Artificial Intelligence

Intelligent Agents
Week 10

Artificial Intelligence, Week 10

Literature

"Artificial Intelligence — A Modern Approach" by Stuart Russel and Peter Norvig (Chapters 1 & 2 for this week)

"Artificial Intelligence in the 21st century" by Steven Lucci and Danny Kopec (Chapter 1)

Slides edited by Benjamin Inden based on slides by Stuart Russel,

http://aima.cs.berkeley.edu

To be sure of a good grade ...

♢ attend all lectures and practicals.

♢ do preparatory reading / rehearsal in Russel & Norvig and / or Lucci & Kopec.

♢ focus on the topic while you are here (you may use books, laptops, websites etc. as long as they are related to the topic of the lecture).

♢ do the practical exercises (only that way you will learn the concepts).

♢ spend about 6 hours 40 minutes per week on the module overall.

Keeping in touch

- Your module team is there to help you
- Use means of contact (email, office hours) early if you get stuck
- Check NOW regularly for updates
- Engage in lectures and practical sessions (don't be afraid of making mistakes)

Outline

- ♢ What is AI?
- ♢ A brief history
- ♢ The state of the art
- ♢ Types of agents and environments

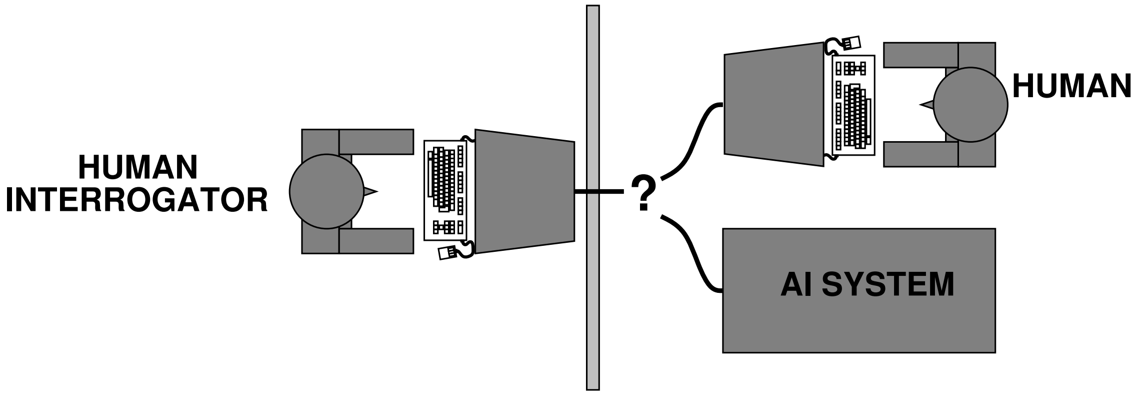
What is AI?

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

Acting humanly: The Turing test

Turing (1950) "Computing machinery and intelligence":

- ♢ "Can machines think?" ⟶ "Can machines behave intelligently?"
- ♢ Operational test for intelligent behavior: the Imitation Game



- ♢ Predicted that by 2000, a machine might have a 30% chance of fooling a lay person for 5 minutes
- ♢ Suggested major components of AI: knowledge, reasoning, language understanding, learning

Thinking humanly: Cognitive Science

1960s "cognitive revolution": information-processing psychology replaced prevailing orthodoxy of behaviorism

Requires scientific theories of internal activities of the brain
  What level of abstraction? "Knowledge" or "circuits"?

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 (or engender) anything resembling human-level general intelligence

Hence, all three fields share one principal direction!

Thinking rationally: Laws of Thought

Normative (or prescriptive) rather than descriptive

Aristotle: what are correct arguments/thought processes?

Several Greek schools developed various forms of logic:
notation and **rules of derivation** for thoughts;

Direct line through mathematics and philosophy to modern AI

Problems:
1) Not all intelligent behavior is mediated by logical deliberation
2) 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

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

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:

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

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**

$\to$ design best program for given machine resources

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
Psychology	adaptation phenomena of perception and motor control experimental techniques (psychophysics, etc.)
Economics	formal theory of rational decisions
Linguistics	knowledge representation grammar
Neuroscience	plastic physical substrate for mental activity
Control theory	homeostatic systems, stability simple optimal agent designs

