# Artificial Intelligence

## Intelligent Agent

(Ch. 1, 2)

## What is Artificial Intelligence?

- The science and engineering of making intelligent machines [John McCarthy].
- The scientific understanding of the mechanisms underlying thought and intelligent behavior and their embodiment in machines [AAAI].
- The study and design of intelligent agents, where an intelligent agent is a system that perceives its environment and takes actions that maximize its chances of success [Russell & Norvig].

## What is AI?

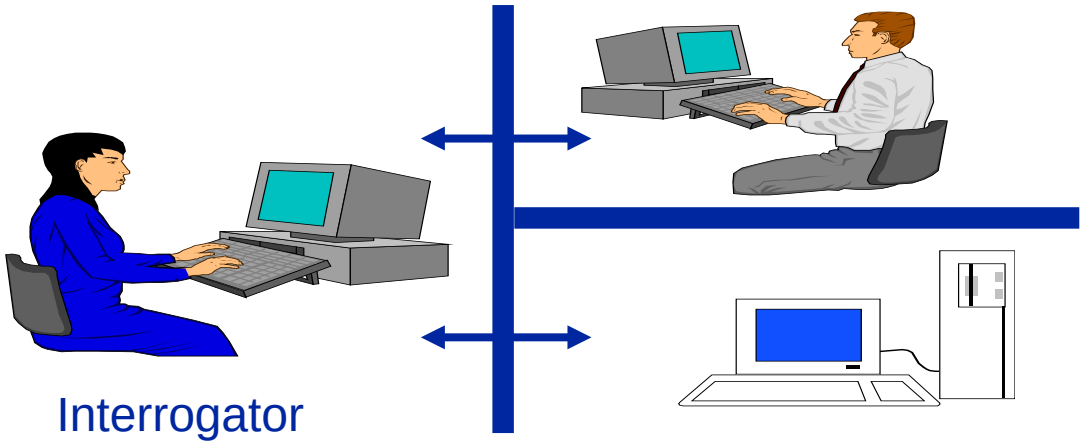
- **Views of AI fall into four categories:**

| | |
|---|---|
| **Thinking humanly** | **Thinking rationally** |
| **Acting humanly** | **Acting rationally** |

- **The textbook advocates "acting rationally"**

# Acting humanly: Turing Test

- **Turing (1950): *Computing machinery and intelligence***




- **Suggested major components of AI:**
  - **Knowledge representation**
  - **Reasoning**
  - **Natural language processing**
  - **Learning**

## Thinking humanly: cognitive modeling

- 1960s "cognitive revolution": information-processing psychology

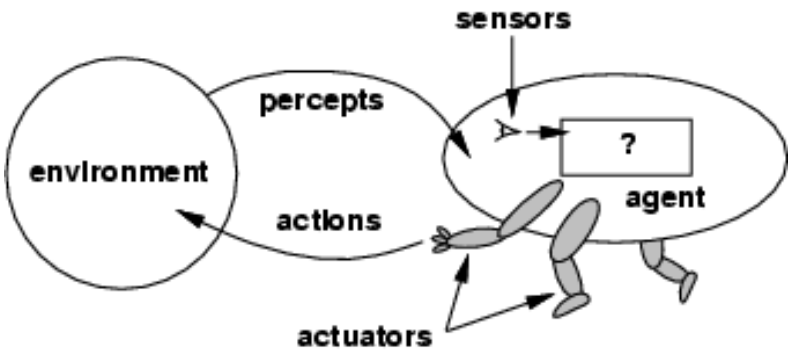
- Cognitive science
  - An interdisciplinary field that combines AI (including algorithms) and psychology
  - The goal:  to determine how the human mind works and to build computer systems that mimic this process
  - Based on experimental investigation of humans and animals

## Thinking rationally: "laws of thought"

- **Logic (both classical and modern) is taken as the basis for intelligence.**
- **Main obstacles**
  - **It’s difficult to translate informal knowledge into logical notation;**
  - **Problems that can be solved in principle may not be solvable in practice using logic.**

