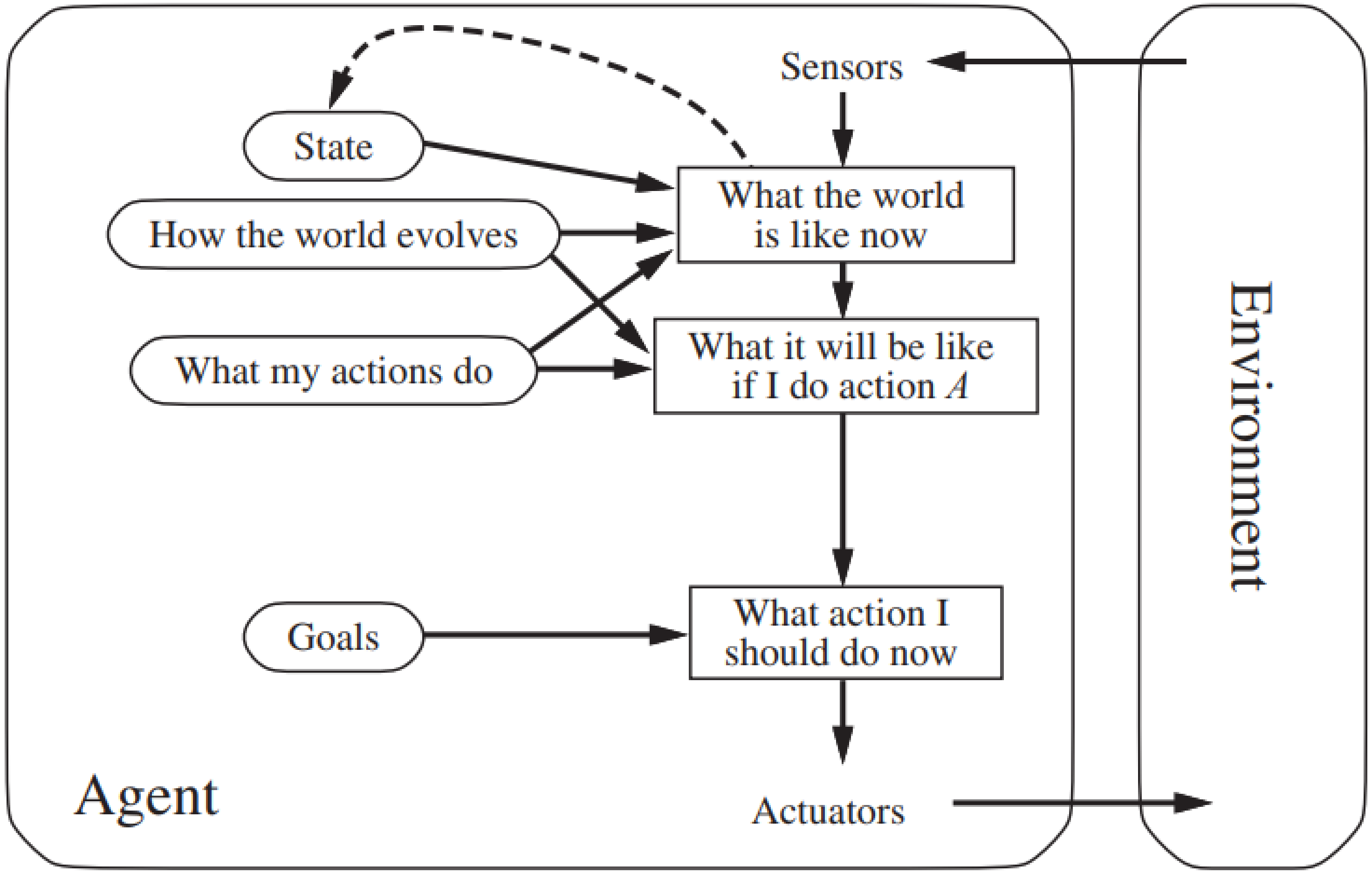


Agent Program

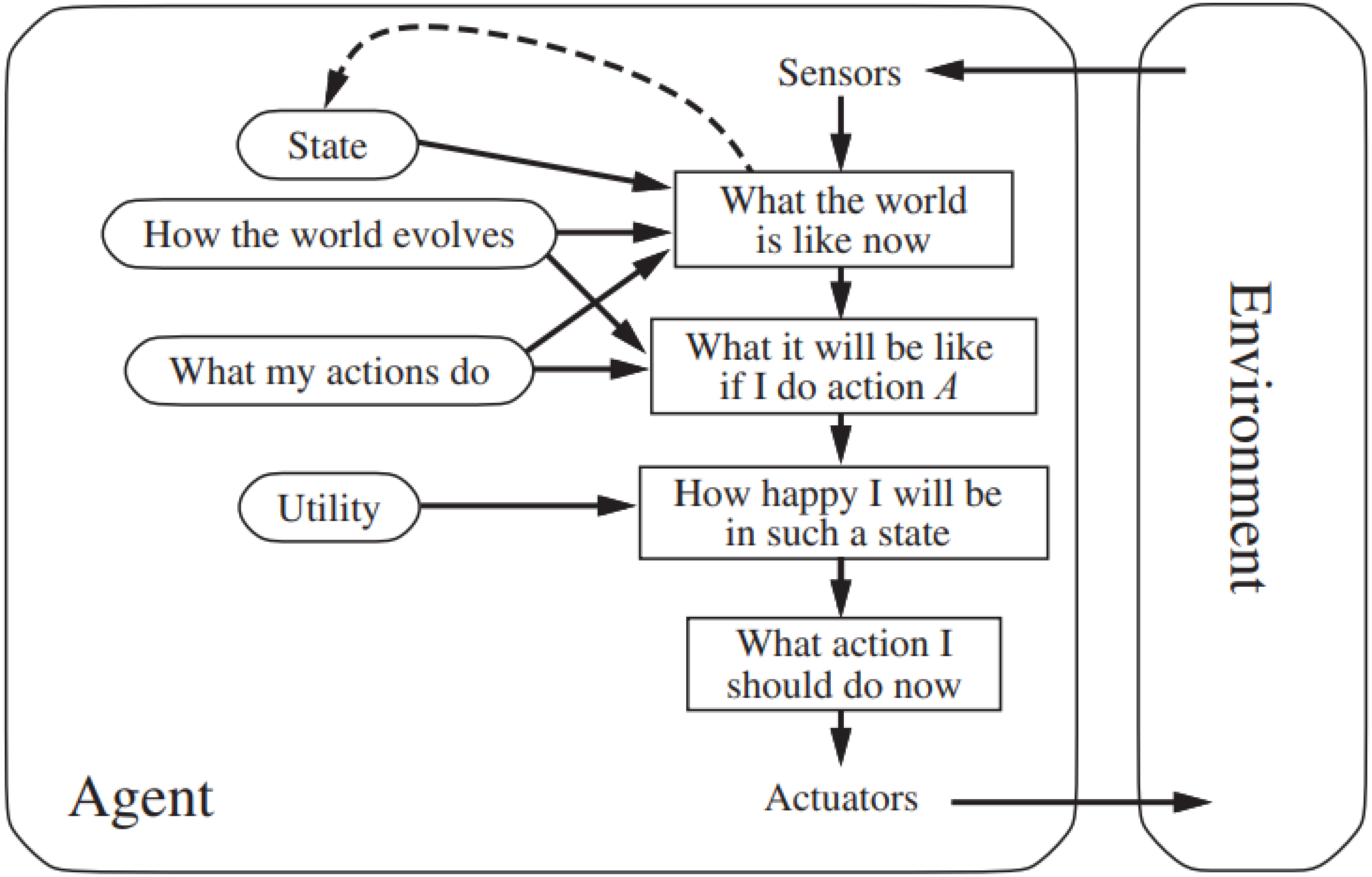
```
function MODEL-BASED-REFLEX-AGENT(percept) returns an action
  persistent: state, the agent's current conception of the world state
              model, a description of how the next state depends on current state and action
              rules, a set of condition–action rules
              action, the most recent action, initially none

  state ← UPDATE-STATE(state, action, percept, model)
  rule ← RULE-MATCH(state, rules)
  action ← rule.ACTION
  return action
```