Potted history of AI

1943	McCulloch & Pitts: Boolean circuit model of brain
1950	Turing's "Computing Machinery and Intelligence"
1952–69	Look, Ma, no hands!
1950s	Early AI programs, including Samuel's checkers program, Newell & Simon's Logic Theorist, Gelernter's Geometry Engine
1956	Dartmouth meeting: "Artificial Intelligence" adopted
1965	Robinson's complete algorithm for logical reasoning
1966–74	AI discovers computational complexity Neural network research almost disappears
1969–79	Early development of knowledge-based systems
1980–88	Expert systems industry booms
1988–93	Expert systems industry busts: "AI Winter"
1985–95	Neural networks return to popularity
1988–	Resurgence of probability; general increase in technical depth "Nouvelle AI": ALife, GAs, soft computing
1995–	Agents, agents, everywhere . . .
2003–	Human-level AI back on the agenda
2012–	Deep learning and "Big data" create the biggest AI hype yet

State of the art

Which of the following can be done at present?

- ♢ Play a decent game of table tennis
- ♢ Drive safely along a curving mountain road
- ♢ Drive safely along Southchurch Drive
- ♢ Buy a week's worth of groceries on the web
- ♢ Buy a week's worth of groceries at a local supermarket
- ♢ Play a decent game of bridge
- ♢ Discover and prove a new mathematical theorem
- ♢ Design and execute a research program in molecular biology
- ♢ Write an intentionally funny story
- ♢ Give competent legal advice in a specialized area of law
- ♢ Translate spoken English into spoken Swedish in real time
- ♢ Converse successfully with another person for an hour
- ♢ Perform a complex surgical operation
- ♢ Unload any dishwasher and put everything away

Examples: Mars Exploration Rover

Examples: Roomba

Examples: Packbot

Examples: Stanley

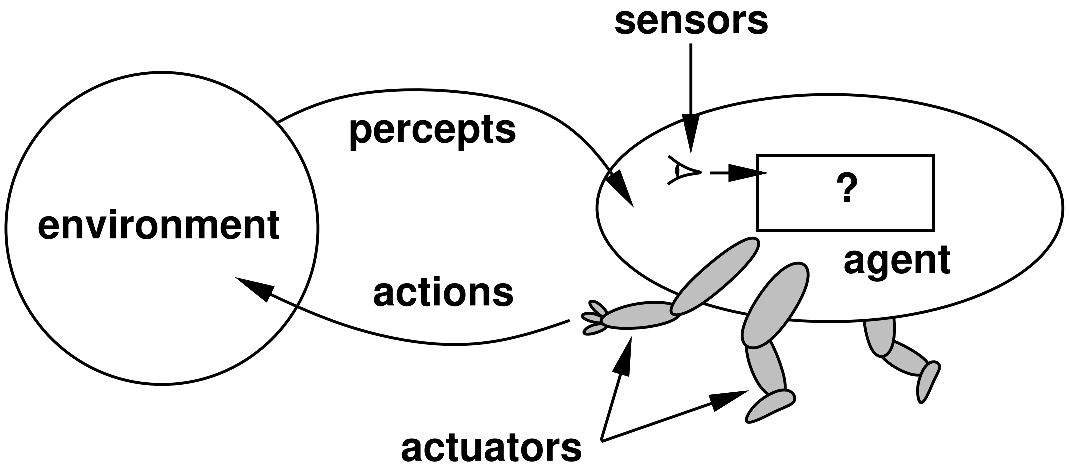
Examples: Watson

Agents

♢ Agents are semi-autonomous computer programs that intelligently assist the user with computer applications.

♢ Agents employ artificial intelligence techniques to assist users with daily computer tasks, such as reading electronic mail, maintaining a calendar, and filing information.

