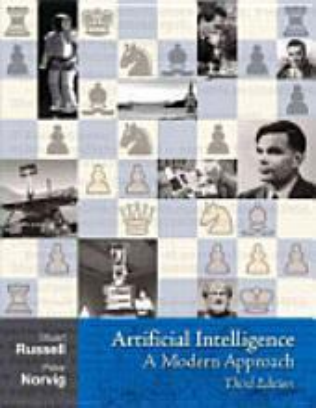


CS371/671 Artificial Intelligence

Professor: Sriraam Natarajan

Email: nataras@wfu.edu

Class Hours: TR 2:00 - 3:15
Office Hours TR 1:00 - 2:00 and by appointment

Textbook

Russell and Norvig, *Artificial Intelligence - A Modern Approach*, Third Edition, Prentice Hall.

Slides

Most of the lectures are based on **AI book slides**. I will edit/modify as needed. You can also refer to the slides at http://aima.eecs.berkeley.edu/slides-pdf/

Tentative Plan

5 Homeworks, 2 Programming Assignments, 1-2 Midterm, 1 Final, 1 Final Project (For graduate students)

What is AI?

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

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"?
- 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

Both share with AI the following characteristic:

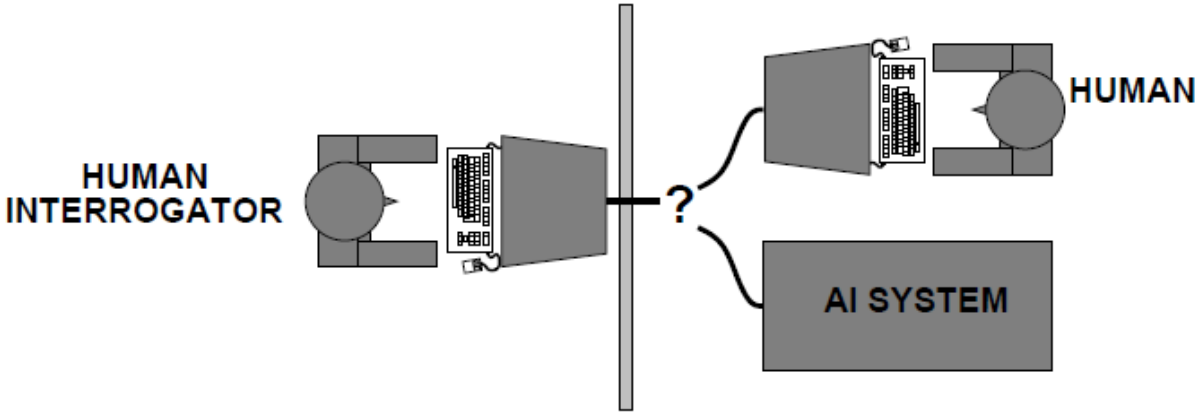
the available theories do not explain (or engender) anything resembling human-level general intelligence

Hence, all three fields share one principal direction!

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



- Predicted that by 2000, a machine might have a 30% chance of fooling a lay person for 5 minutes
- Anticipated all major arguments against AI in following 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 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;
may or may not have proceeded to the idea of mechanization

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

Aristotle (Nicomachean Ethics):

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

State of the art

What can be done?
- Play Chess
- Play Jeopardy
- Play a decent game of table tennis
- Clean a room
- Buy items from the web
- Search web automatically to provide recommendations
- Recognize and understand human speech
- Autonomously drive through a desert
- Autonomously drive in north California
- Provide “specialized” advice
- …

What cannot be done?
- Play baseball
- Write poems
- Play tennis
- Clean a room with obstacles, children and pets
- Buy items from local grocery store
- Full personalization
- Understand Intentions
- Autonomously drive in Boston
- Cross a busy road in India
- Diagnose and cure patients
- …

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:

