What is Artificial Intelligence (AI)?

- Computational models of human behavior?
  - Programs that behave (externally) like humans
- Computational models of human "thought" processes?
  - Programs that operate (internally) the way humans do
- Computational systems that behave intelligently?
  - What does it mean to behave intelligently?
- Computational systems that behave rationally!

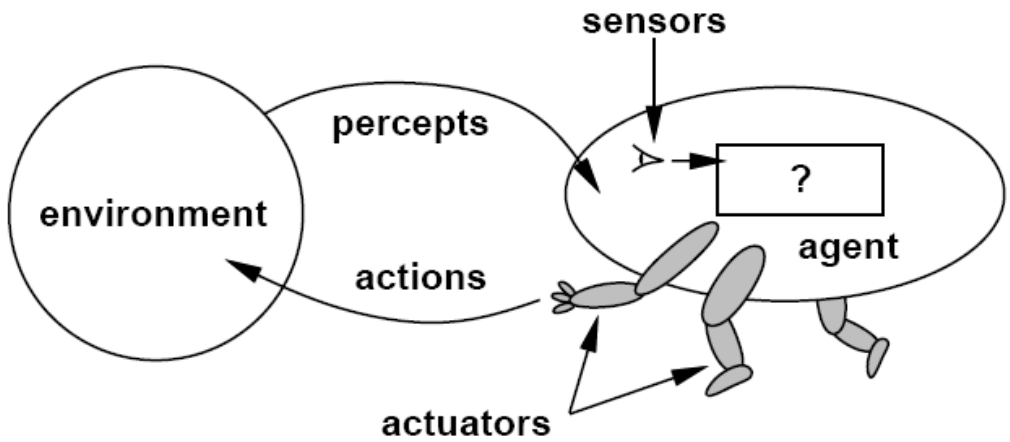
i.e. Do a "good job" of doing what they're supposed to do

Agents

- Definition
  - An (Intelligent) Agent perceives its environment via sensors and acts *rationally* upon that environment with its effectors.

    Hence an agent gets percepts one at a time, and maps the percept sequence to actions.

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

Agent Characteristics

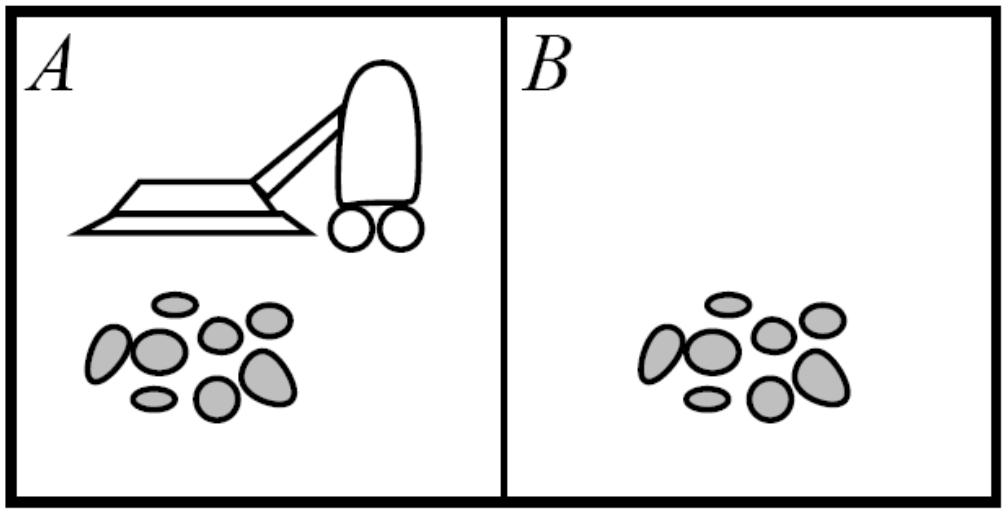
- Situatedness
  The agent receives some form of sensory input from its environment, and it performs some action that changes its environment in some way. Examples of environments: the physical world and the Internet.
- Autonomy
  The agent can act without direct intervention by humans or other agents and that it has control over its own actions and internal state.
- Adaptivity
  The agent is capable of (1) reacting flexibly to changes in its environment; (2) taking goal-directed initiative (i.e., is pro-active), when appropriate; and (3) learning from its own experience, its environment, and interactions with others.
- Sociability
  The agent is capable of interacting in a peer-to-peer manner with other agents or humans.

