Agents

Artificial Intelligence @ Allegheny College

Janyl Jumadinova

January 23 - 27, 2023

Weak AI vs. Strong AI

- **Strong (General) AI:** Computer software + hardware alone can emulate a human mind. There is no fundamental difference between man and machine.

Weak AI vs. Strong AI

- **Strong (General) AI:** Computer software + hardware alone can emulate a human mind. There is no fundamental difference between man and machine.
- **Weak (Narrow) AI:** Computer software + hardware alone can simulate every aspect of a human mind. Only people can think, machines cannot.

What is AI?

THOUGHT	**Systems that think like humans**	**Systems that think rationally**
BEHAVIOUR	**Systems that act like humans**	**Systems that act rationally**
	HUMAN	RATIONAL

Acting humanly: The Turing test

Turing (1950) "Computing machinery and intelligence":

- "*Can machines think?*" $\longrightarrow$ "*Can machines behave intelligently?*"
- Operational test for intelligent behavior: the *Imitation Game*

Thinking humanly: Cognitive Science

Requires scientific theories of internal activities of the brain

- What level of abstraction? "*Knowledge*" or "*circuits*"?
- How to validate? Requires
  1. Predicting and testing behavior of human subjects (top-down) or
  2. Direct identification from neurological data (bottom-up)

Both approaches (roughly, *Cognitive Science* and *Cognitive Neuroscience*) are now distinct from AI

Thinking and Acting rationally

- *Normative* (or *prescriptive*) rather than *descriptive*
- Direct line through mathematics and philosophy to modern AI

Thinking and Acting rationally

- *Normative* (or *prescriptive*) rather than *descriptive*
- Direct line through mathematics and philosophy to modern AI
- Rational behavior: **doing the right thing**
- The right thing: that which is expected to maximize goal achievement, given the available information

What is AI?

Systems that think like humans	**Systems that think rationally**
Systems that act like humans	**Systems that act rationally**

Agents and environments

An agent is something that acts in an environment.

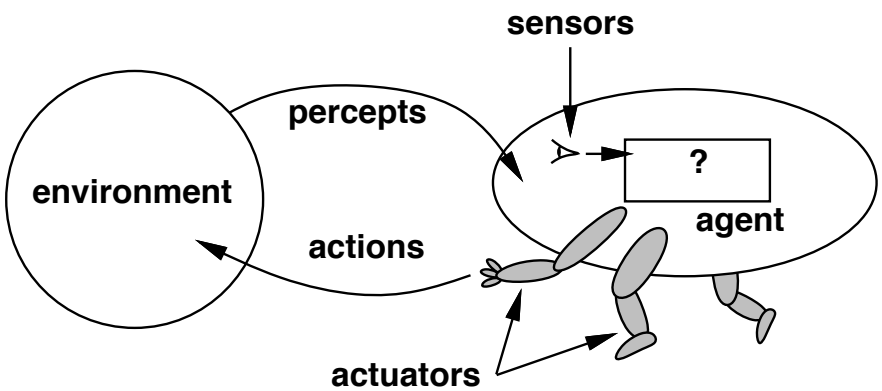
Agents and environments

An agent is something that acts in an environment.

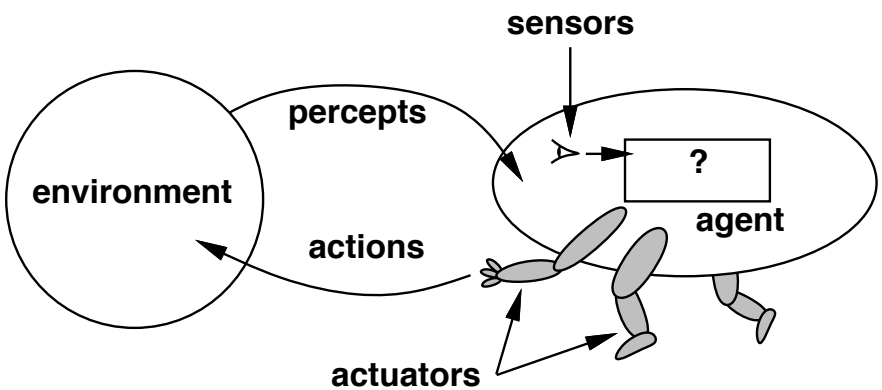
An agent acts *intelligently* if:

- 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.

Agents and environments

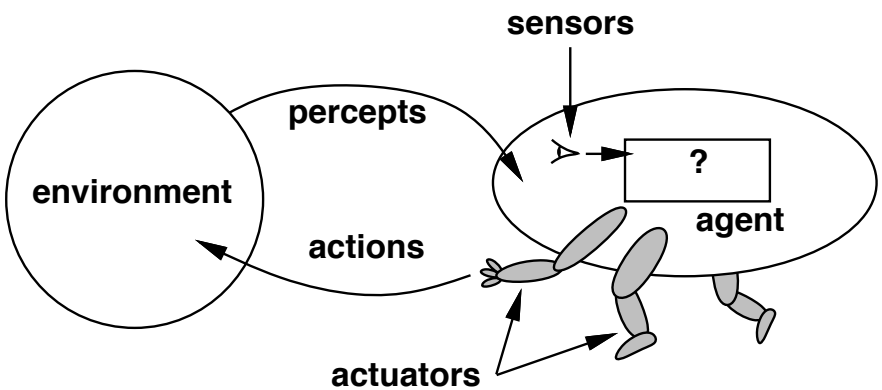


Agents and environments



Agents include humans, robots, softbots, thermostats, etc.