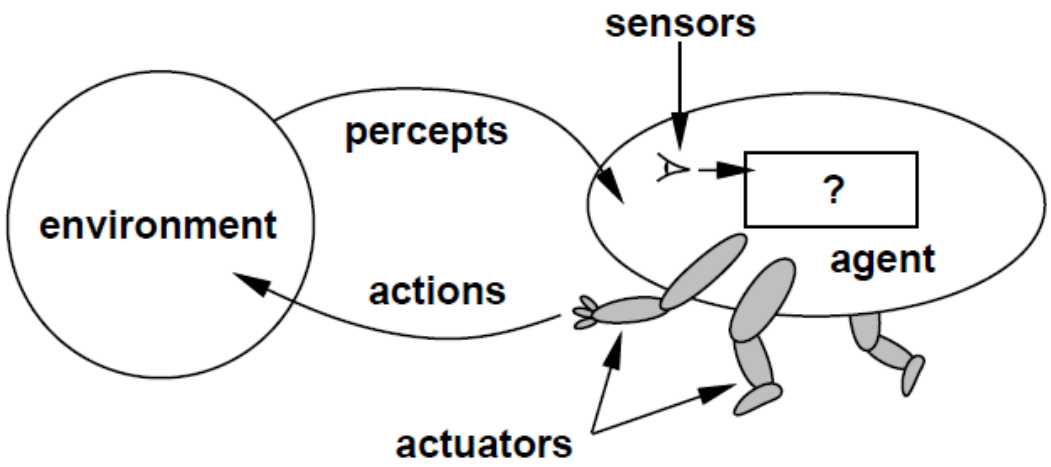
$$f : \mathcal{P}^* \rightarrow \mathcal{A}$$

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

Agents and environments



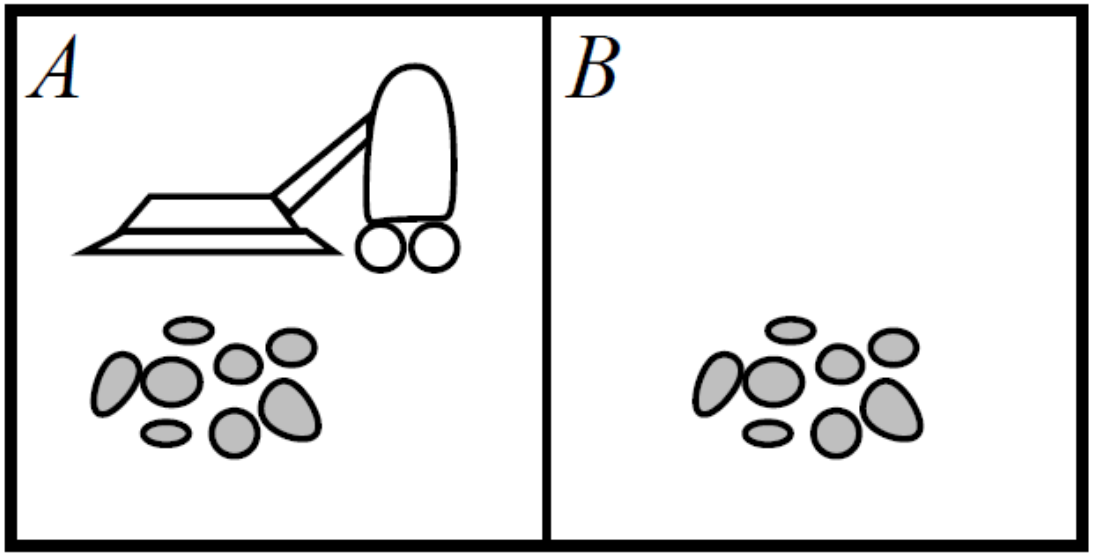
Agents include humans, robots, softbots, thermostats, etc.

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

PEAS

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??

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)

