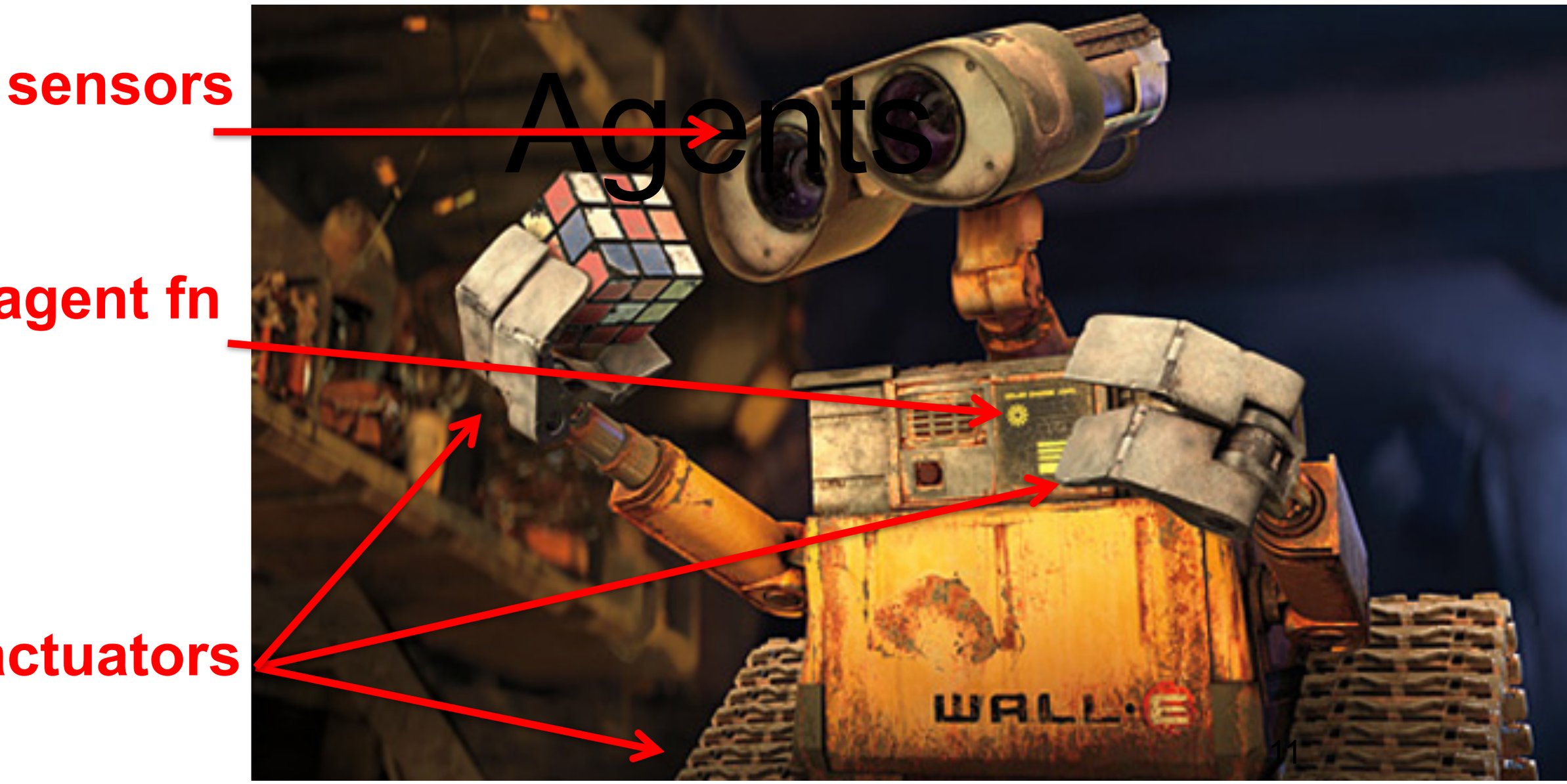
Intelligent Agents

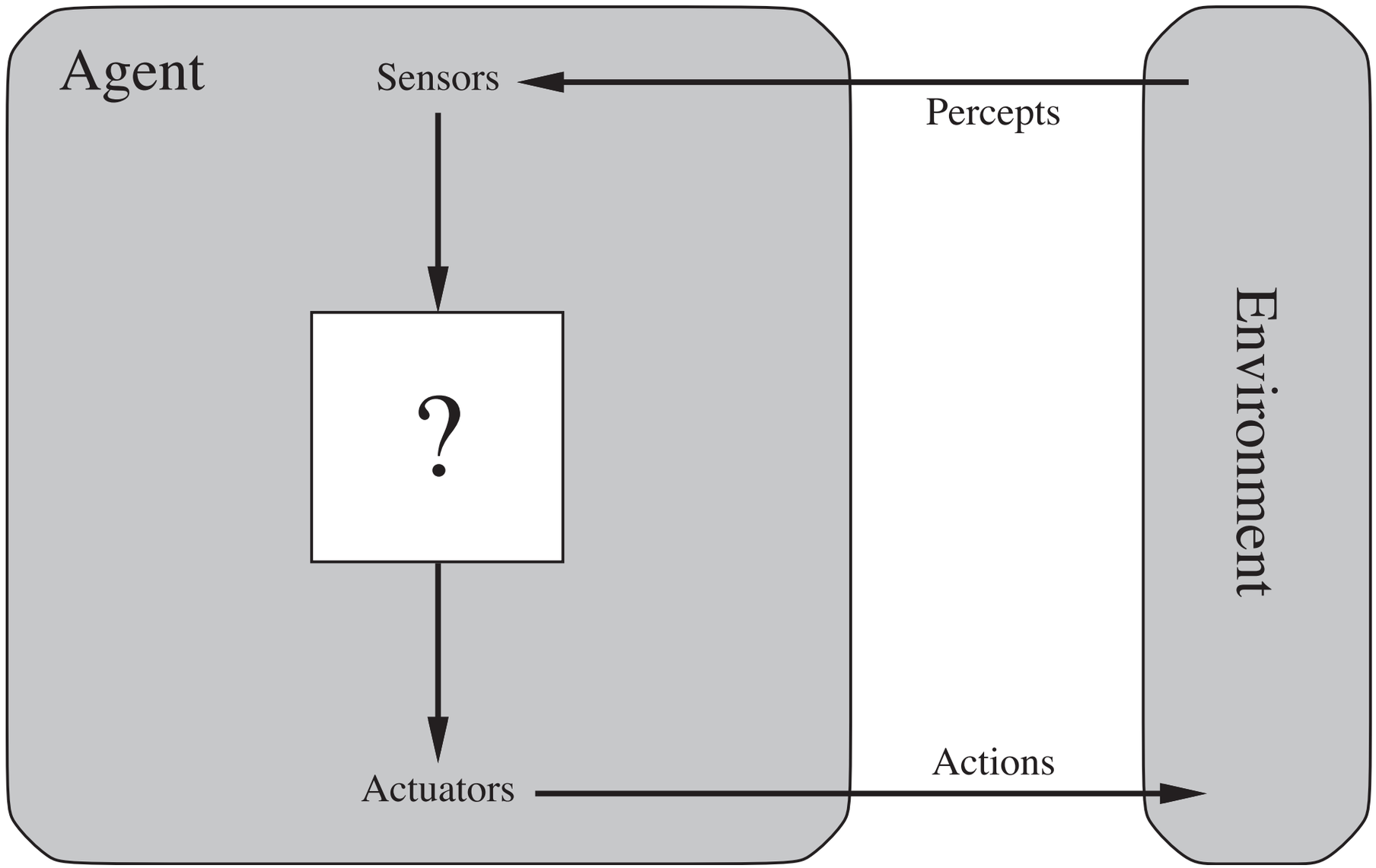
Chapter 2

AI as Agents

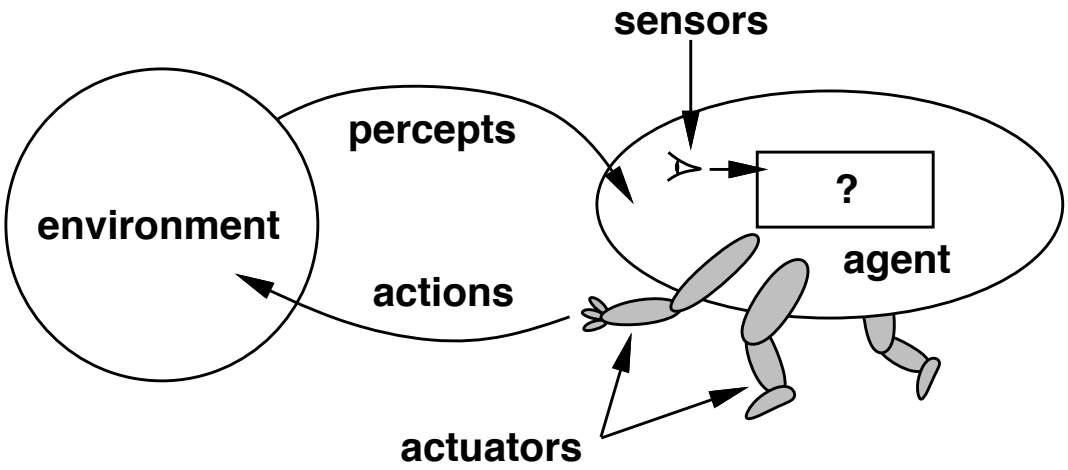