Agents and environments

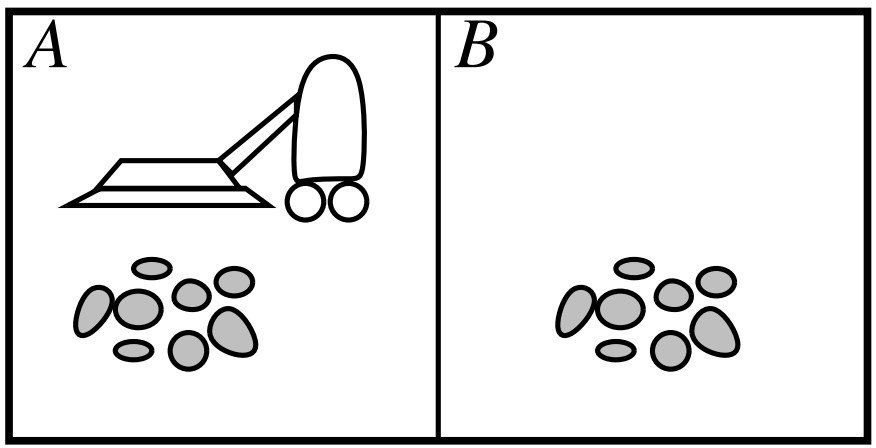


Agents include humans, robots, softbots, thermostats, etc.
The *agent function* maps from percept histories to actions:

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

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

A vacuum cleaner agent



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

A vacuum cleaner agent

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*
⋮	⋮

What is the right function?

What makes an agent good or bad, intelligent or stupid?

Agents and environments

For any given class of environments and tasks, we seek the agent (or class of agents) with the best performance.

Agents and environments

For any given class of environments and tasks, we seek the agent (or class of agents) with the best performance.

Caveat: computational limitations make perfect rationality unachievable.

Agents and environments

For any given class of environments and tasks, we seek the agent (or class of agents) with the best performance.

Caveat: computational limitations make perfect rationality unachievable.
$\longrightarrow$ design best program for given machine resources.

Restricting the definition of an **agent**: an ideal agent

- **Autonomy**: The ability to operate without the direct intervention of humans or others, and have some kind of control over their actions and internal state.

Restricting the definition of an **agent**: an ideal agent

- **Autonomy**: The ability to operate without the direct intervention of humans or others, and have some kind of control over their actions and internal state.
- (Structural) **Reactivity**: The ability to perceive the environment, and respond regularly to changes that occur in it.

Restricting the definition of an **agent**: an ideal agent

- **Autonomy**: The ability to operate without the direct intervention of humans or others, and have some kind of control over their actions and internal state.
- (Structural) **Reactivity**: The ability to perceive the environment, and respond regularly to changes that occur in it.
- **Social Ability**: The ability to interact with other agents (and possibly humans).

Restricting the definition of an **agent**: an ideal agent

- **Autonomy**: The ability to operate without the direct intervention of humans or others, and have some kind of control over their actions and internal state.
- (Structural) **Reactivity**: The ability to perceive the environment, and respond regularly to changes that occur in it.
- **Social Ability**: The ability to interact with other agents (and possibly humans).
- **Pro-Activity**: The ability to exhibit goal-directed behavior by taking the initiative instead of just acting in response.

Restricting the definition of an **agent**: an ideal agent

Other attributes:

- **Mobility**: The ability to move around an electronic network.

Restricting the definition of an **agent**: an ideal agent

Other attributes:

- **Mobility**: The ability to move around an electronic network.
- **Veracity**: The assumption of not communicating false information knowingly.

Restricting the definition of an **agent**: an ideal agent

Other attributes:

- **Mobility**: The ability to move around an electronic network.
- **Veracity**: The assumption of not communicating false information knowingly.
- **Benevolence**: The assumption of not having conflicting goals.

Restricting the definition of an **agent**: an ideal agent

Other attributes:

- **Mobility**: The ability to move around an electronic network.
- **Veracity**: The assumption of not communicating false information knowingly.
- **Benevolence**: The assumption of not having conflicting goals.
- **Rationality**: The assumption of acting with a view to achieve its goals, instead of preventing them.

Agents vs. Objects

OOP versus AOP

	OOP	AOP
Basic unit	object	agent
Parameters defining state of basic unit	unconstrained	beliefs, commitments, capabilities, choices, ...
Process of computation	message passing and response methods	message passing and response methods
Types of message	unconstrained	inform, request, offer, promise, decline, ...
Constraints on methods	none	honesty, consistency, ...

"Agent-Oriented Programming", Y. Shoham

Agents vs. Objects

- **Object-Oriented Design**: objects have identity, state and behaviour and communicate via messages.
- **Agent-Oriented Approach**: agents have identity, state (knowledge, beliefs, desires, intentions) and behaviour(goal-achieving, actions, reactions) and communication abilities.

