EDAP01: Artificial Intelligence
Agents (Chapter 2 of AIMA)

Jacek Malec
Dept. of Computer Science, Lund University, Sweden

January 20th, 2021

Plan for the 2nd hour

- What is an agent?
- PEAS (Performance measure, Environment, Actuators, Sensors)
- Agent architectures.
- Environments
- Multi-agent systems.

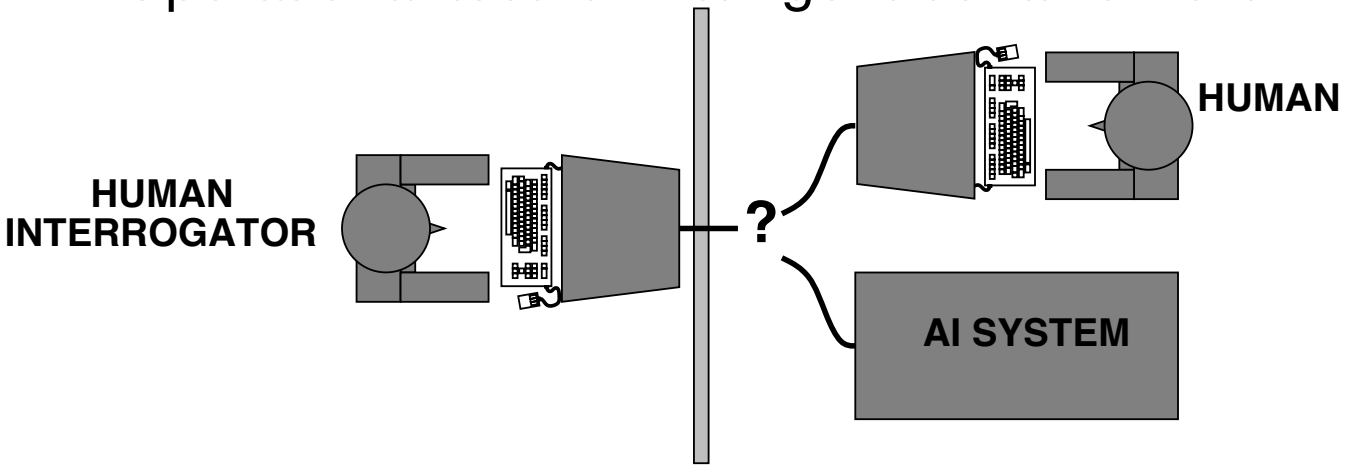
What is AI

Systems that think like humans	
Systems that act like humans	

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*



- Anticipated all major arguments against AI in last 50 years
- Suggested major components of AI: knowledge, reasoning, language understanding, learning

Problem: Turing test is not *reproducible*, *constructive*, or amenable to *mathematical analysis*

Thinking humanly: cognitive science

1960s "*cognitive revolution*": information-processing psychology replaced the then prevailing orthodoxy of *behaviorism*
Requires scientific theories of internal activities of the brain

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

Both approaches (roughly, *Cognitive Science* and *Cognitive Neuroscience*) are now distinct from AI
Both share with AI the following characteristic: *the available theories do not explain anything resembling human-level general intelligence*

What is AI

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

Thinking rationally: laws of thought

Aristotle: what are correct arguments/thought processes?

Several Greek schools developed various forms of **logic**:
notation and *rules of derivation* for thoughts;
may or may not have proceeded to the idea of mechanization

Direct line through mathematics and philosophy to modern AI

Problems:

- Not all intelligent behavior is mediated by logical deliberation
- What is the purpose of thinking? What thoughts *should* I have out of all the thoughts (logical or otherwise) that I *could* have?

Acting rationally

Rational behavior: doing the right thing

Acting rationally

Rational behavior: doing the right thing

The right thing:
that which is expected to **maximize goal achievement**,
given the **available information**

Acting rationally

Rational behavior: doing the right thing

The right thing:
that which is expected to **maximize goal achievement**,
given the **available information**

Doesn't necessarily involve thinking—e.g., blinking reflex—but thinking should be in the service of rational action

Aristotle (Nicomachean Ethics):

> *Every art and every inquiry, and similarly every action and pursuit, is thought to aim at some good*