***act* = AgentFn(*percept*)**



Agents and Environments



Agents and environments



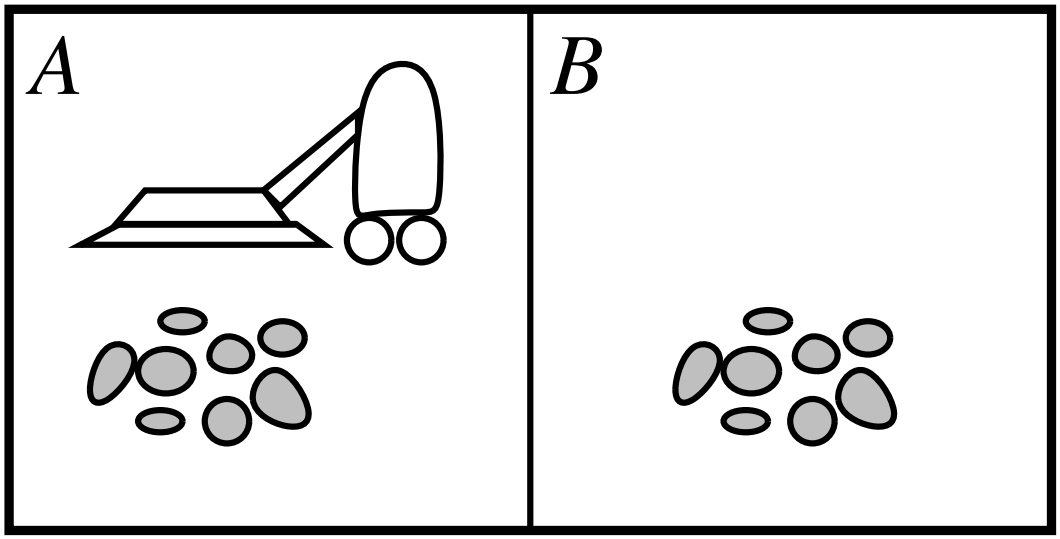
Agents include humans, robots, softbots, thermostats, etc.

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

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?