Agents and environments: a more general view



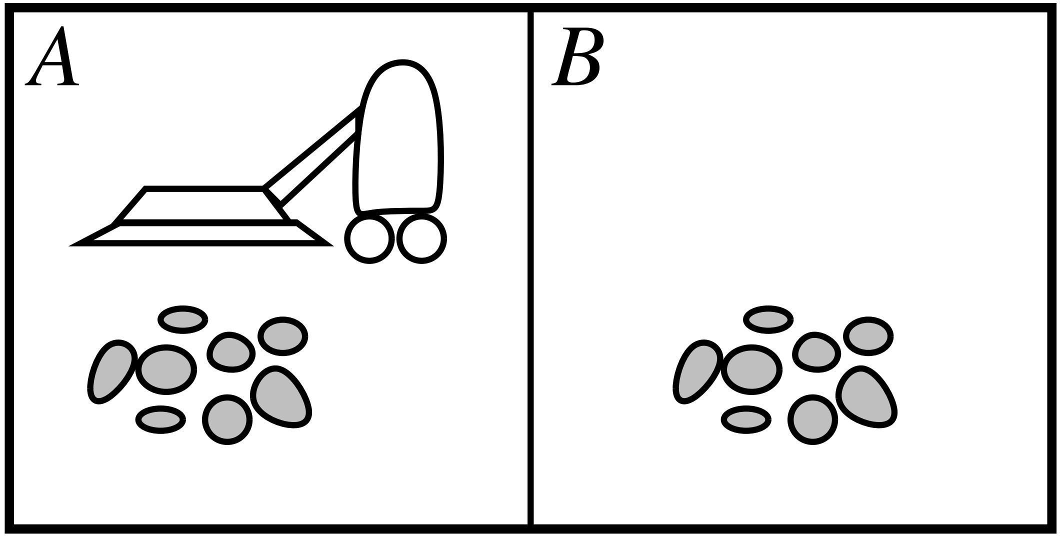
An Agent is anything that can be viewed as perceiving its environment through sensors and acting upon that environment through actuators. Agents include humans, robots, softbots, thermostats, etc.

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

An example: The 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*

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 $\neq$ omniscient

– percepts may not supply all relevant information

Rational $\neq$ clairvoyant

– action outcomes may not be as expected

Hence, rational $\neq$ successful

The PEAS approach

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

Consider, e.g., the task of designing an automated taxi:

Performance measure??

Environment??

Actuators??

Sensors??

The PEAS approach

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, . . .

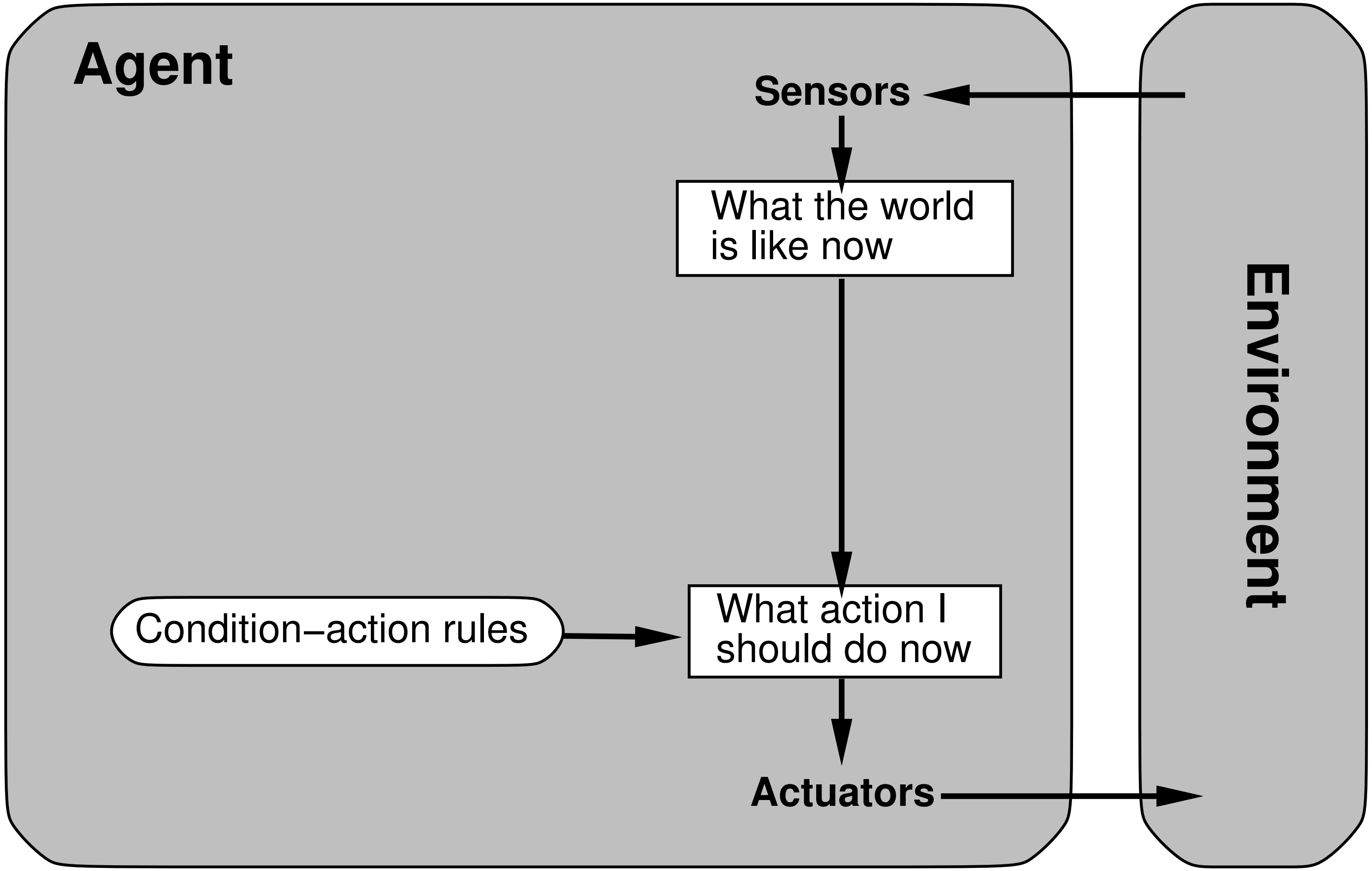
Sensors?? video, accelerometers, gauges, engine sensors, keyboard, GPS, . . .