Rationality

- A rational agent takes actions it believes will achieve its goals.
  - Assume I don't like to get wet, so I bring an umbrella. Is that rational?
  - Depends on the weather forecast and whether I've heard it. If I've heard the forecast for rain (and I believe it) then bringing the umbrella is rational.

"Best possible" sequence of actions – maximize chosen performance metric

Vacuum-cleaner world



Percepts: location and contents, e.g., $[A, Dirty]$

Actions: $Left$, $Right$, $Suck$, $NoOp$

A vacuum-cleaner agent

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

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** $Left$

What is the **right** function?
Can it be implemented in a small agent program?

Examples of Agents

Agent Type	Percepts	Actions	Goals	Environment
Bin-Picking Robot	Images	Grasp objects; Sort into bins	Parts in correct bins	Conveyor belt
Medical Diagnosis	Patient symptoms, tests	Tests and treatments	Healthy patient	Patient & hospital
Web product finder	Web pages	navigate web, gather relevant products	Find best price for a product	Internet
Webcrawler Softbot	Web pages	Follow links, pattern matching	Collect info on a subject	Internet
Financial forecasting software	Financial data	Gather data on companies	Pick stocks to buy & sell	Stock market, company reports

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

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)

Classes of Environments

- Accessible (vs. Inaccessible)
  - Can you see the state of the world directly?
- Deterministic (vs. Non-Deterministic)
  - Does an action map one state into a single other state?
- Static (vs. Dynamic)
  - Can the world change while you are thinking?
- Discrete (vs. Continuous)
  - Are the percepts and actions discrete (like integers) or continuous (like reals)?

Environment types

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

The environment type largely determines the agent design

The real world is (of course) partially observable, stochastic, sequential, dynamic, continuous, multi-agent

Agent types

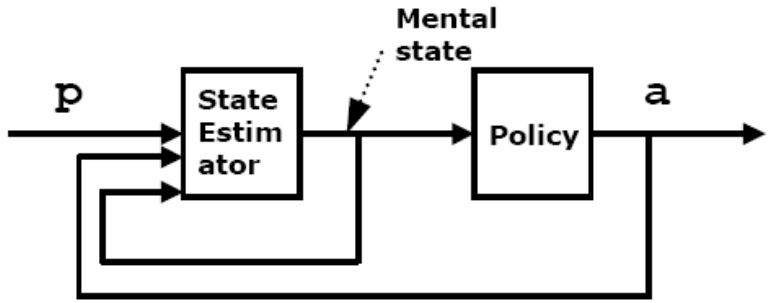
Four basic types in order of increasing generality:

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

All these can be turned into learning agents

Structures of Agents

- Agent with memory

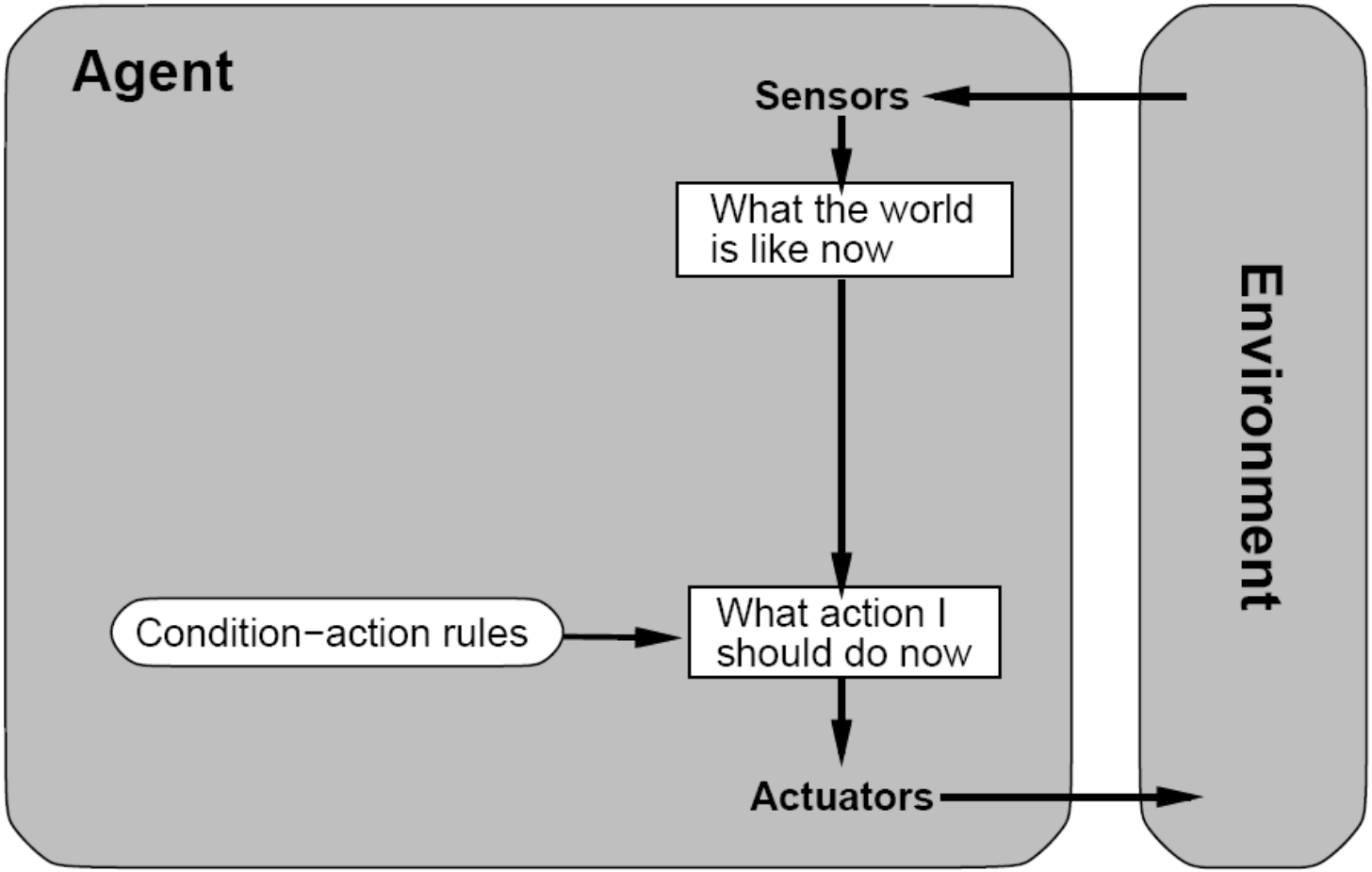


- State estimator/Memory
  - What we've chosen to remember from the history of percepts
  - Maps what you knew before, what you just perceived and what you just did, into what you know now.
- Problem of behavior: Given my mental state, what action should I take?

The second component is called the *policy*.