Agents vs. Objects

- **Object-Oriented Design**: objects have identity, state and behaviour and communicate via messages.
- **Agent-Oriented Approach**: agents have identity, state (knowledge, beliefs, desires, intentions) and behaviour(goal-achieving, actions, reactions) and communication abilities.

Then, are objects agents?

Agents vs. Objects

- Agents exhibit autonomy, they have control over their state, execution and behavior.
- Agents exhibit goal-directed, reactive and social behavior.
- Agents are persistent, self-aware and able to learn and adapt.
- Control in multi-agent systems is distributed.

Agents vs. Objects

- Agents exhibit autonomy, they have control over their state, execution and behavior.
- Agents exhibit goal-directed, reactive and social behavior.
- Agents are persistent, self-aware and able to learn and adapt.
- Control in multi-agent systems is distributed.

Objects do not have these characteristics.

Agents vs. Objects

- Agents have the quality of volition.
  - using AI techniques, intelligent agents are able to judge their results, and then modify their behavior (and thus their own internal structure) to improve their perceived fitness.

Agents vs. Objects

- Agents have the quality of volition.
  - using AI techniques, intelligent agents are able to judge their results, and then modify their behavior (and thus their own internal structure) to improve their perceived fitness.
- Objects are abstractions of things like invoices. Agents are abstractions of intelligent beings – they are essentially anthropomorphic.

  Note that this does not mean that agents are intelligent in the human sense, only that they are modeled after an anthropomorphic architecture, with beliefs, desires, etc.

Agents vs. Objects

Group Think Tank

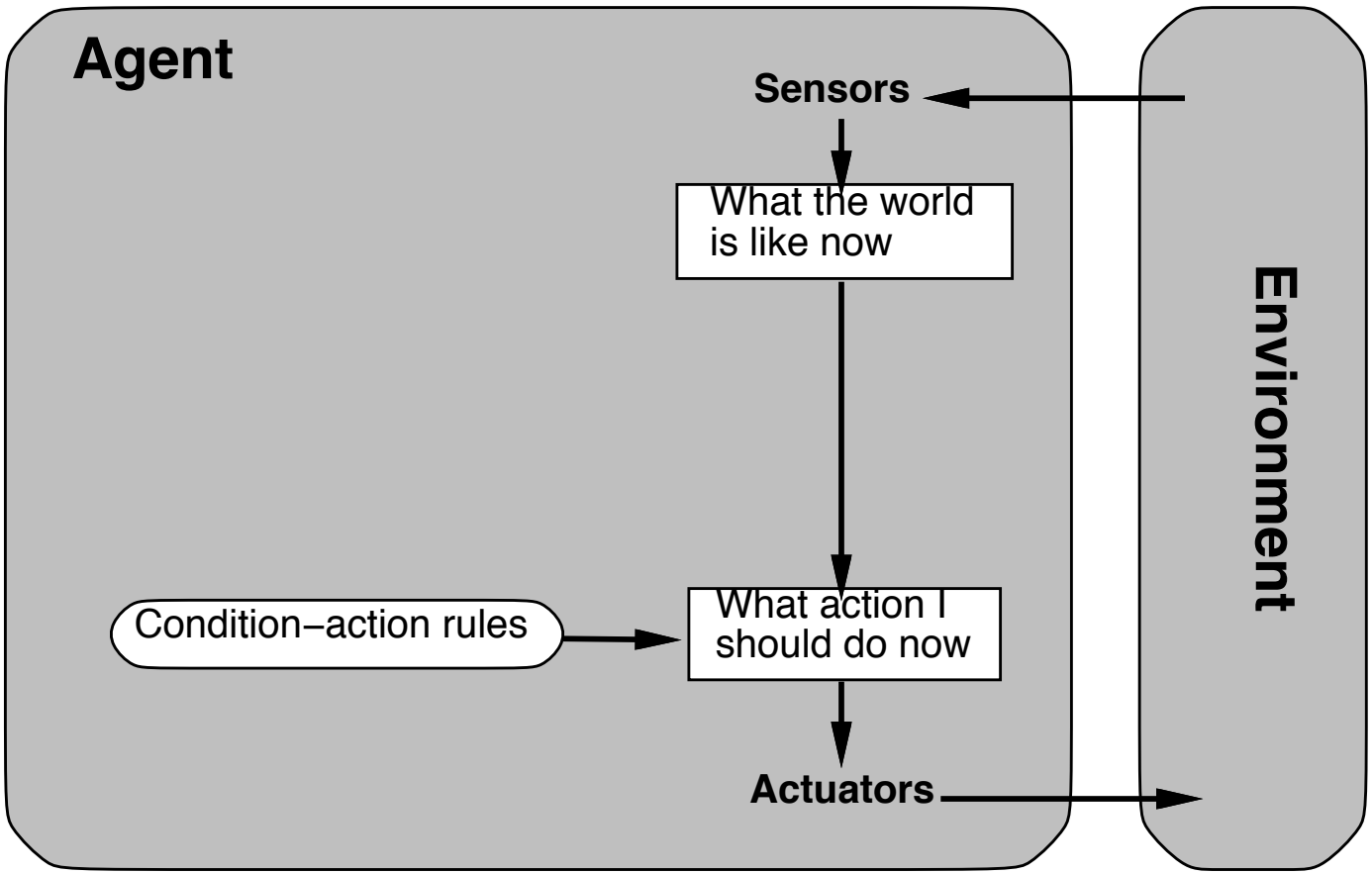
- Design an object-oriented solution and an agent-oriented solution for a car wash task.
- Identify why it is an object-oriented or an agent-oriented solution.
- List agents and objects for both solutions.