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

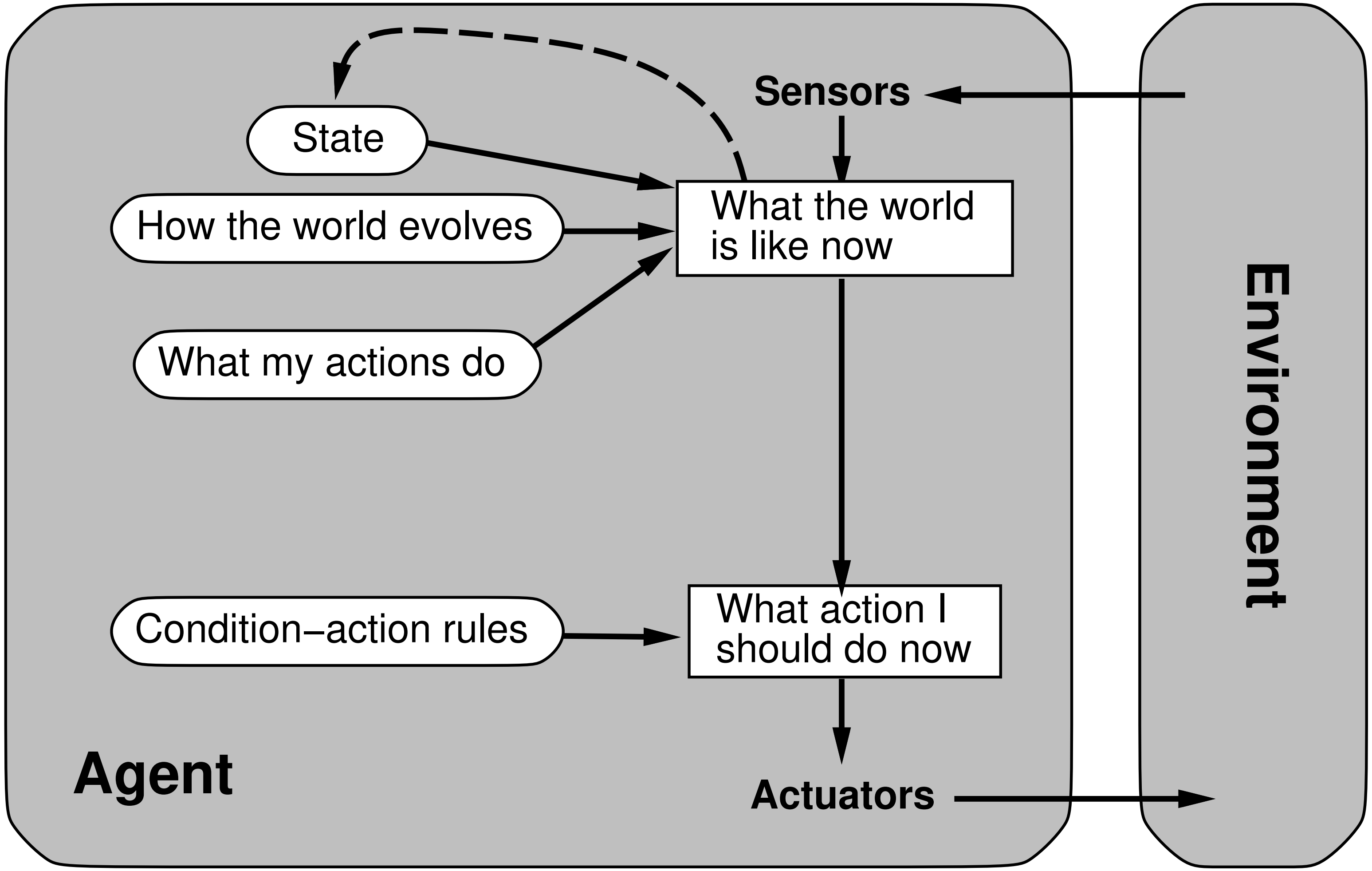
Simple reflex agents



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
```

Reflex agents with state