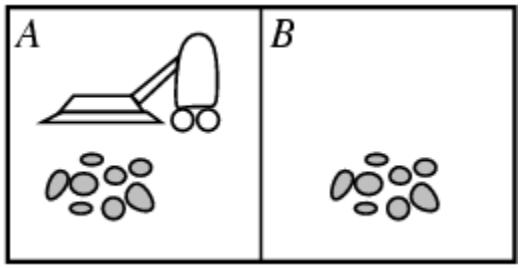
# Acting rationally: rational agent

- AI is viewed as the study and construction of rational agents
- An agent is anything that can be viewed as perceiving its environment through sensors and acting upon that environment through actuators




  - Human agent: eyes, ears, and other organs for sensors; hands, legs, mouth, and other body parts for actuators
  - Robotic agent: cameras and infrared range finders for sensors; various motors for actuators

# Vacuum-cleaner world




- **Percepts: location and contents, e.g., [A, Dirty]**
- **Actions: *Left*, *Right*, *Suck*, *NoOp***

## Acting rationally: rational agent

- An agent should strive to "do the right thing", based on what it can perceive and the actions it can perform
- The right action is the one that will cause the agent to be most successful
  - Performance measure: An objective criterion for success of an agent's behavior
- The agent function maps from percept histories to actions:

$$[f: \mathcal{P}^* \rightarrow \mathcal{A}]$$

- The agent program runs on the physical architecture to produce *f*

## Rational agents

- A rational agent chooses whichever action maximizes the expected value of the performance measure given the percept sequence to date
- Rationality ≠ omniscience
  - Percepts may not supply all relevant information
- Rationality ≠ clairvoyant
  - Action outcomes may not be as expected
- Hence, rationality ≠ successful

## PEAS

- To design a rational agent, we must first specify the setting for intelligent agent design
- PEAS: Performance measure, Environment, Actuators, Sensors
- Consider, e.g., a medical diagnosis system
  - Performance measure:
  - Environment:
  - Actuators:
  - Sensors:

## PEAS

- **An automated taxi driver:**
  - **Performance measure**
  - **Environment**
  - **Actuators**
  - **Sensors**

## Environment types

- **Fully observable (vs. partially observable):**
  - An agent's sensors give it access to the complete state of the environment at each point in time.
- **Deterministic (vs. stochastic):**
  - The next state of the environment is completely determined by the current state and the action executed by the agent. (If the environment is deterministic except for the actions of other agents, then the environment is strategic)
- **Episodic (vs. sequential):**
  - The agent's experience is divided into atomic "episodes" (each episode consists of the agent perceiving and then performing a single action), and the choice of action in each episode depends only on the episode itself.

## Environment types

- **Static (vs. dynamic):**
  - The environment is unchanged while an agent is deliberating. (The environment is semidynamic if the environment itself does not change with the passage of time but the agent's performance score does)
- **Discrete (vs. continuous):**
  - A limited number of distinct, clearly defined percepts and actions.
- **Single agent (vs. multiagent):**
  - An agent operating by itself in an environment.