Dictates the (best) action choic based on the current state of the memory

Simple reflex agents



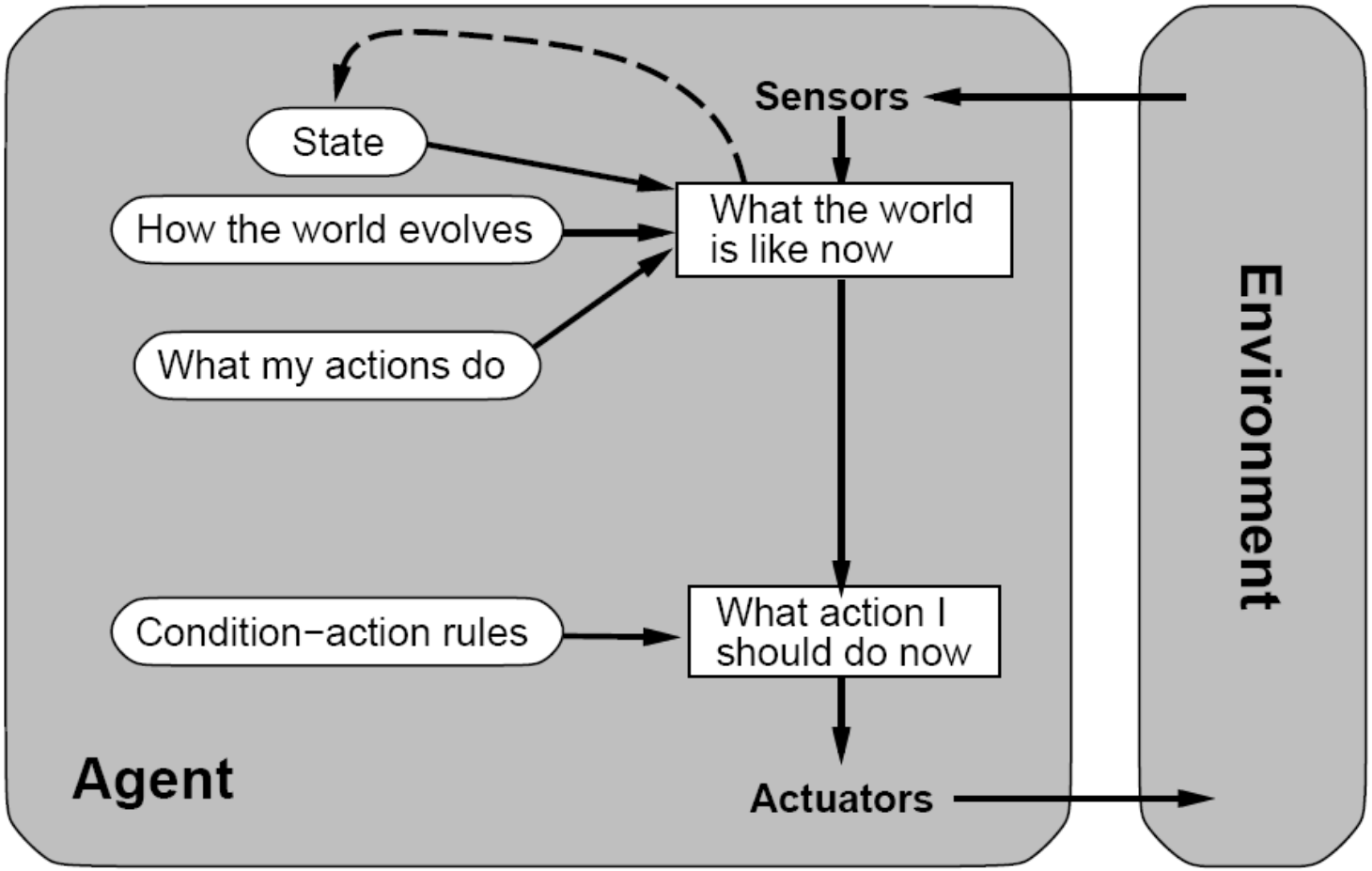
Example

```
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
```

Simple Reflex Agent

- Table lookup of percept-action pairs defining all possible condition-action rules necessary to interact in an environment
- Problems
  - Too big to generate and to store (Chess has about 10^{120} states, for example)
  - No knowledge of non-perceptual parts of the current state
  - Not adaptive to changes in the environment; requires entire table to be updated if changes occur
  - Looping: Can't make actions conditional