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

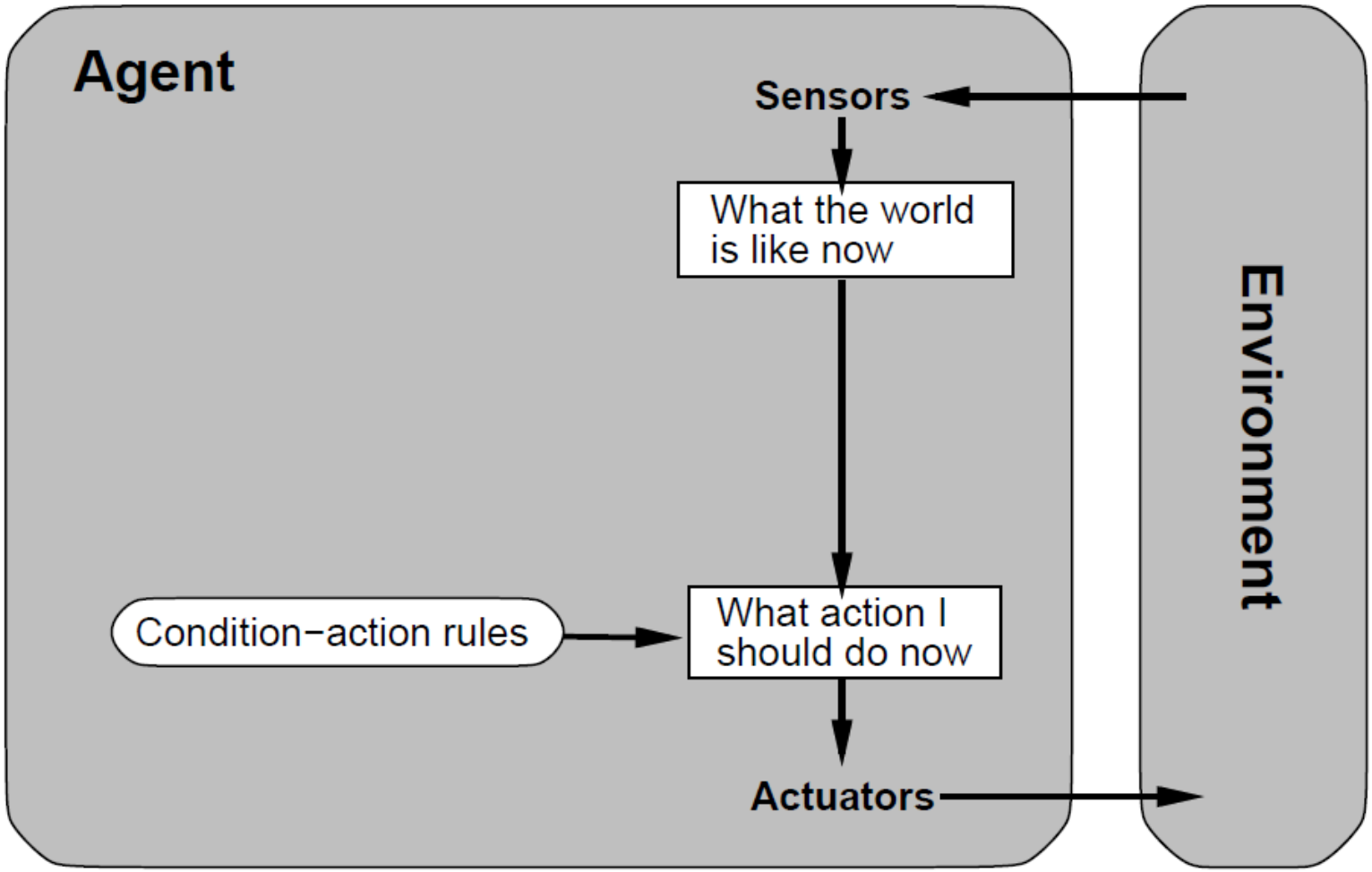
Agent types

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

Simple reflex agents



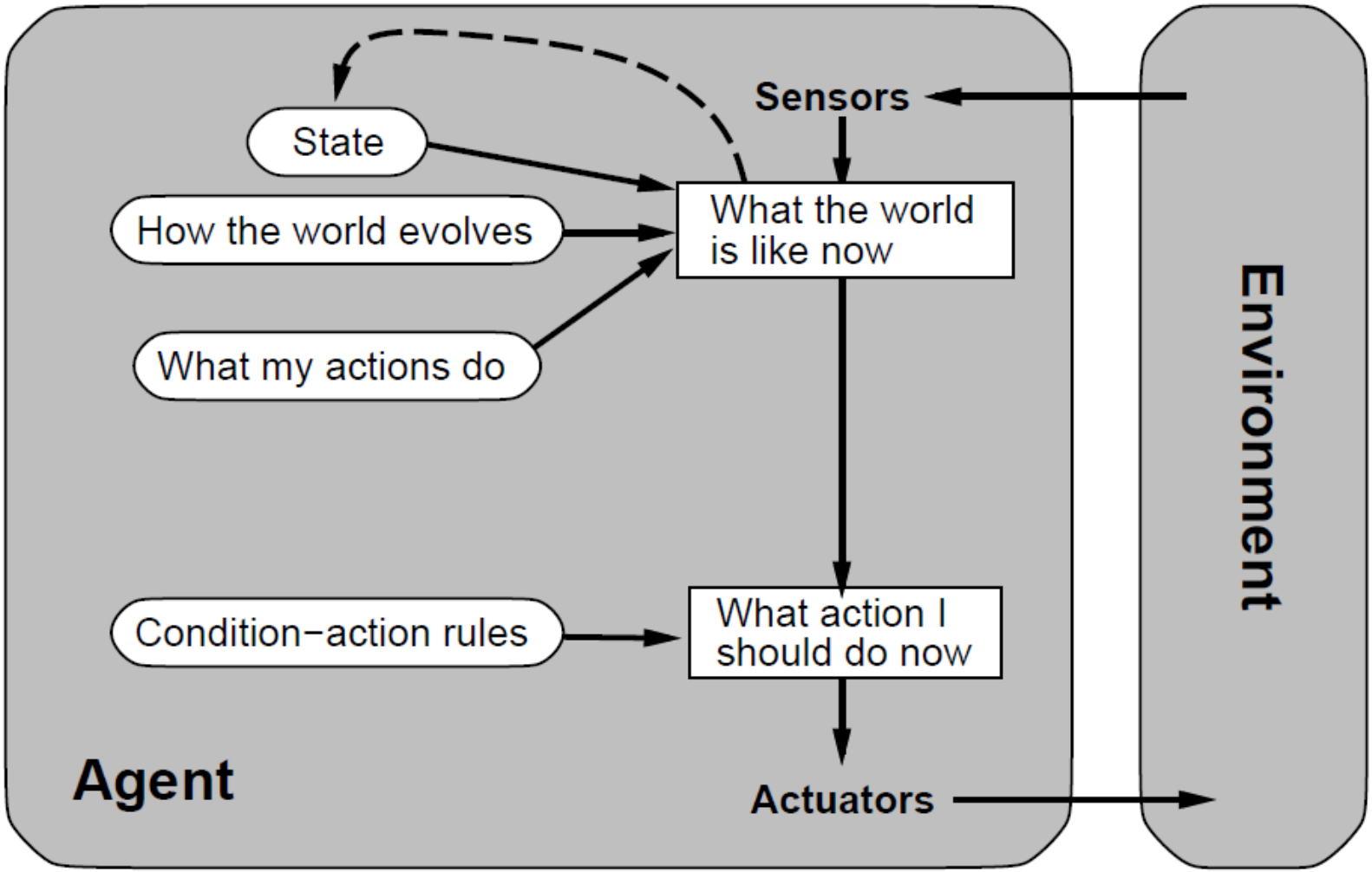
Example Function

function Simple_Reflex_Vaccum_Agent([*location,roomStatus]*) **returns** action

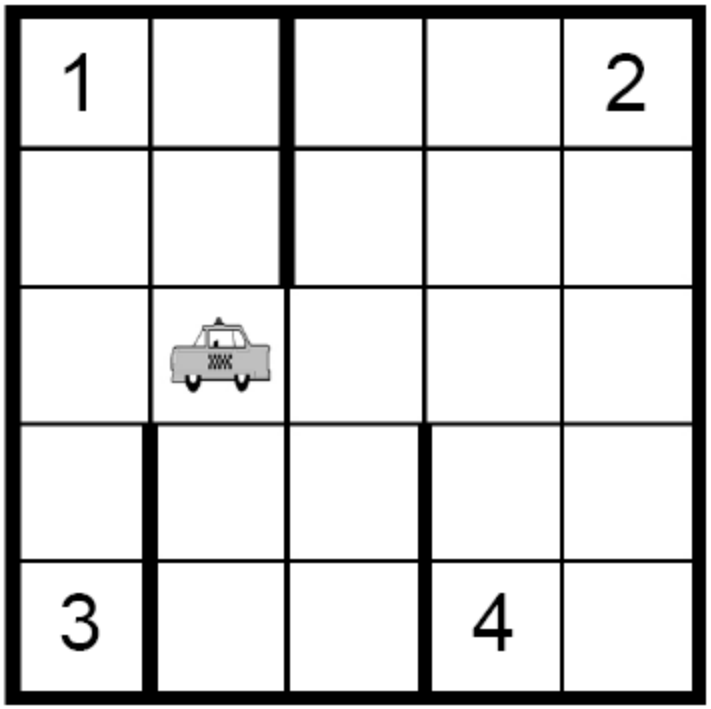
if *roomStatus == Dirty* **then return** *Clean*
else if *location == A* **then return** *MoveRight*
else if *location == B* **then return** *MoveLeft*

The input is a sequence of location and the status of the room and the output is the corresponding sequence of actions. The location and status are simple sensory information.