# Environment types

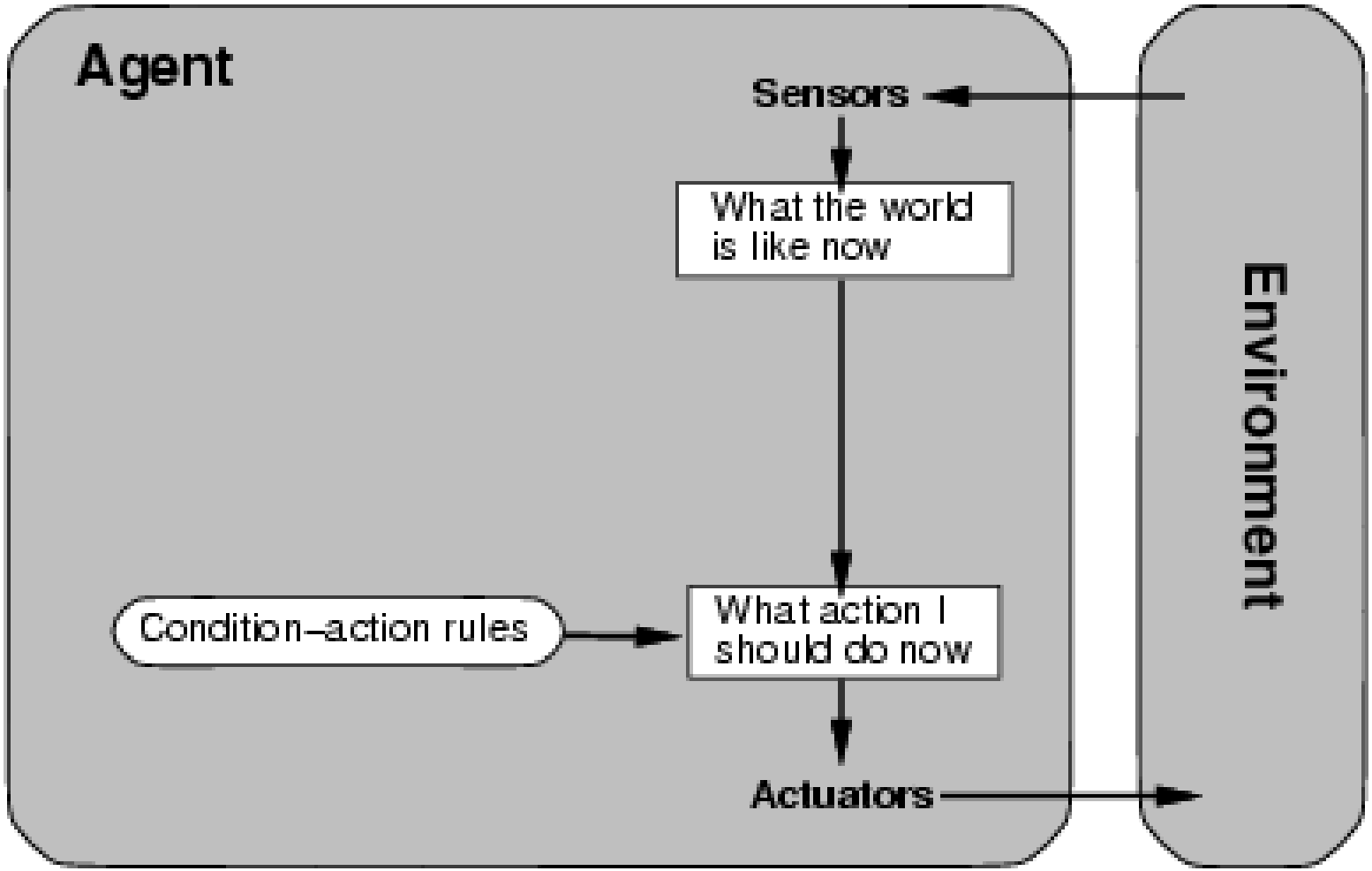
| | Go | Bridge | Taxi driving |
|---|---|---|---|
| Fully observable | | | |
| Deterministic | | | |
| Episodic | | | |
| Static | | | |
| Discrete | | | |
| Single agent | | | |

- The environment type largely determines the agent design
- The real world is (of course) partially observable, stochastic, sequential, dynamic, continuous, multi-agent