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

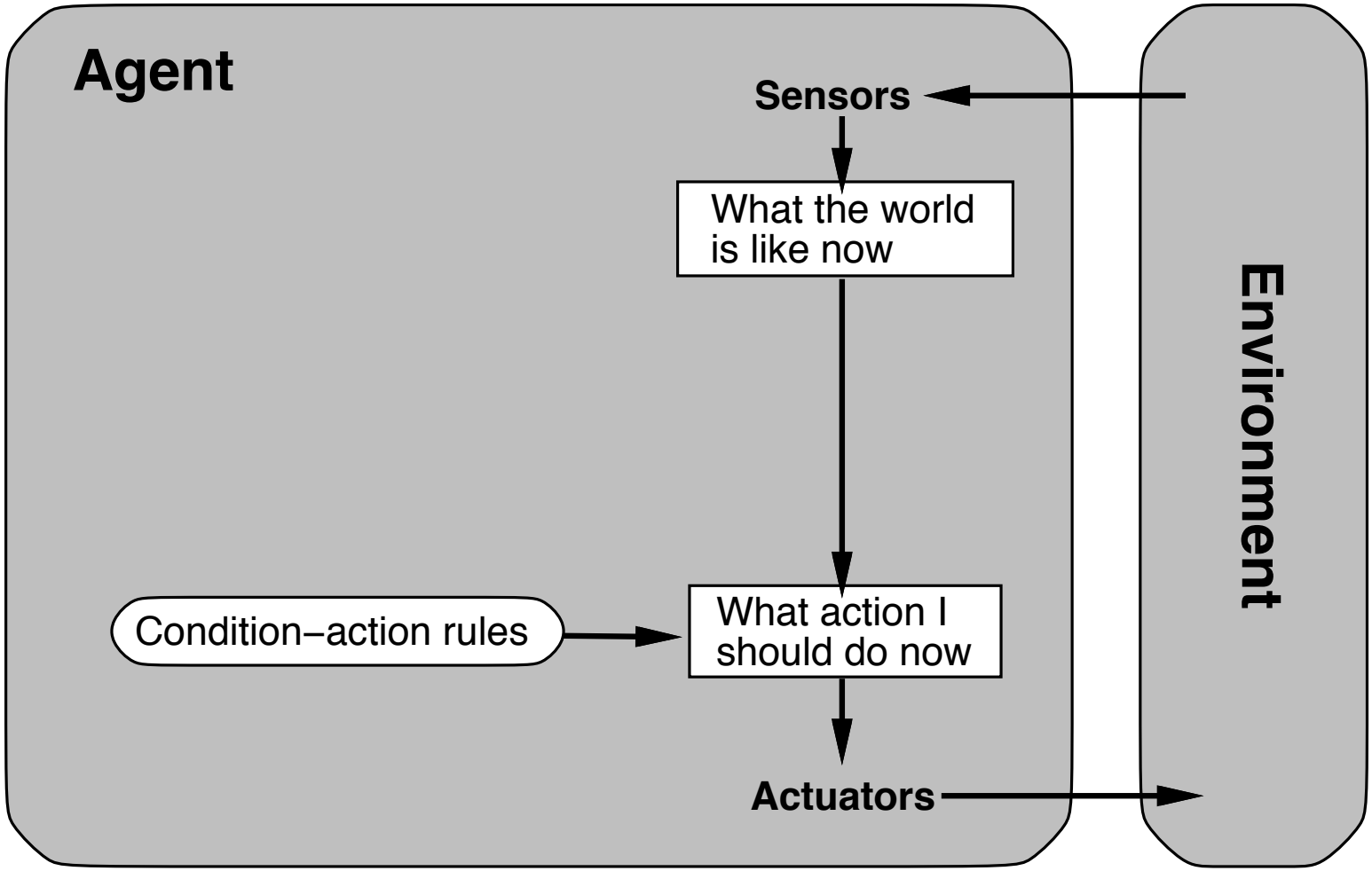
Hence, rational $\neq$ successful

Rational $\Rightarrow$ exploration, learning, autonomy