Reflex agents with state



A More Sophisticated Domain



- A more complex example – Taxi domain – Goal is to pick up passengers in a grid and drop them at their destination
- Merely capturing sensory information is not sufficient
- Need to “reason” about changes in the world i.e., when a passenger is picked up, he/she is no longer in the square, a new passenger may appear somewhere else etc.

Example Function

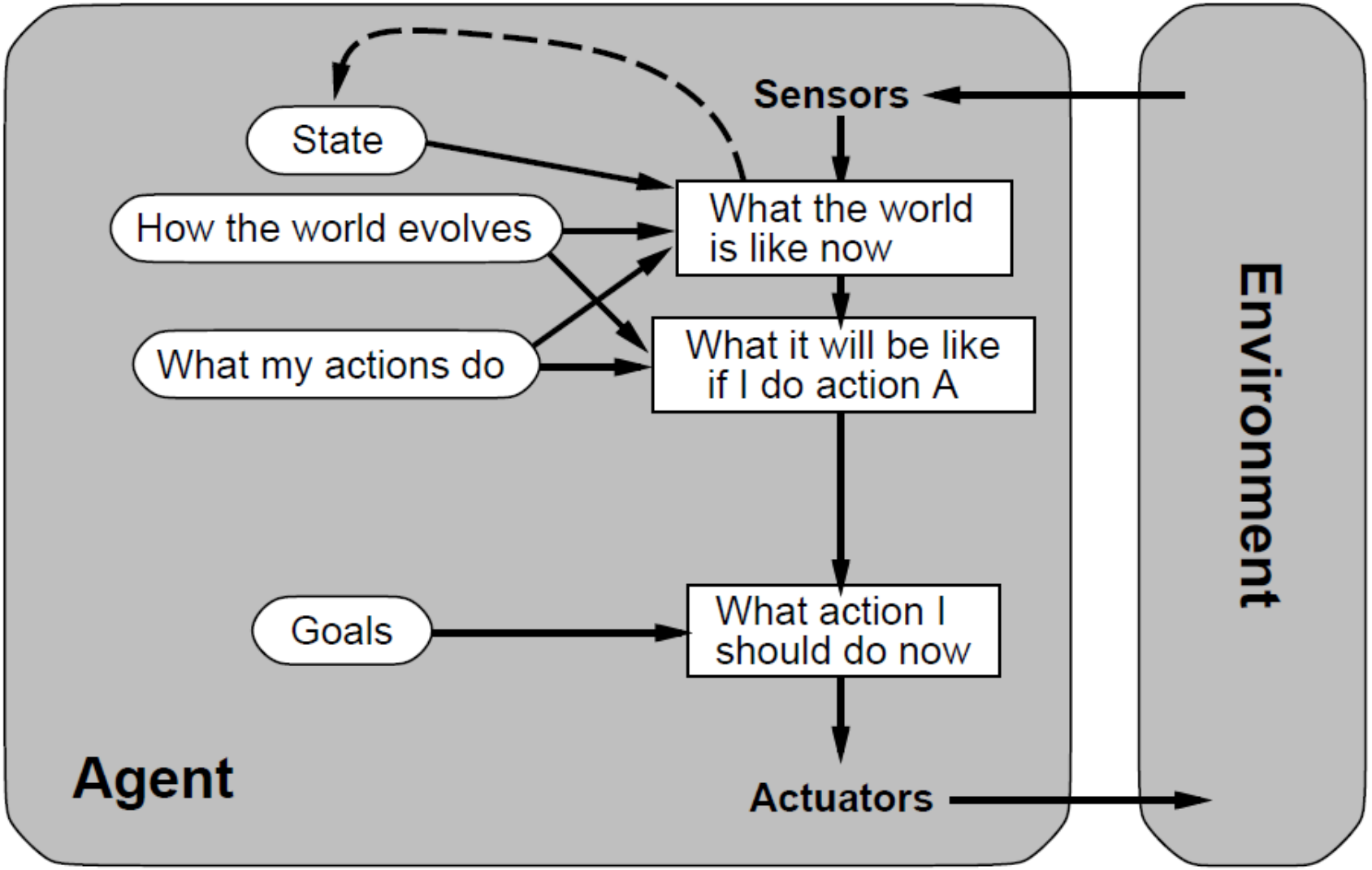
The location and status are simple sensory information – These are called percepts
State – Snapshot of the world – captures the internal history of the agent
Model – How does the world evolve? A description of next state given the current state

What would be the percepts, states and model in the grid world?