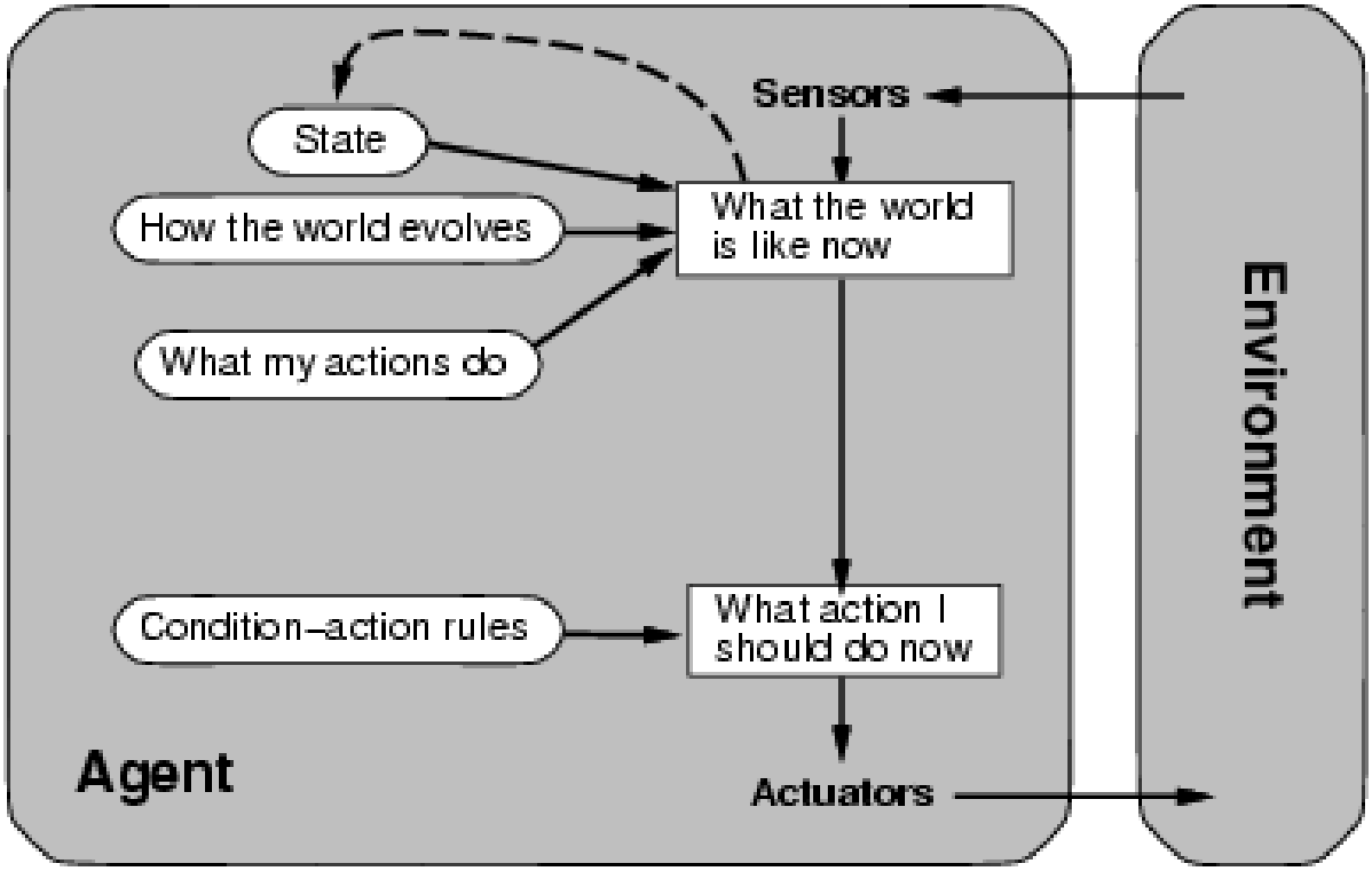
## Agent types

- Four basic types in order of increasing generality:
  - Simple reflex agents
  - Model-based reflex agents
  - Goal-based agents
  - Utility-based agents
- All these can be turned into learning agents