Rational agents

An *agent* is an entity that perceives and acts

This course is about designing *rational agents*

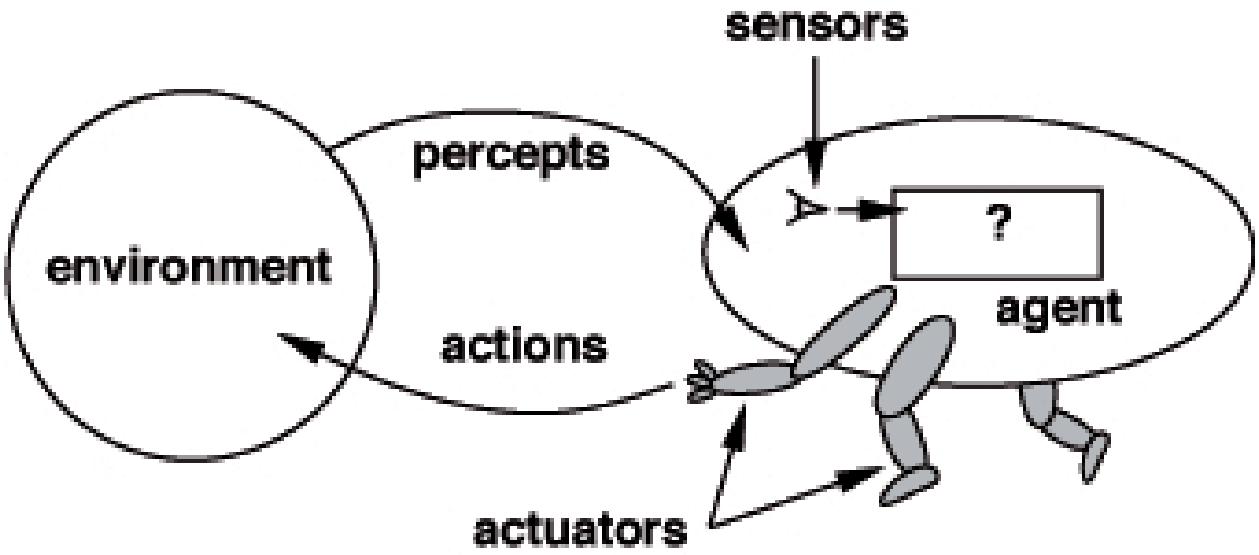
Abstractly, an agent is a function from percept histories to actions:

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

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*
$\rightarrow$ design best **program** for given machine resources

Agent



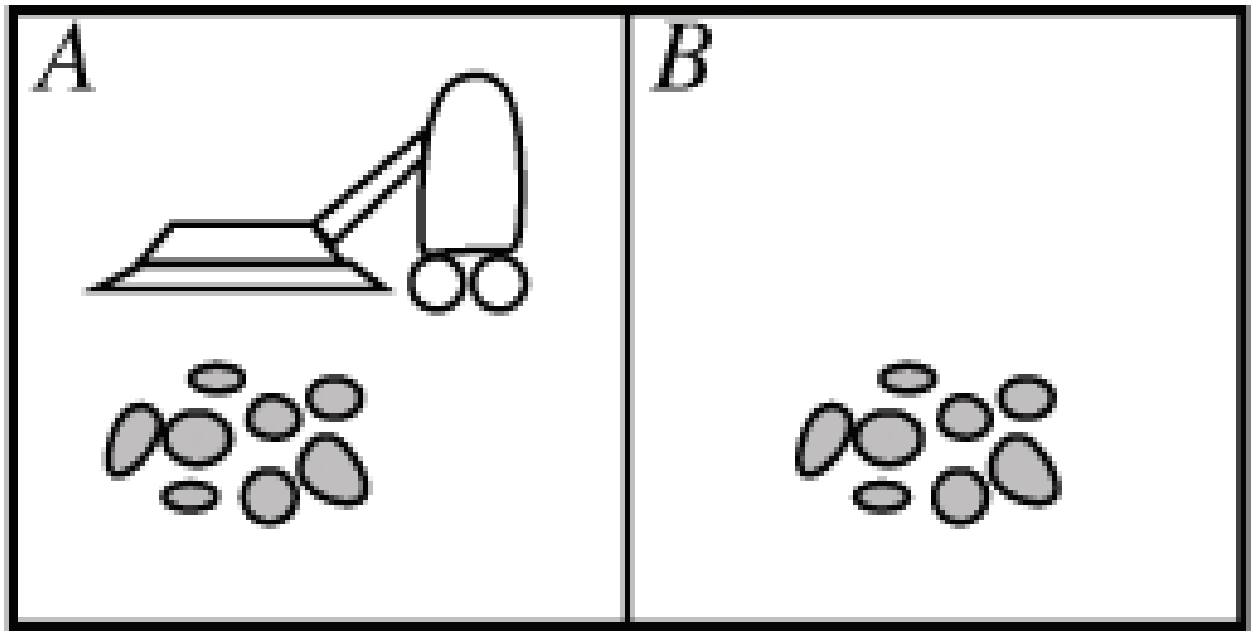
Agents include humans, robots, web-crawlers, thermostats, etc.