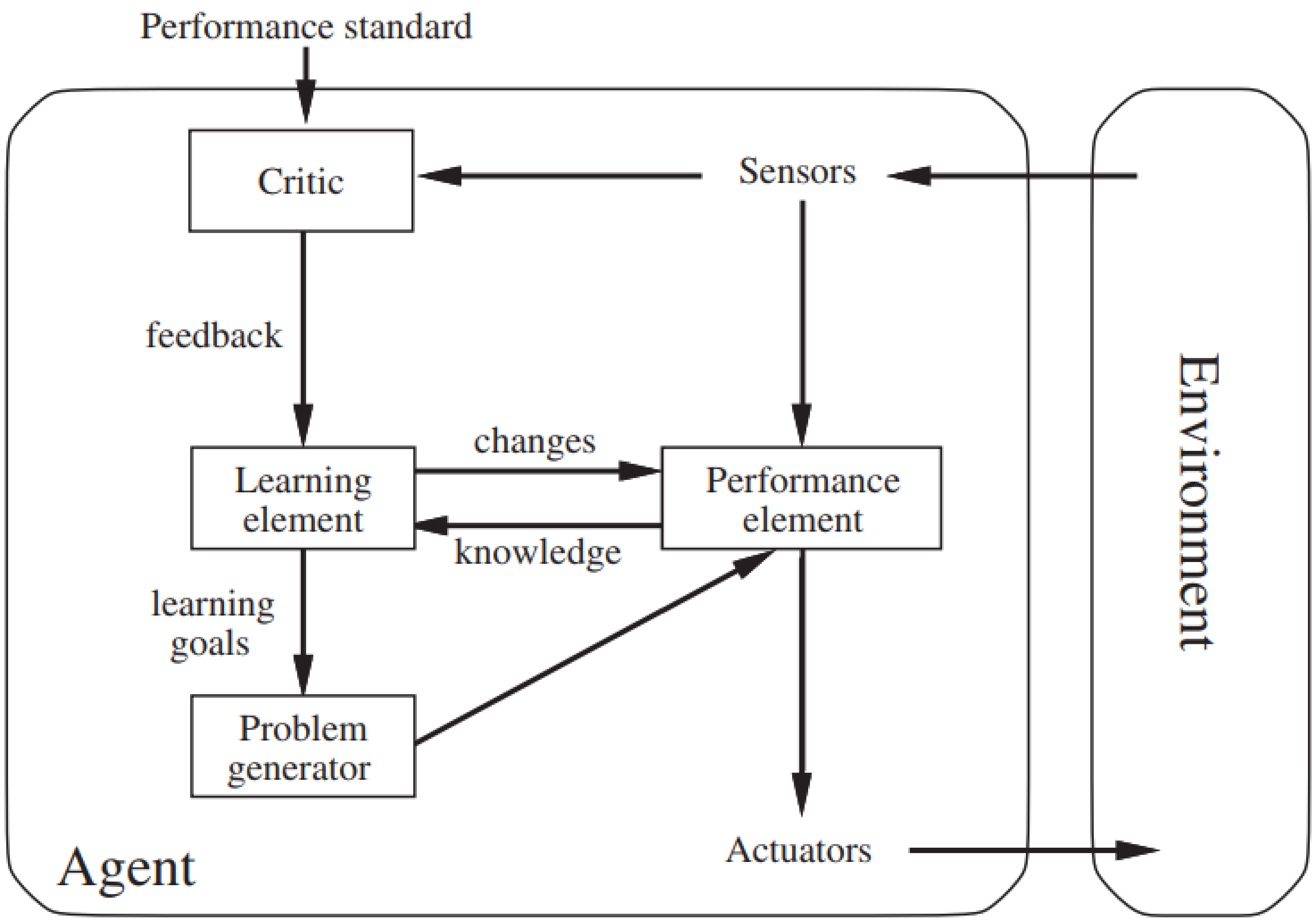
3. Goal-based agents



4. Utility based Agent

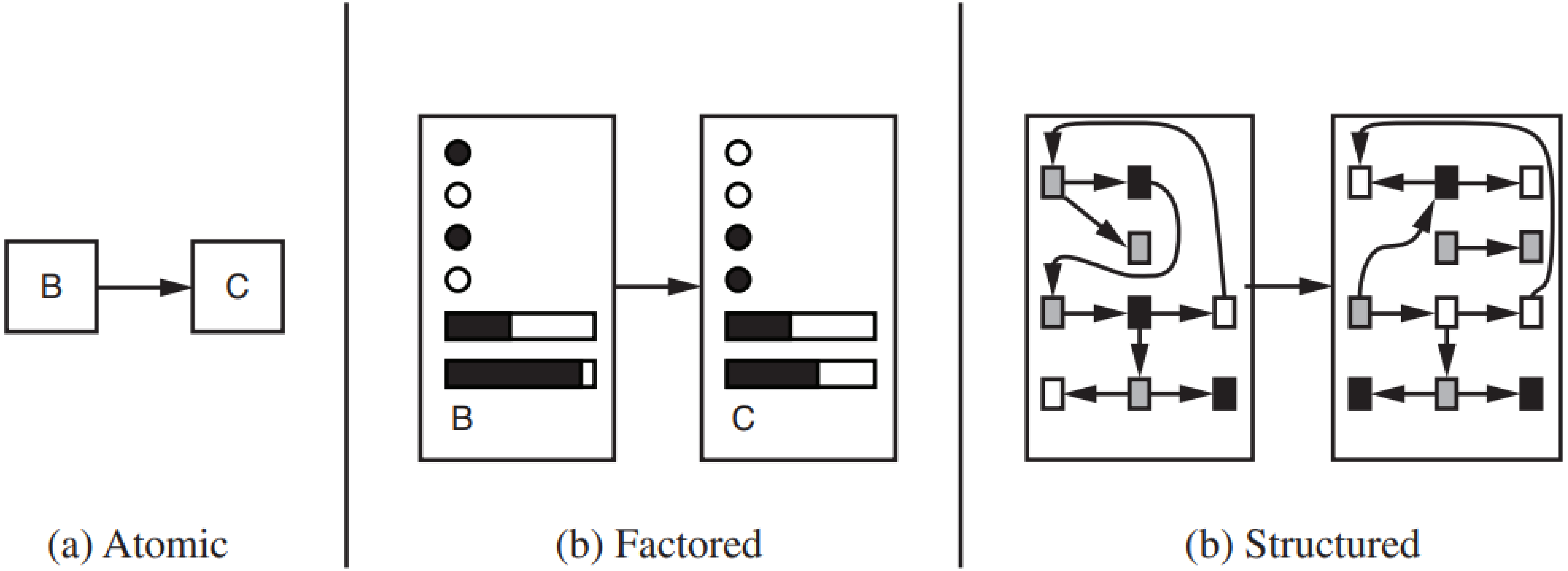


5. Learning agents



How the components of agent programs work ?

The representations of components of agents can be categorized along an axis of increasing complexity and expressive power: **atomic, factored, and structured**.



Topics

- What is AI?
- Foundations of AI
- History of AI,
- Agents, The structure of agents
- Concept of Rationality,
- The nature of environment.

End of Module1