Getting to an ideal agent

Agent types in order of increasing generality:

- simple reflex agents
- reflex agents with state
- goal-based agents
- utility-based agents
- learning agents

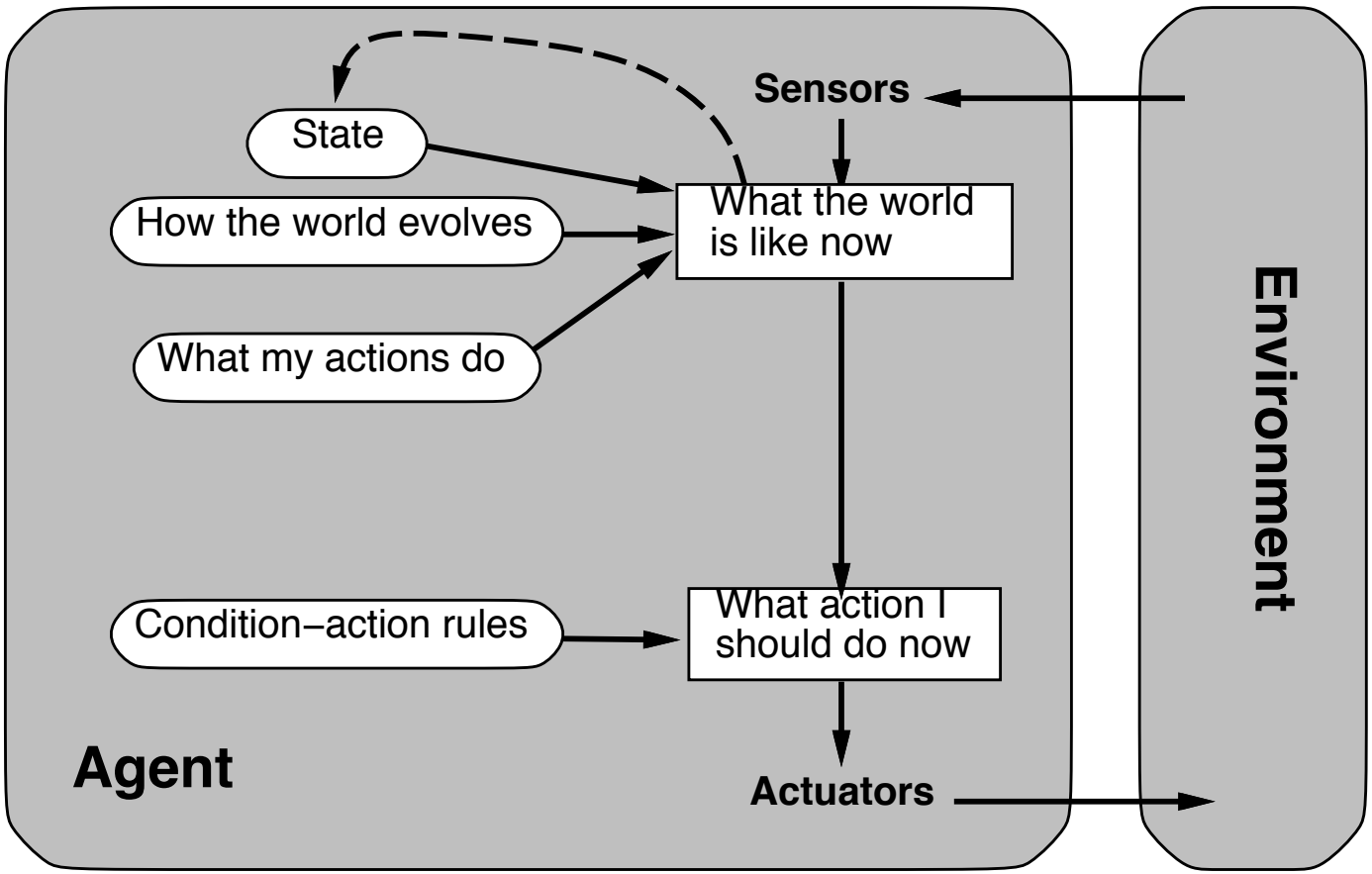
Simple Reflex Agent



Simple Reflex Agent - An Example

```
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
```