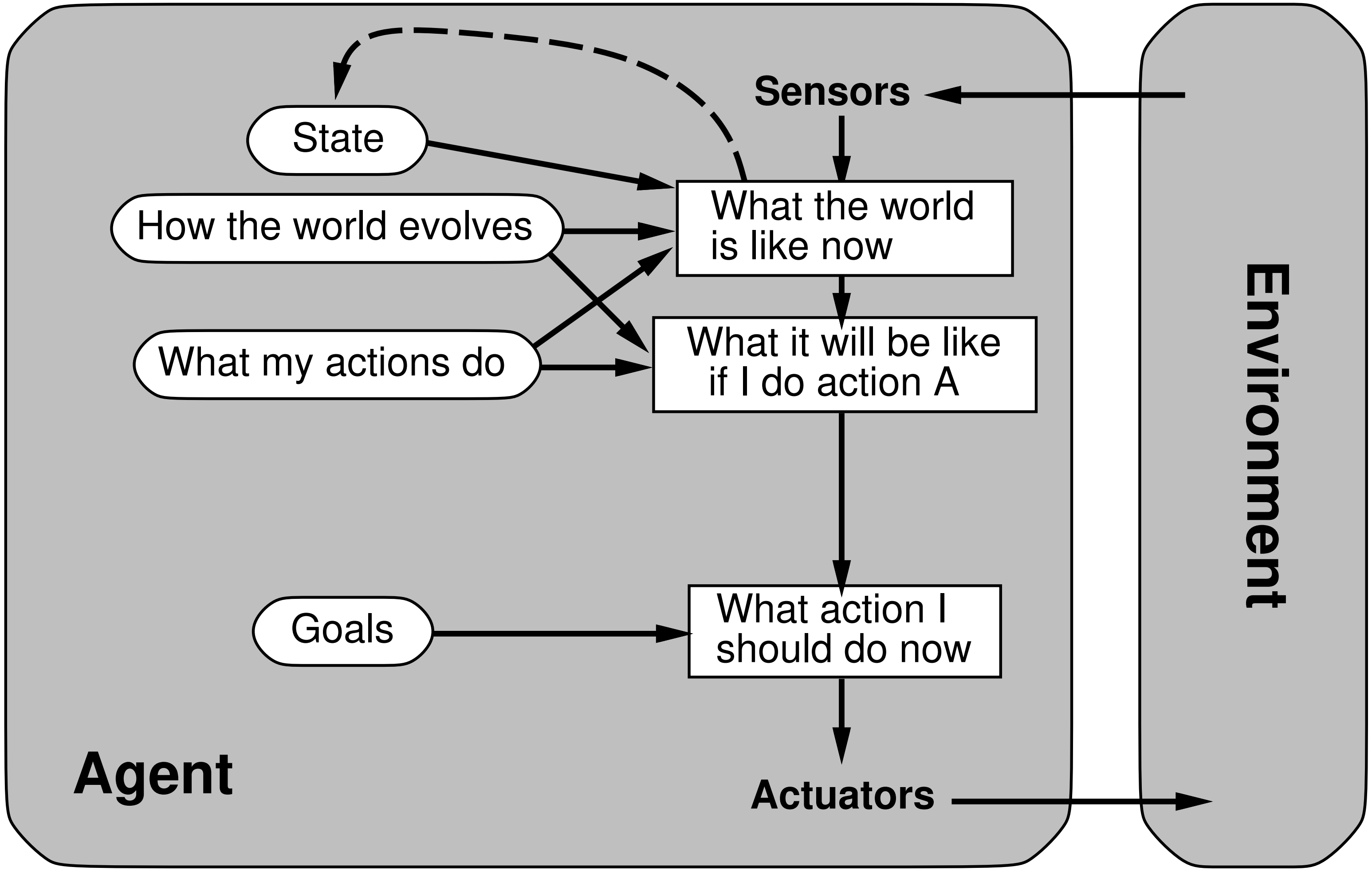


Example

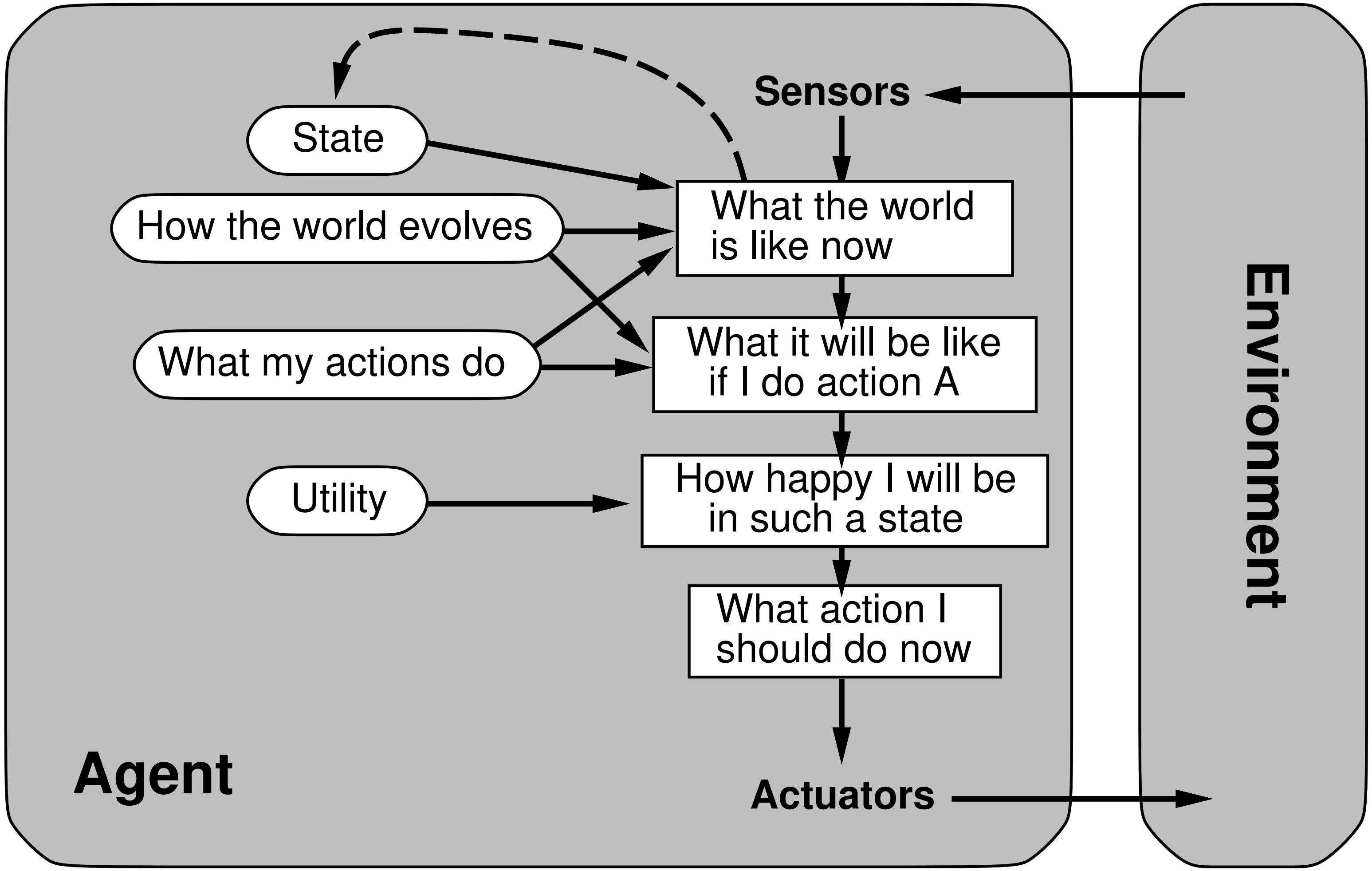
function REFLEX-VACUUM-AGENT([*location*,*status*]) **returns** an action
static: $last\_A,\ last\_B$, numbers, initially ∞