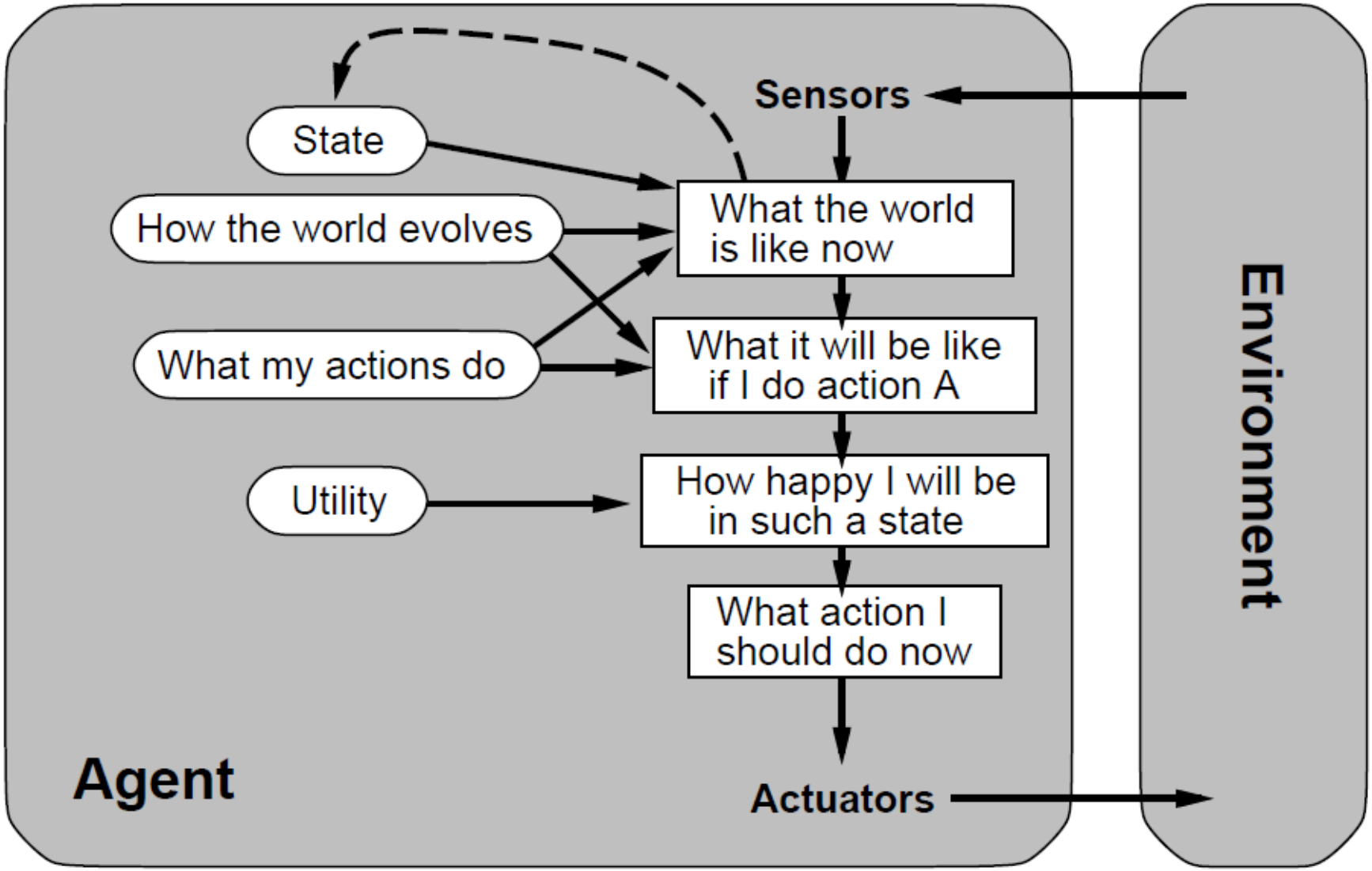
function Model_Reflex_Agent([*percepts]*) **returns** action
static: *state,model,rules,action*