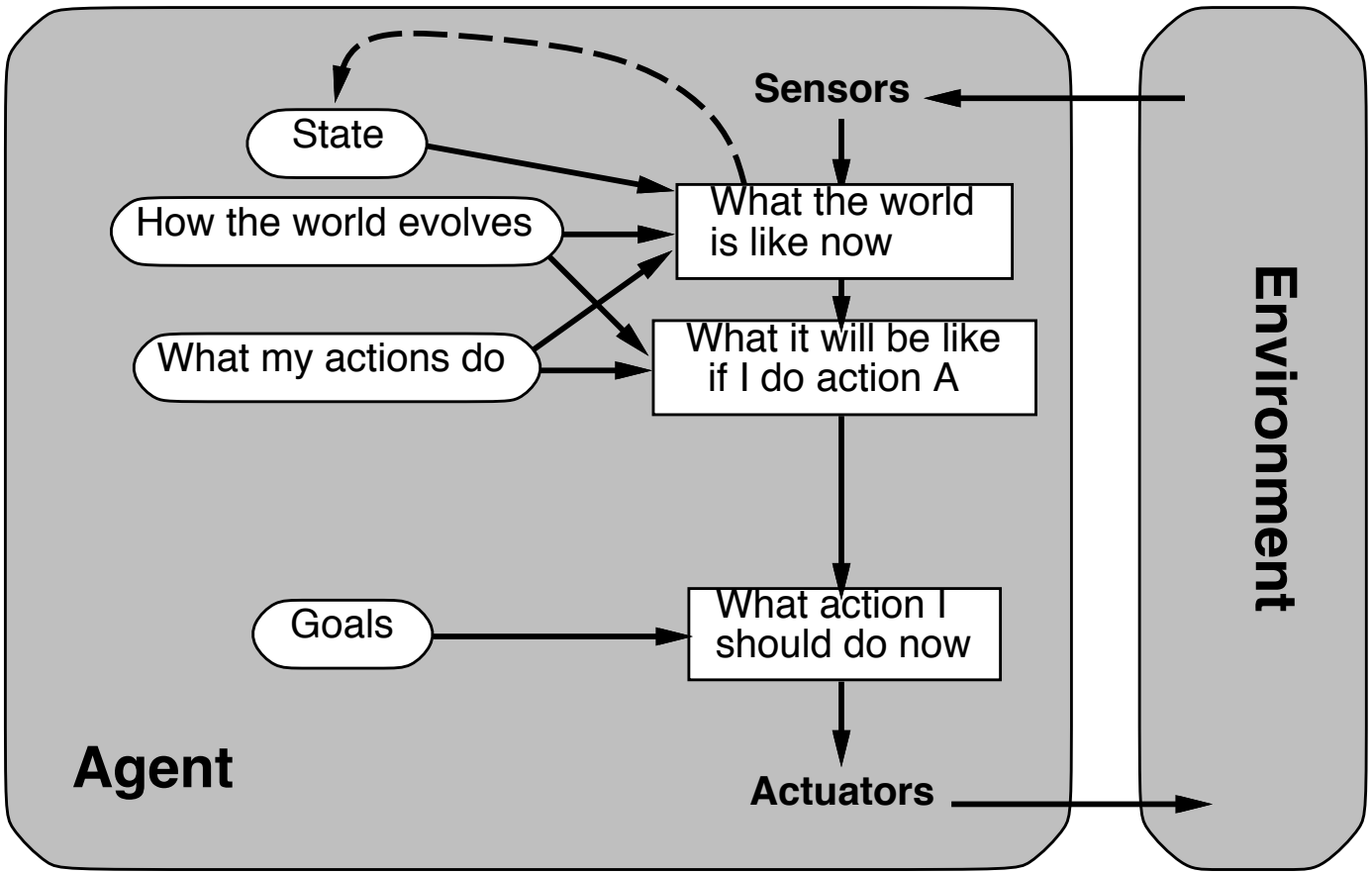
Reflex Agent with State



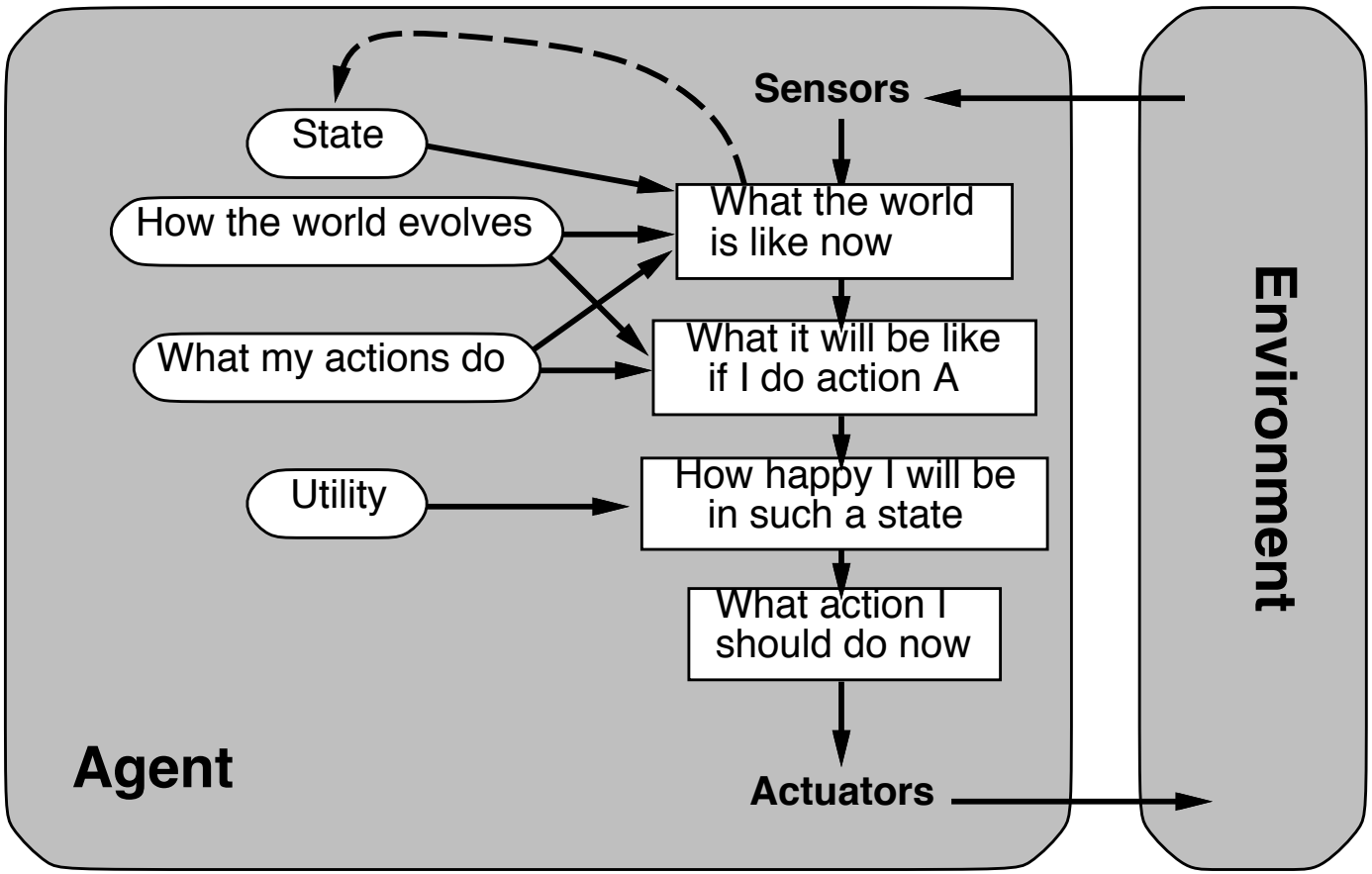
Reflex Agent with State - An Example

```
function REFLEX-VACUUM-AGENT([location,status]) returns an action
static: last_A, last_B, numbers, initially ∞
    if status = Dirty then ...
```