The *agent function* maps from percept histories to actions:

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

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

The vacuum-cleaning world



Percepts: location and contents, e.g. $< A, Dirty >$

Actions: *Left, Right, Suck, NoOp*

A vacuum-cleaning 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 vacuum-cleaning 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)
  if status == Dirty  then return Suck
  if location == A    then return Right
  if location == B    then return Left
```

A vacuum-cleaning 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)
  if status == Dirty  then return Suck
  if location == A    then return Right
  if location == B    then return Left
```

What is the *RIGHT* function?

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

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*

Rational is not omniscient
as percepts may not supply all relevant information

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*

Rational is not omniscient
as percepts may not supply all relevant information
Rational is not clairvoyant
as action outcomes may not be as expected

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*

Rational is not omniscient
as percepts may not supply all relevant information
Rational is not clairvoyant
as action outcomes may not be as expected
Hence, rational is **not necessarily successful**!

A rational agent

[Wooldridge, 2000]

An agent is said to be *rational* if it chooses to perform actions that are in its own best interests, given the beliefs it has about the world.

Properties of rational agents:

- Autonomy (they decide);
- Proactiveness (they try to achieve their goals);
- Reactivity (they react to changes in the environment);
- Social ability (they negotiate and cooperate with other agents).

PEAS

- PEAS: Performance measure, Environment, Actuators, Sensors
- Must first specify the setting for intelligent agent design
- Consider, e.g., the task of designing an automated taxi driver:
  - Performance measure
  - Environment
  - Actuators
  - Sensors

PEAS, example

AUTOMATED TAXI DRIVER:

- Performance measure: Safe, fast, legal, comfortable trip, maximize profits
- Environment: Roads, other traffic, pedestrians, customers
- Actuators: Steering, accelerator, brake, signal, horn
- Sensors: Cameras, radars, speedometer, GPS, odometer, engine sensors, car-human interface, car-car interface

Autonomous agents

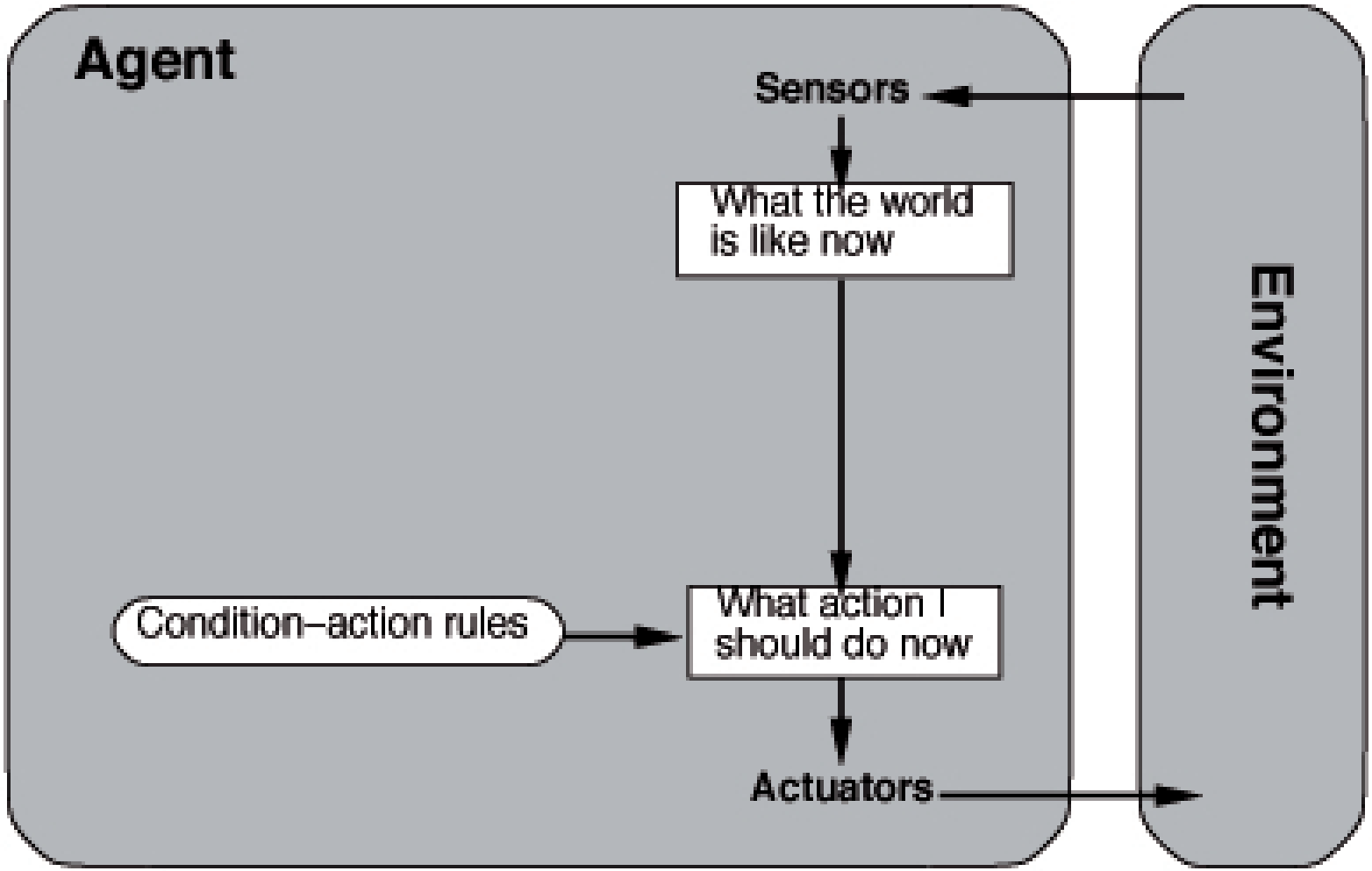
Can make decisions on their own.
Why do they need to? Because of the following properties of real environments (cf. Russell and Norvig):