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

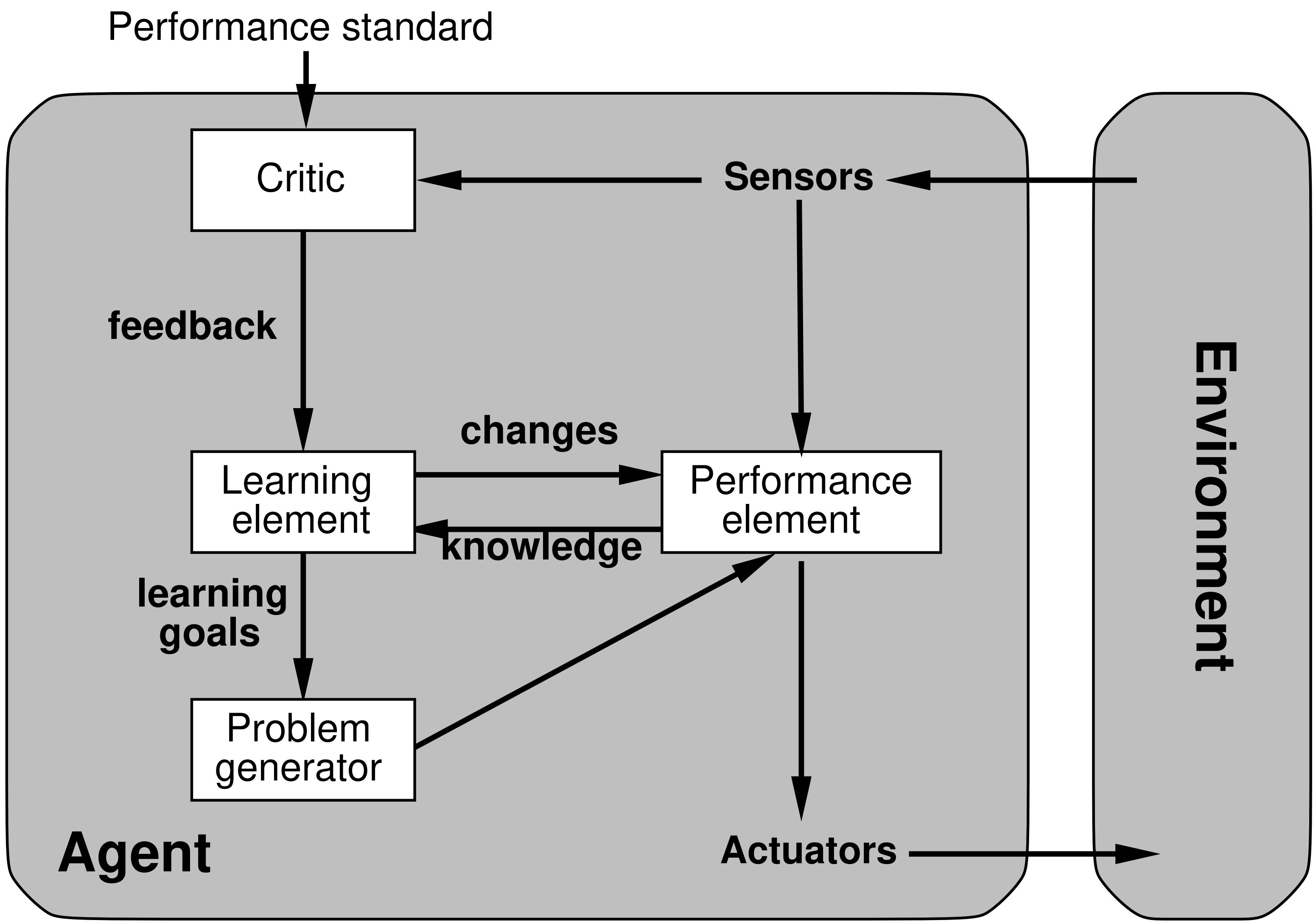
Goal-based agents



Utility-based agents



Learning agents



Summary

Artificial intelligence can be described as systems that think / act like humans / rationally.

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

Several basic agent architectures exist:
reflex, reflex with state, goal-based, utility-based