Goal-based Agent

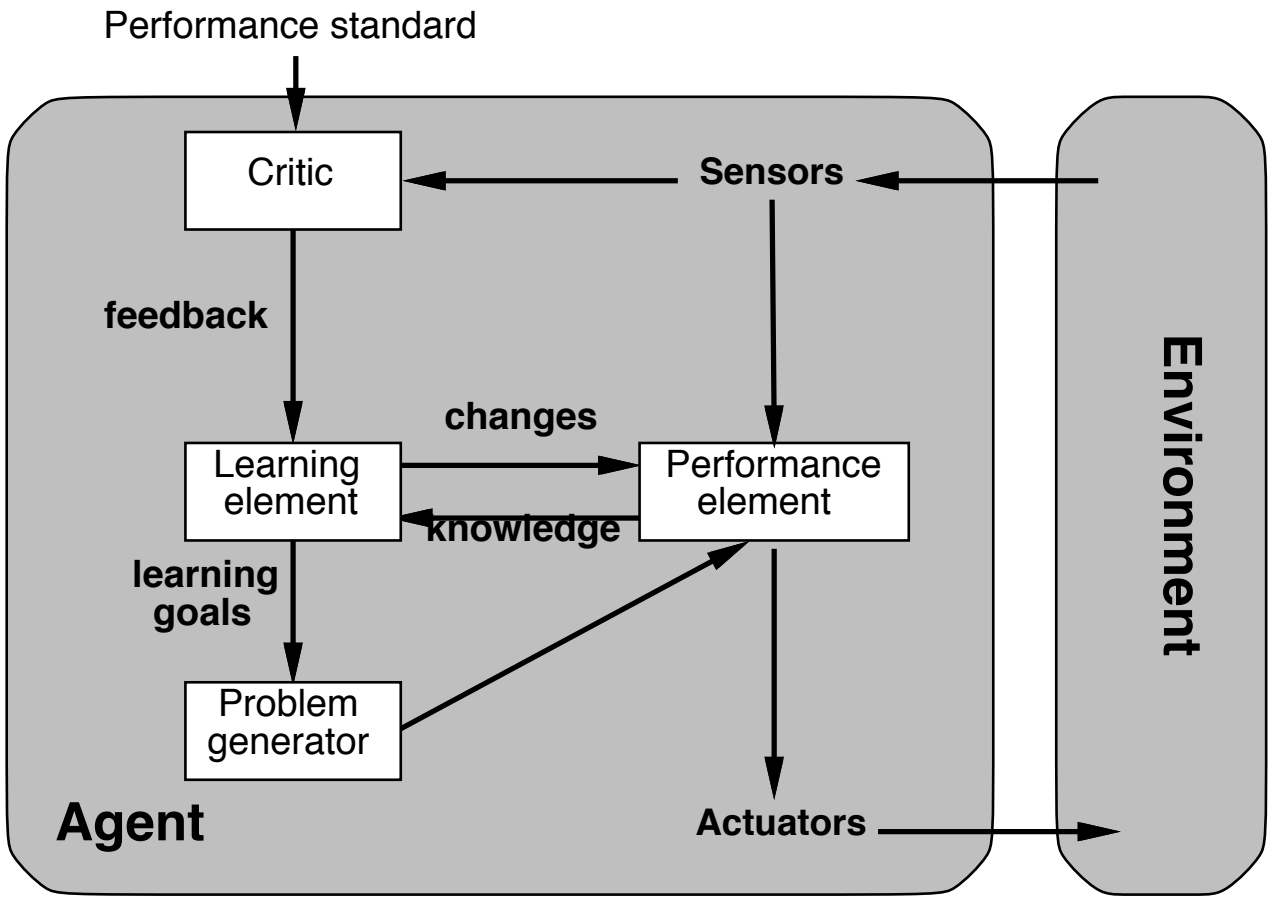


Utility-based Agent



Learning Agent

All the previous agents can be turned into learning agents



Rational Agents

- A rational agent chooses whichever action maximizes the **expected** value of the performance measure **given the percept sequence to date**.
- A system is rational if it does the "right thing", given what it knows.

Rationality

Fixed performance measure evaluates the *environment sequence*

- one point per square cleaned up in time T?
- one point per clean square per time step, minus one per move?
- penalize for $> k$ dirty squares?

Rationality

Fixed performance measure evaluates the *environment sequence*

- one point per square cleaned up in time T?
- one point per clean square per time step, minus one per move?
- penalize for $> k$ dirty squares?

A rational agent chooses whichever action maximizes the **expected** value of the performance measure **given the percept sequence to date**

Rationality

Rational $\neq$ omniscient

– percepts may not supply all relevant information

Rationality

Rational $\neq$ omniscient
– percepts may not supply all relevant information
Rational $\neq$ clairvoyant
– action outcomes may not be as expected

Rationality

Rational $\neq$ omniscient
– percepts may not supply all relevant information
Rational $\neq$ clairvoyant
– action outcomes may not be as expected
Hence, rational $\neq$ successful

Rationality

Rational $\neq$ omniscient
– percepts may not supply all relevant information
Rational $\neq$ clairvoyant
– action outcomes may not be as expected
Hence, rational $\neq$ successful
Rational $\implies$ exploration, learning, autonomy

PEAS

To design a rational agent, we must specify the task environment:
Performance measure
Environment
Actuators
Sensors

PEAS

To design a rational agent, we must specify the task environment.

Consider, e.g., the task of designing an **automated taxi**:
Performance measure safety, destination, profits, legality, comfort, ...

PEAS

To design a rational agent, we must specify the task environment.

Consider, e.g., the task of designing an **automated taxi**:
Performance measure safety, destination, profits, legality, comfort, ...
Environment

PEAS

To design a rational agent, we must specify the task environment.

Consider, e.g., the task of designing an **automated taxi**:
Performance measure safety, destination, profits, legality, comfort, ...
Environment US streets/freeways, traffic, pedestrians, weather, ...