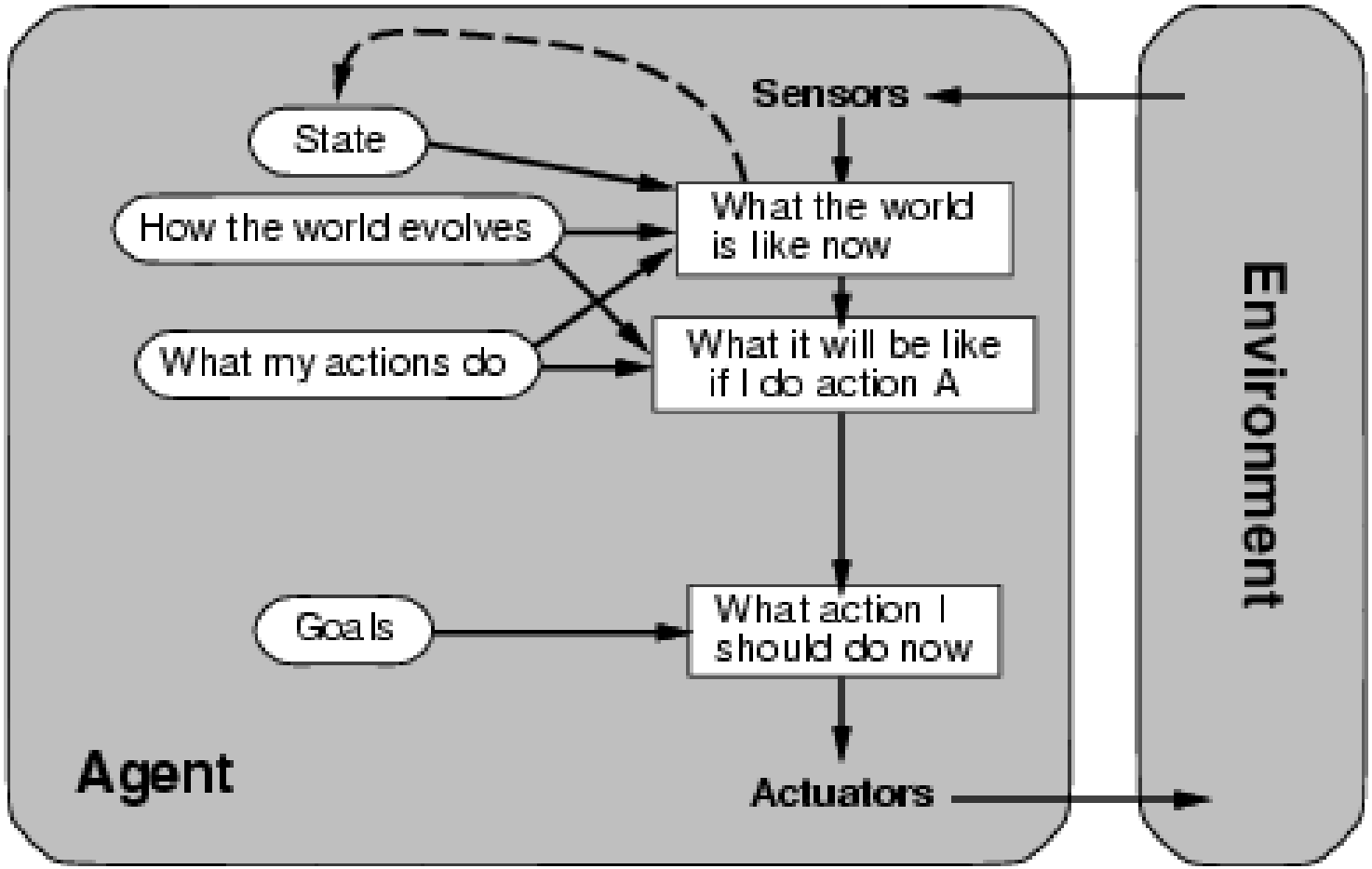
# Simple reflex agents



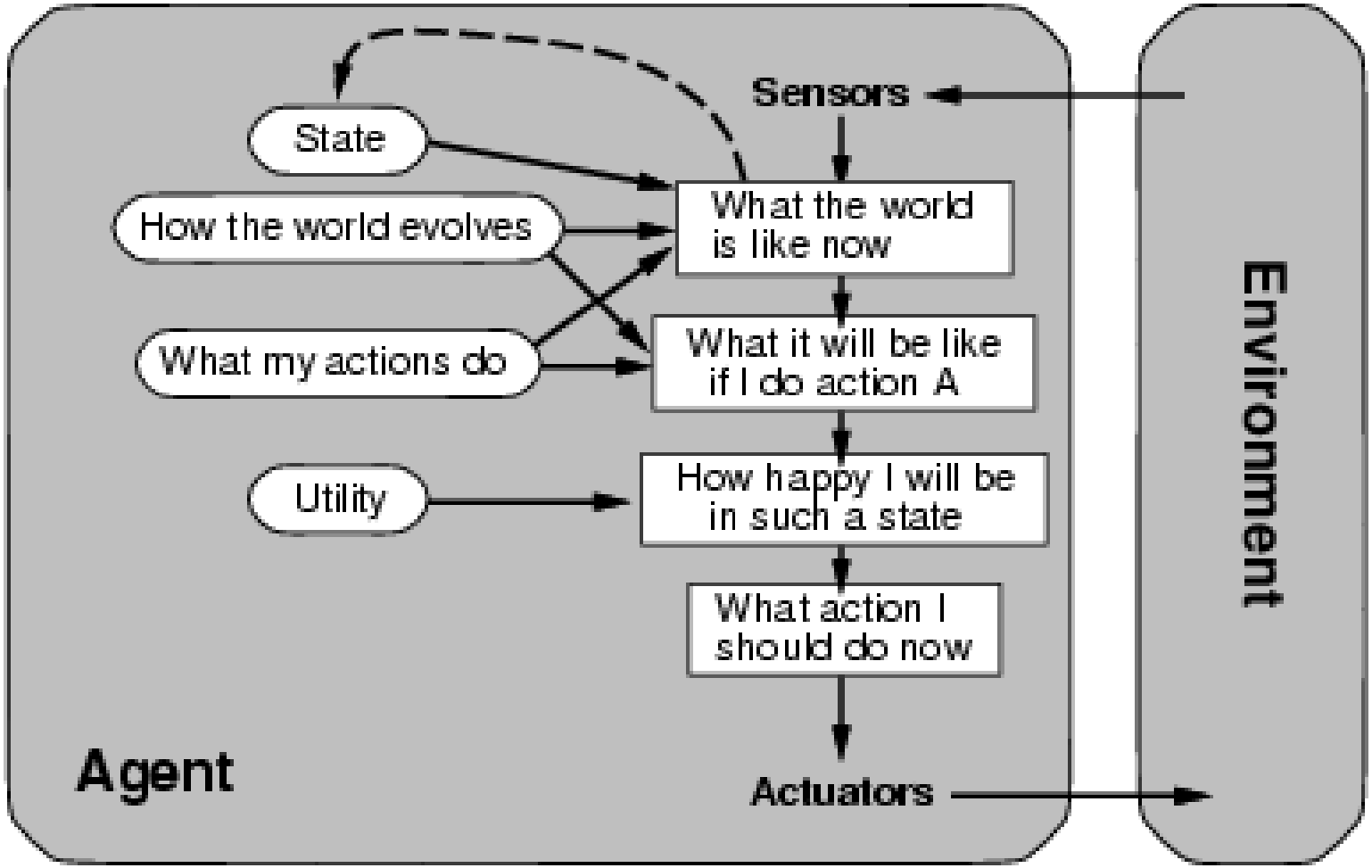
# Model-based reflex agents



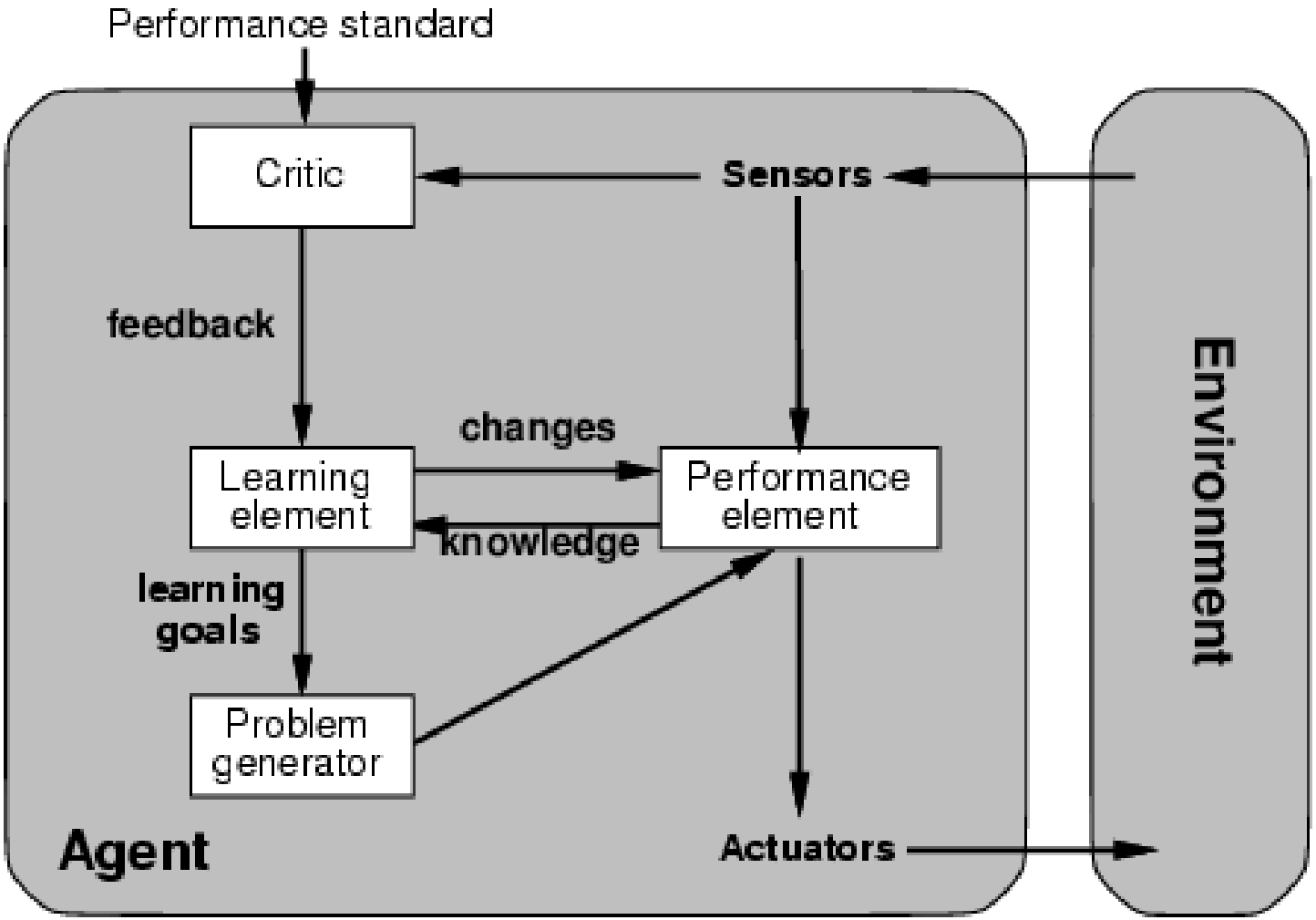
# Goal-based agents



# Utility-based agents



# Learning agents



# AI prehistory

| | |
|---|---|
| • Philosophy | Logic, methods of reasoning, mind as physical system foundations of learning, language, rationality |
| • Mathematics | Formal representation and proof algorithms, computation, (un)decidability, (in)tractability, probability |
| • Economics | utility, decision theory |
| • Neuroscience | physical substrate for mental activity |
| • Psychology | phenomena of perception and motor control, experimental techniques |
| • Computer engineering | building fast computers |