- the real world is inaccessible (partially observable);
- the real world is nondeterministic (stochastic, sometimes strategic);
- the real world is nonepisodic (sequential);
- the real world is dynamic (non-static);
- the real world is continuous (non-discrete).

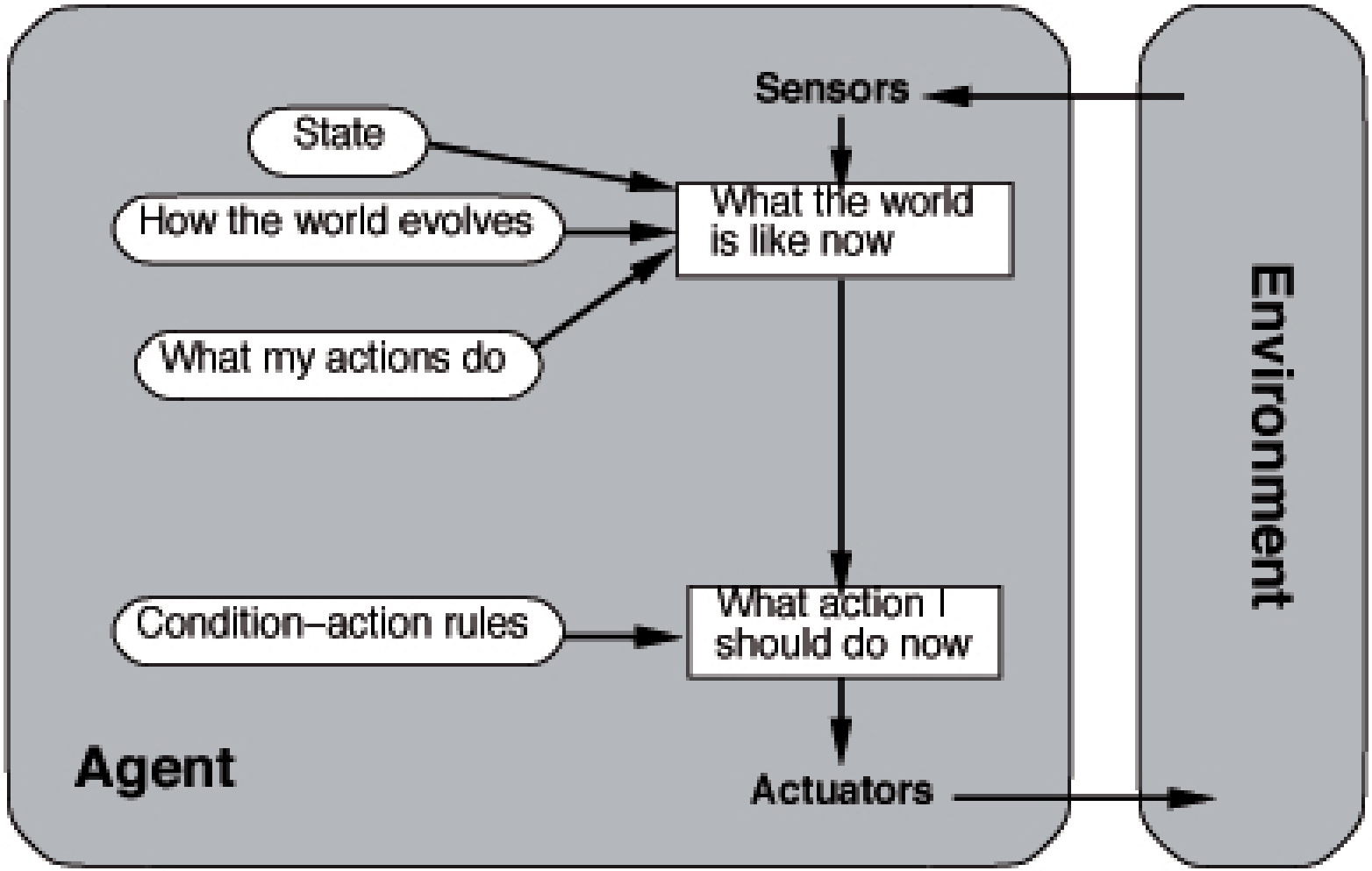
Agent taxonomy

- simple reflex agents
- reflex agents with state
- goal-based agents
- utility-based agents
-
- learning agents - independent property from the list above