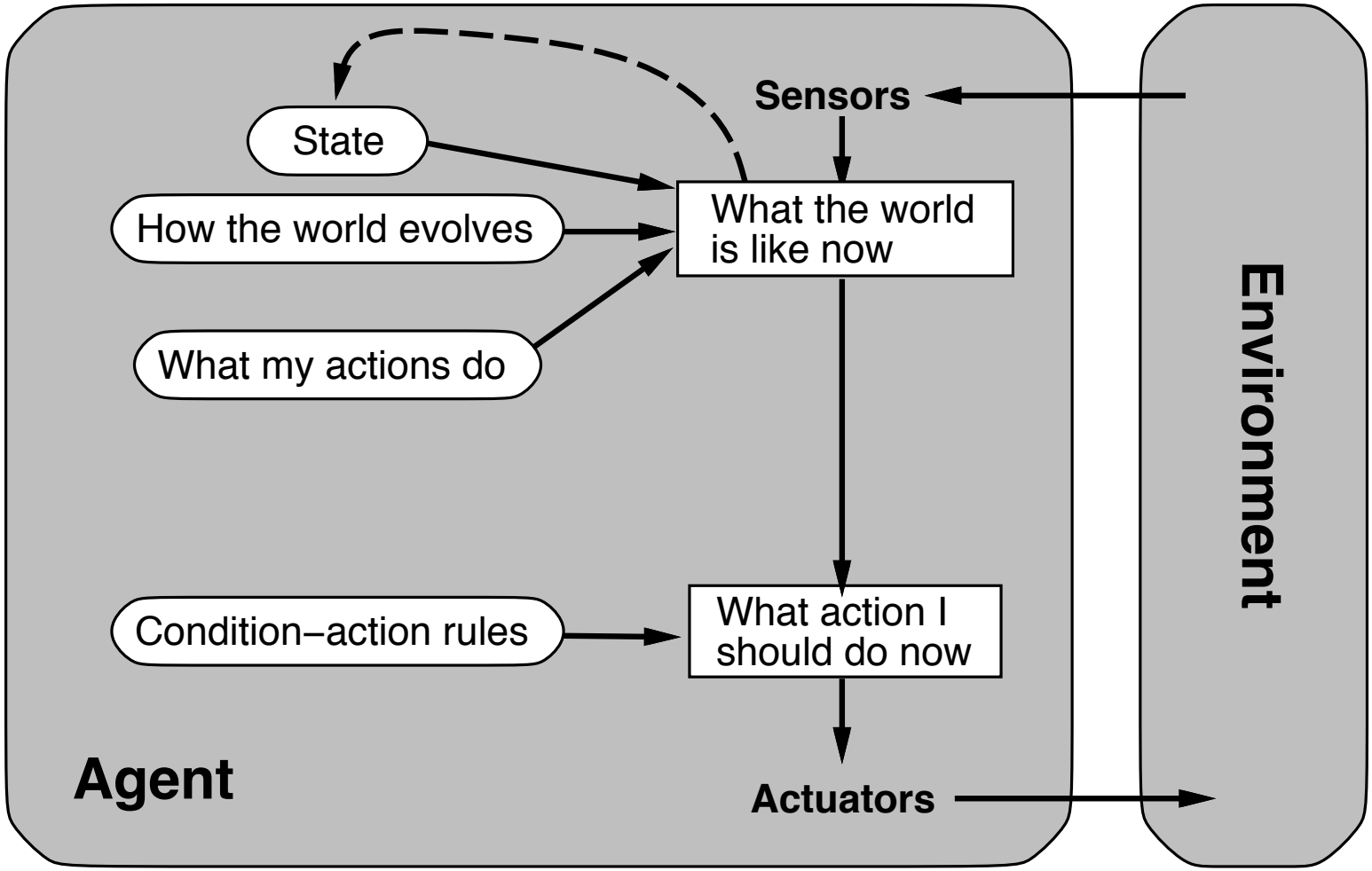
What choices do we have for building an agent program *f* that, given percept *p*, current state *s* and model of the world *m* chooses action *a*?