Reflex agents with state



Example

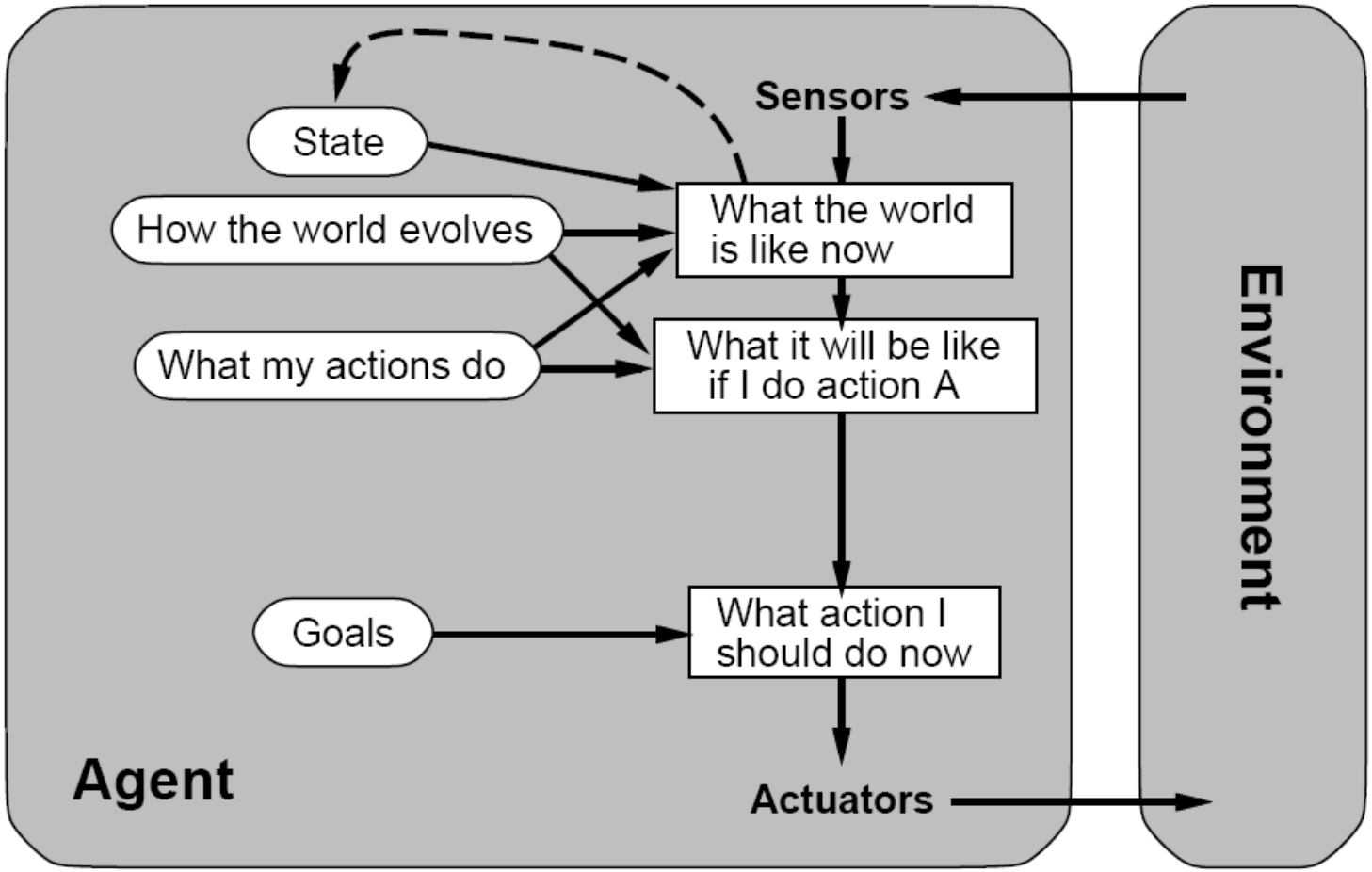
function REFLEX-VACUUM-AGENT([*location*,*status*]) **returns** an action
static: *last_A*, *last_B*, numbers, initially ∞

if *status* = *Dirty* **then** ...

Reflex Agent with Internal State

- Encode "internal state" of the world to remember the past as contained in earlier percepts
- Needed because sensors do not usually give the entire state of the world at each input, so perception of the environment is captured over time. "State" used to encode different "world states" that generate the same immediate percept.
- Requires ability to represent change in the world; one possibility is to represent just the latest state, but then can't reason about hypothetical courses of action
- Example: Rodney Brooks's Subsumption Architecture
  Main idea: build complex, intelligent robots by decomposing behaviors into a hierarchy of skills, each completely defining a complete percept-action cycle for one very specific task. For example, avoiding contact, wandering, exploring, recognizing doorways, etc. Each behavior is modeled by a finite-state machine with a few states (though each state may correspond to a complex function or module). Behaviors are loosely-coupled, asynchronous interactions.