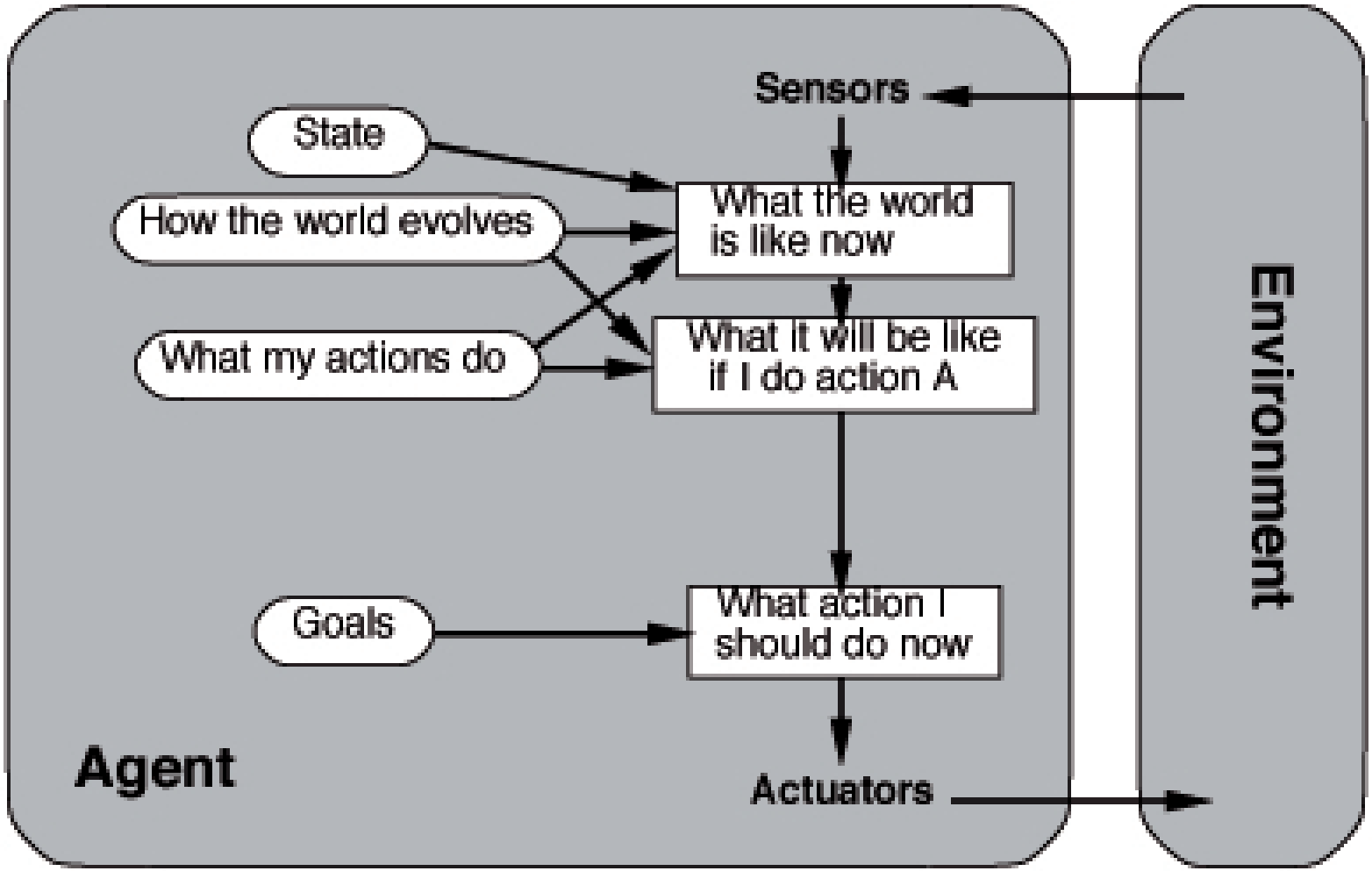
Simple reflex agent



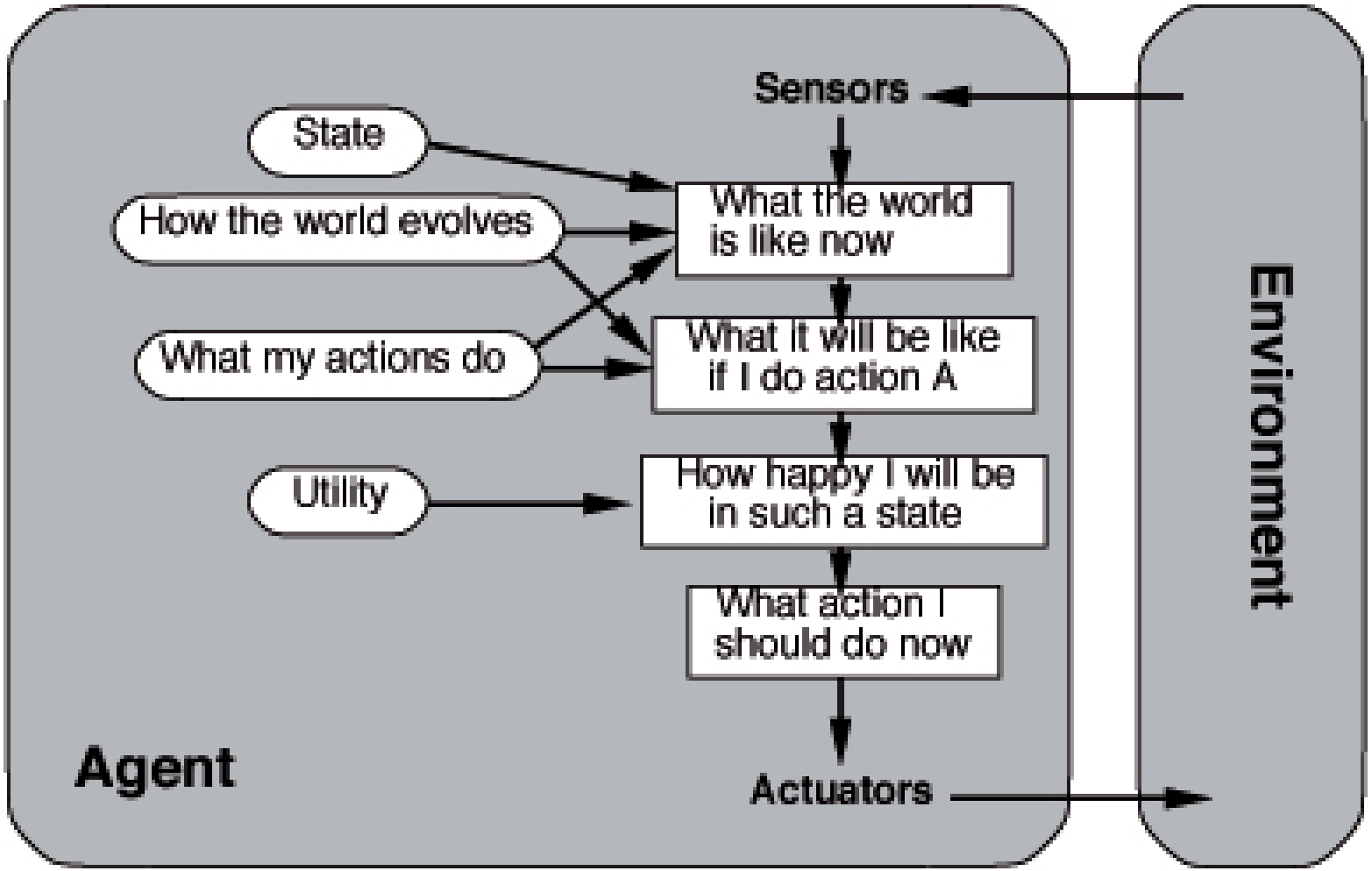
Reflex agent with state



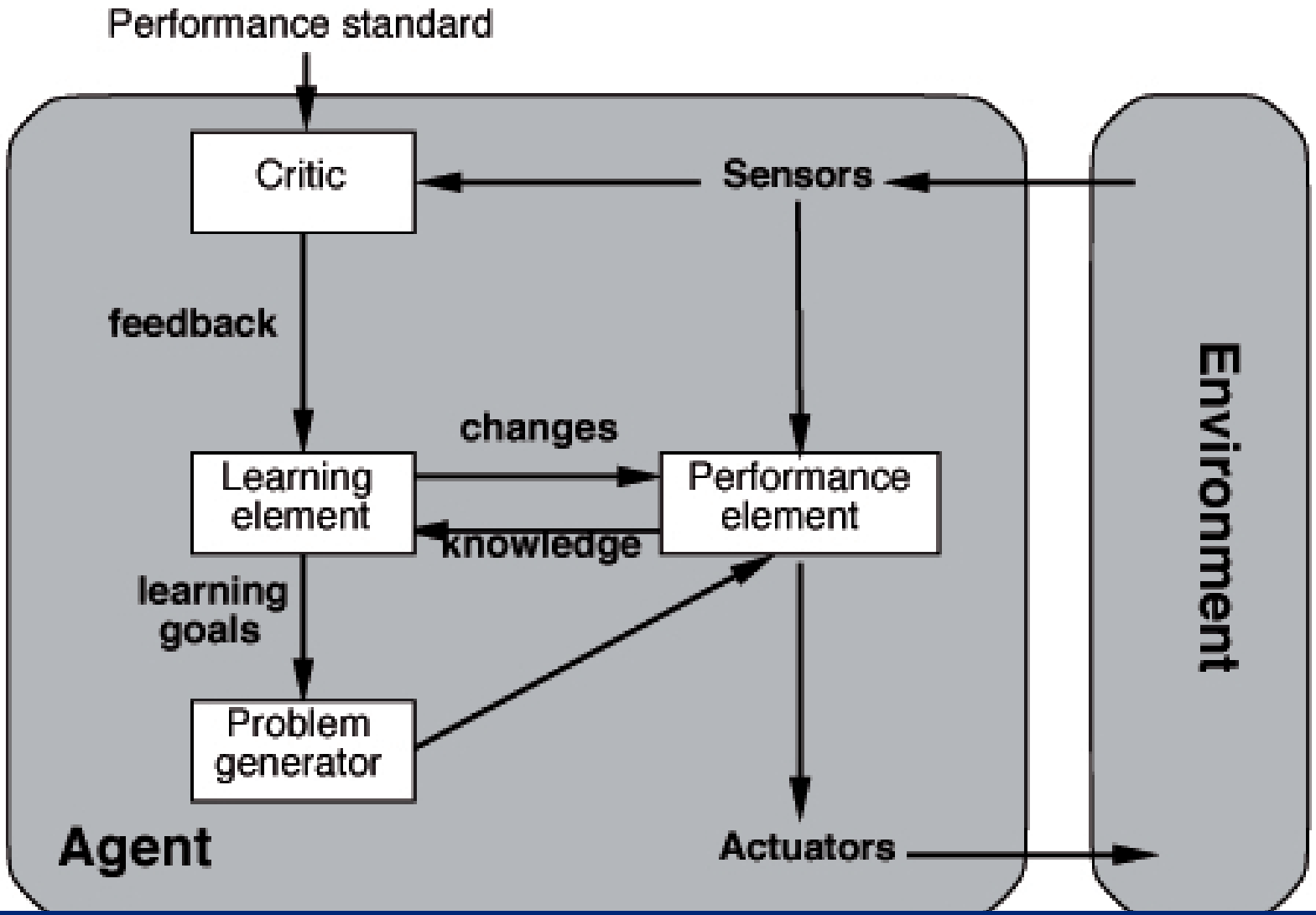
Goal-based agent



Utility-based agent



Learning agent



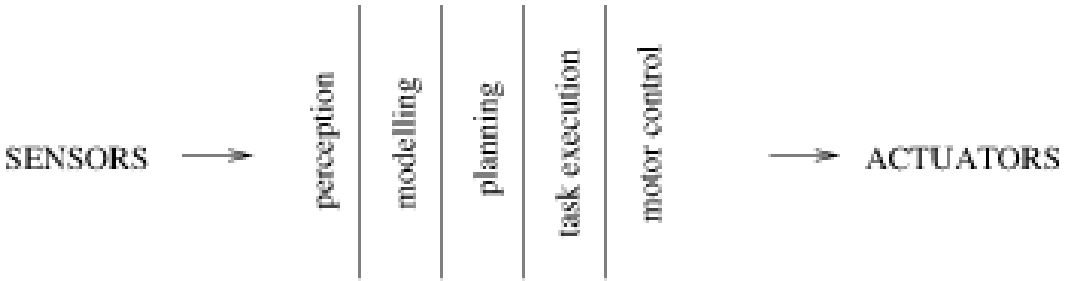
Rationality: John McCarthy 1956

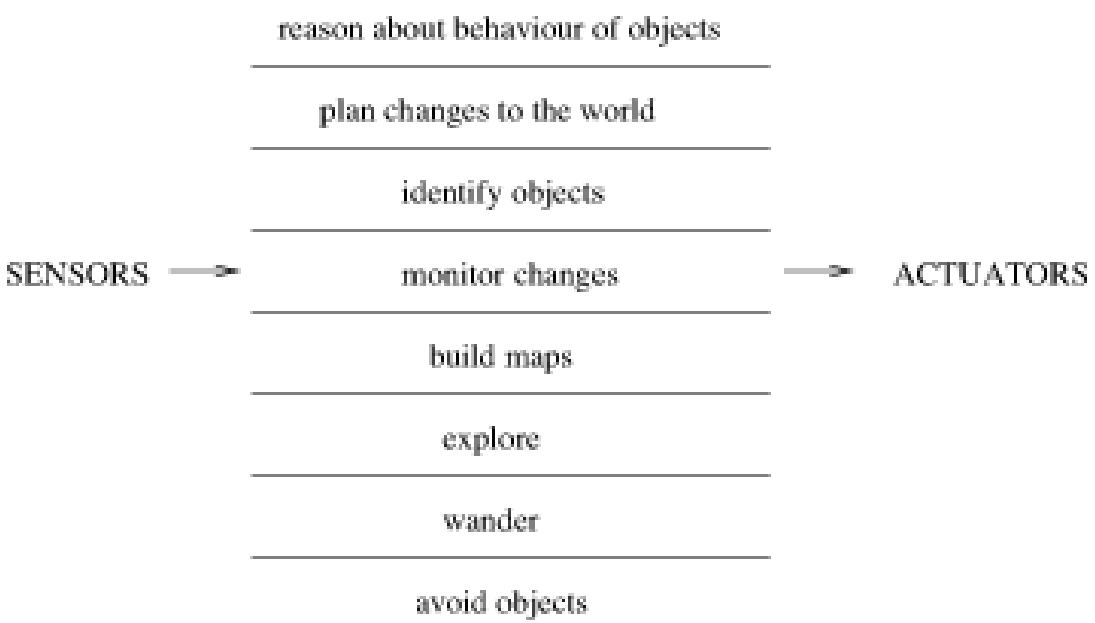
Rationality is a very powerful assumption.

It allows us to compute things we wouldn't otherwise be able to dream of!

30+ first years of AI were based solely on this assumption.

Subsumption: Rodney Brooks, 1985





Physical Grounding Hypothesis

- situatedness
  "the world is its own best model"
- embodiment
- intelligence
  "intelligence is determined by the dynamics of interaction with the world"
- emergence
  "intelligence is in the eye of the observer"

Summary

Agents interact with *environments* through *actuators and sensors*

Summary

Agents interact with *environments* through *actuators and sensors*

The *agent function* describes what the agent does in all circumstances

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

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

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

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

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?

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