PEAS

To design a rational agent, we must specify the task environment.

Consider, e.g., the task of designing an **automated taxi**:
Performance measure safety, destination, profits, legality, comfort, ...
Environment US streets/freeways, traffic, pedestrians, weather, ...
Actuators

PEAS

To design a rational agent, we must specify the task environment.

Consider, e.g., the task of designing an **automated taxi**:
Performance measure safety, destination, profits, legality, comfort, ...
Environment US streets/freeways, traffic, pedestrians, weather, ...
Actuators steering, accelerator, brake, horn, speaker/display, ...

PEAS

To design a rational agent, we must specify the task environment.

Consider, e.g., the task of designing an **automated taxi**:
Performance measure safety, destination, profits, legality, comfort, ...
Environment US streets/freeways, traffic, pedestrians, weather, ...
Actuators steering, accelerator, brake, horn, speaker/display, ...
Sensors

PEAS

To design a rational agent, we must specify the task environment.

Consider, e.g., the task of designing an **automated taxi**:
Performance measure safety, destination, profits, legality, comfort, ...
Environment US streets/freeways, traffic, pedestrians, weather, ...
Actuators steering, accelerator, brake, horn, speaker/display, ...
Sensors video, accelerometers, gauges, engine sensors, keyboard, GPS, ...

Internet shopping agent?

Internet shopping agent?

Performance measure

Internet shopping agent?

Performance measure price, quality, appropriateness, efficiency, . . .

Internet shopping agent?

Performance measure price, quality, appropriateness, efficiency, . . .
Environment

Internet shopping agent?

Performance measure price, quality, appropriateness, efficiency, ...
Environment current and future WWW sites, vendors, shippers, ...

Internet shopping agent?

Performance measure price, quality, appropriateness, efficiency, ...
Environment current and future WWW sites, vendors, shippers, ...
Actuators

Internet shopping agent?

Performance measure price, quality, appropriateness, efficiency, ...
Environment current and future WWW sites, vendors, shippers, ...
Actuators display to user, follow URL, fill in form, ...

Internet shopping agent?

Performance measure price, quality, appropriateness, efficiency, ...
Environment current and future WWW sites, vendors, shippers, ...
Actuators display to user, follow URL, fill in form, ...
Sensors

Internet shopping agent?