// update current state based on previous state, action,percepts and the model
state = updateState(*state,action,percepts,model*)
// Find the best rule that matches
rule = matchRule(*state,rules*)
action = *rule*.ACTION
return *action*

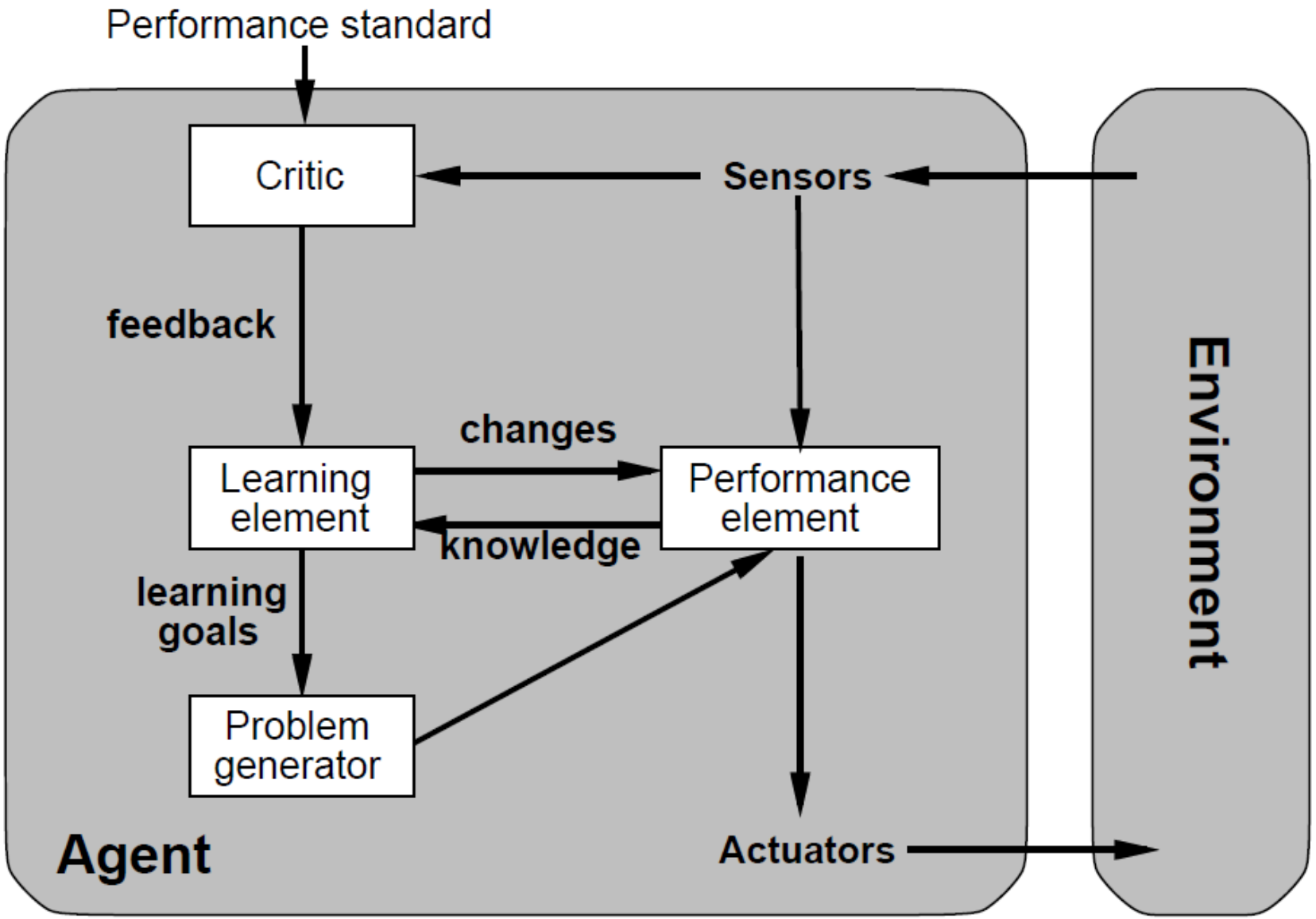
Goal-based agents



Utility-based agents



Learning agents



Thought Exercise

Can you create an example from the taxi domain where the agent is one of (1) goal-based, (2) utility-based and (3) learning. What are the states and model? Can you think of the action rule set in each of these domains?