| • Control theory | design systems that maximize an objective function over time |
| • Linguistics | knowledge representation, grammar |

# Abridged history of AI

| | |
|---|---|
| • 1943 | McCulloch & Pitts: Boolean circuit model of brain |
| • 1950 | Turing's "Computing Machinery and Intelligence" |
| • 1956 | Dartmouth meeting: "Artificial Intelligence" adopted |
| • 1952—69 | Look, Ma, no hands! |
| • 1950s | Early AI programs, including Samuel's checkers program, Newell & Simon's Logic Theorist, Gelernter's Geometry Engine |
| • 1965 | Robinson's complete algorithm for logical reasoning |
| • 1966—73 | AI discovers computational complexity<br>Neural network research almost disappears |
| • 1969—79 | Early development of knowledge-based systems |
| • 1980-- | AI becomes an industry |
| • 1986-- | Neural networks return to popularity |
| • 1987-- | AI becomes a science |
| • 1995-- | The emergence of intelligent agents |

# State of the art

- Alpha Go beats human champion go players in 2016/7
- Watson defeated human Jeopardy champions in 2011
- Deep Blue defeated the reigning world chess champion Garry Kasparov in 1997
- Proved a mathematical conjecture (Robbins conjecture) unsolved for decades
- No hands across America (driving autonomously 98% of the time from Pittsburgh to San Diego)
- During the 1991 Gulf War, US forces deployed an AI logistics planning and scheduling program that involved up to 50,000 vehicles, cargo, and people
- NASA's on-board autonomous planning program controlled the scheduling of operations for a spacecraft
- Proverb solves crossword puzzles better than most humans
- … …