Performance measure price, quality, appropriateness, efficiency, ...
Environment current and future WWW sites, vendors, shippers, ...
Actuators display to user, follow URL, fill in form, ...
Sensors HTML pages (text, graphics, scripts), ...

Environment Types

- Fully Observable: the agent can observe the state of the world, *vs.* Partially Observable: there can be a number states that are possible given the agent's observations

Environment Types

- Fully Observable: the agent can observe the state of the world, *vs.*
  Partially Observable: there can be a number states that are possible given the agent's observations
- Deterministic: the resulting state is determined from the action and the state, *vs.*
  Stochastic: there is uncertainty about the resulting state

Environment Types

- Fully Observable: the agent can observe the state of the world, *vs.*
  Partially Observable: there can be a number states that are possible given the agent's observations
- Deterministic: the resulting state is determined from the action and the state, *vs.*
  Stochastic: there is uncertainty about the resulting state
- Episodic: agent's experience is divided into atomic episodes, *vs.*
  Sequential: the current decision could affect all future decisions

Environment Types

- Static: environment does not change, *vs.*
  Dynamic: the environment can change while an agent is deliberating, *vs.*
  Semi: the environment itself does not change with the passage of time but the agent's performance score does

Environment Types

- Static: environment does not change, *vs.*
  Dynamic: the environment can change while an agent is deliberating, *vs.*
  Semi: the environment itself does not change with the passage of time but the agent's performance score does
- Discrete vs. Continuous: applies to the state of the environment, to the way time is handled, and to the percepts and actions of the agent

Environment Types

- Static: environment does not change, *vs.*
  Dynamic: the environment can change while an agent is deliberating, *vs.*
  Semi: the environment itself does not change with the passage of time but the agent's performance score does
- Discrete vs. Continuous: applies to the state of the environment, to the way time is handled, and to the percepts and actions of the agent
- Single-agent vs. Multi-agent

Environment Types

- Static: environment does not change, *vs.*
  Dynamic: the environment can change while an agent is deliberating, *vs.*
  Semi: the environment itself does not change with the passage of time but the agent's performance score does
- Discrete vs. Continuous: applies to the state of the environment, to the way time is handled, and to the percepts and actions of the agent
- Single-agent vs. Multi-agent

The environment type largely determines the agent design

Environment Types

	Solitaire
Observable	Yes
Deterministic	Yes
Episodic	No
Static	Yes
Discrete	Yes
Single-agent	Yes

Environment Types

	Solitaire	Internet shopping
Observable	Yes	No
Deterministic	Yes	Partly
Episodic	No	No
Static	Yes	Semi
Discrete	Yes	Yes
Single-agent	Yes	Yes (except auctions)

Environment Types

	Solitaire	Internet shopping	Taxi
Observable	Yes	No	No
Deterministic	Yes	Partly	No
Episodic	No	No	No
Static	Yes	Semi	No
Discrete	Yes	Yes	No
Single-agent	Yes	Yes (except auctions)	No

Environment Types

	Solitaire	Internet shopping	Taxi
Observable	Yes	No	No
Deterministic	Yes	Partly	No
Episodic	No	No	No
Static	Yes	Semi	No
Discrete	Yes	Yes	No
Single-agent	Yes	Yes (except auctions)	No

The real world is partially observable, stochastic, sequential, dynamic, continuous, multi-agent

Agent Summary

- Agents interact with environments through actuators and sensors

Agent Summary

- Agents interact with environments through actuators and sensors
- The **agent function** describes what the agent does in all circumstances

Agent Summary

- Agents interact with environments through actuators and sensors
- The **agent function** describes what the agent does in all circumstances
- The **performance measure** evaluates the environment sequence

Agent Summary

- Agents interact with environments through actuators and sensors
- The **agent function** describes what the agent does in all circumstances
- The **performance measure** evaluates the environment sequence
- A **perfectly rational** agent maximizes expected performance

Agent Summary

- Agents interact with environments through actuators and sensors
- The **agent function** describes what the agent does in all circumstances
- The **performance measure** evaluates the environment sequence
- A **perfectly rational** agent maximizes expected performance
- **Agent programs** implement (some) agent functions

Agent Summary

- Agents interact with environments through actuators and sensors
- The **agent function** describes what the agent does in all circumstances
- The **performance measure** evaluates the environment sequence
- A **perfectly rational** agent maximizes expected performance
- **Agent programs** implement (some) agent functions
- **PEAS** descriptions define task environments

Agent Summary

- Agents interact with environments through actuators and sensors
- The **agent function** describes what the agent does in all circumstances
- The **performance measure** evaluates the environment sequence
- A **perfectly rational** agent maximizes expected performance
- **Agent programs** implement (some) agent functions
- **PEAS** descriptions define task environments
- Environments are categorized along several dimensions:
  observable? deterministic? episodic? static? discrete? single-agent?

Agent Summary

- Agents interact with environments through actuators and sensors
- The **agent function** describes what the agent does in all circumstances
- The **performance measure** evaluates the environment sequence
- A **perfectly rational** agent maximizes expected performance
- **Agent programs** implement (some) agent functions
- **PEAS** descriptions define task environments
- Environments are categorized along several dimensions:
  observable? deterministic? episodic? static? discrete? single-agent?
- Several basic agent architectures exist:
  reflex, reflex with state, goal-based, utility-based, learning