- Reflex: $a = f(p)$
- History-based: $a = f([p_1, p_2, \ldots, p_n,])$
- Model-based: $a = f_m(p)$ and may update m
- Goal-based agents
  $a = f_m(p, s)$ where a best achieves goal
- Utility-based agents
  $a = f_m(p, s)$ where a maximizes expected utility

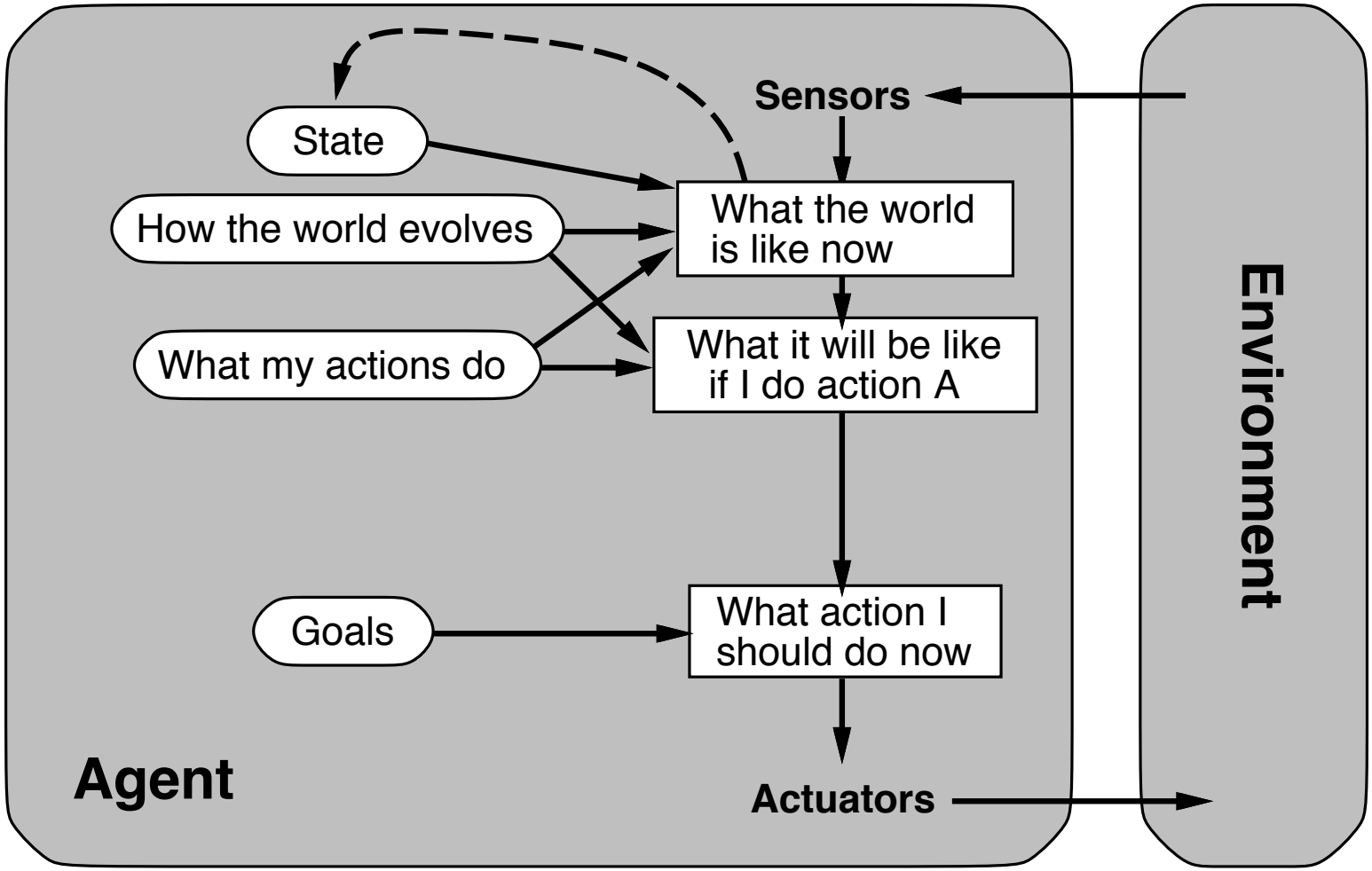
Simple reflex agents



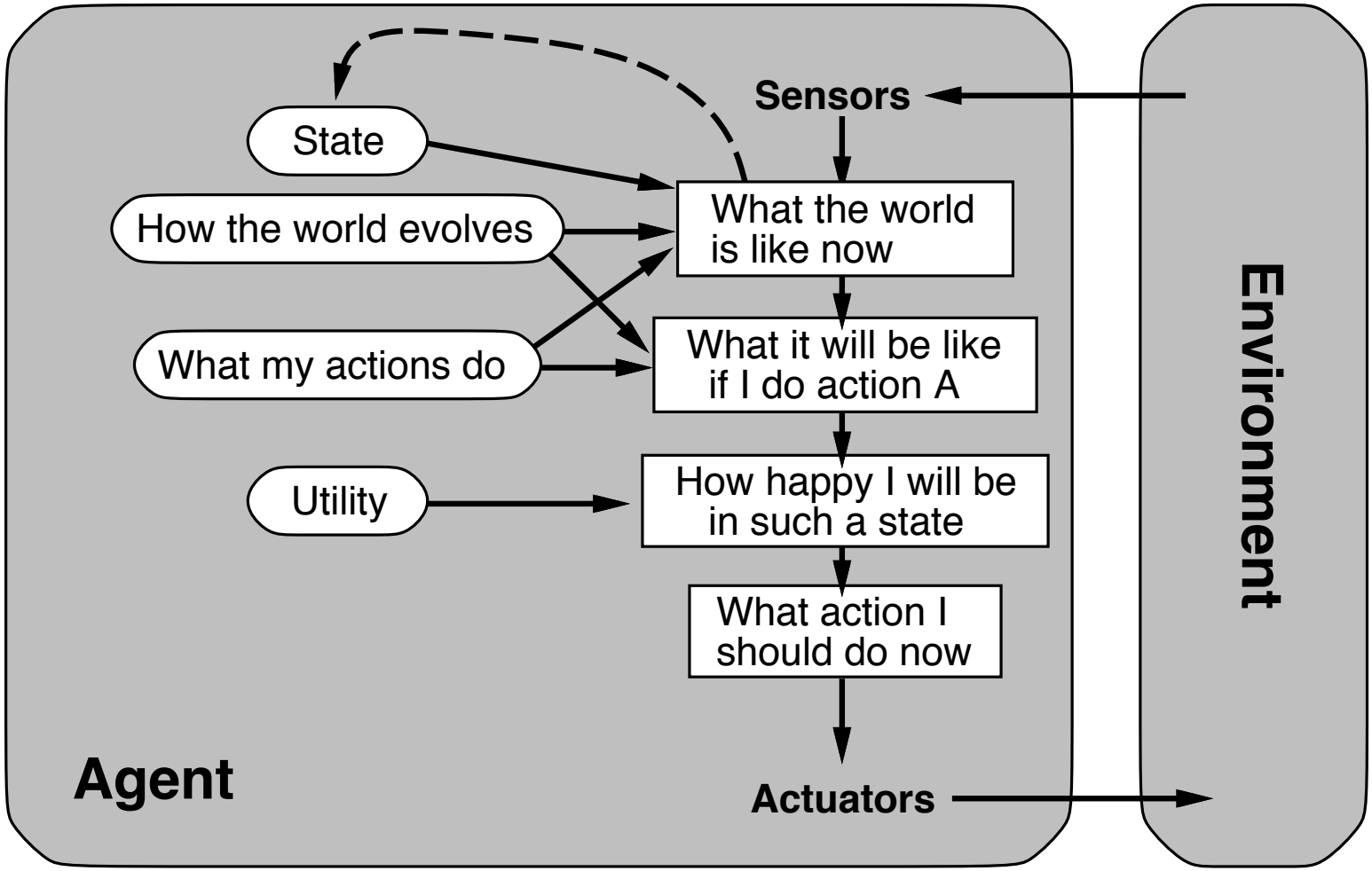
Reflex agents with state



Goal-based agents



Utility-based agents



Learning agents