Goal-based agents

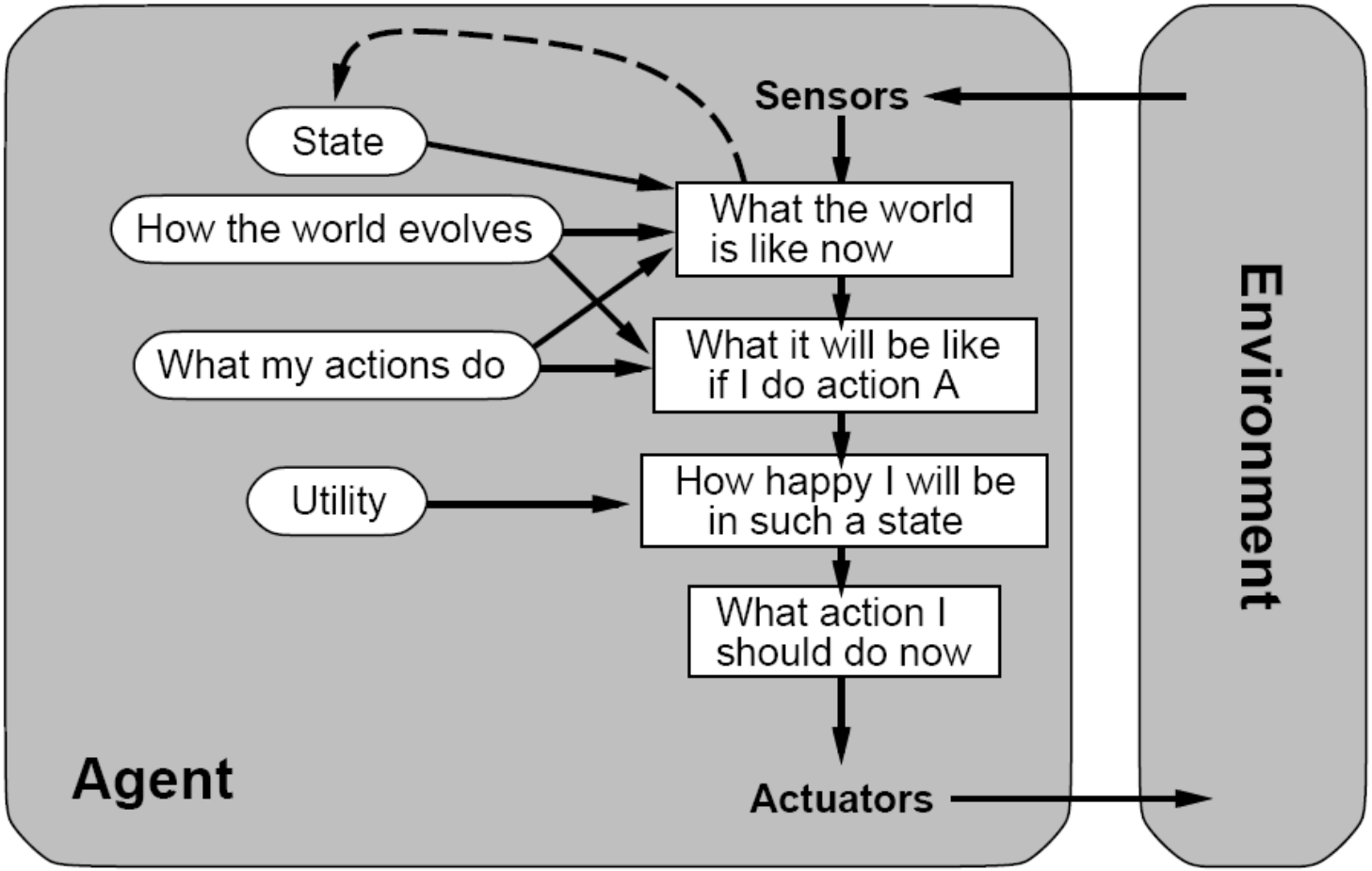


Goal-Based Agent

- Choose actions so as to achieve a (given or computed) goal= a description of a desirable situation
- Keeping track of the current state is often not enough--- need to add goals to decide which situations are good
- Deliberative instead of reactive

- May have to consider long sequences of possible actions before deciding if goal is achieved--- involves consideration of the future, "what will happen if I do...?"

Utility-based agents



Utility-Based Agent

- When there are multiple possible alternatives, how to decide which one is best?
- A goal specifies a crude distinction between a happy and unhappy state, but often need a more general performance measure that describes "degree of happiness"
- Utility function U: State --> Reals
  indicating a measure of success or happiness when at a given state
- Allows decisions comparing choice between conflicting goals, and choice between likelihood of success and importance of goal (if achievement is uncertain)

Learning agents