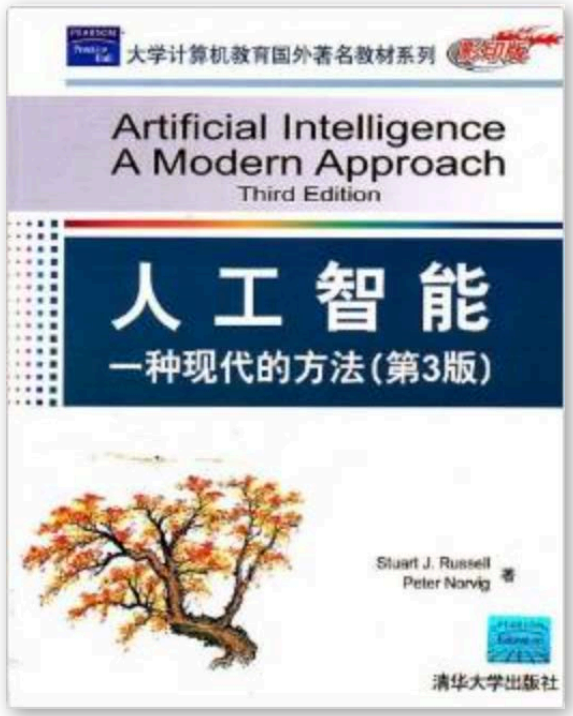
Lecture 1: Introduction

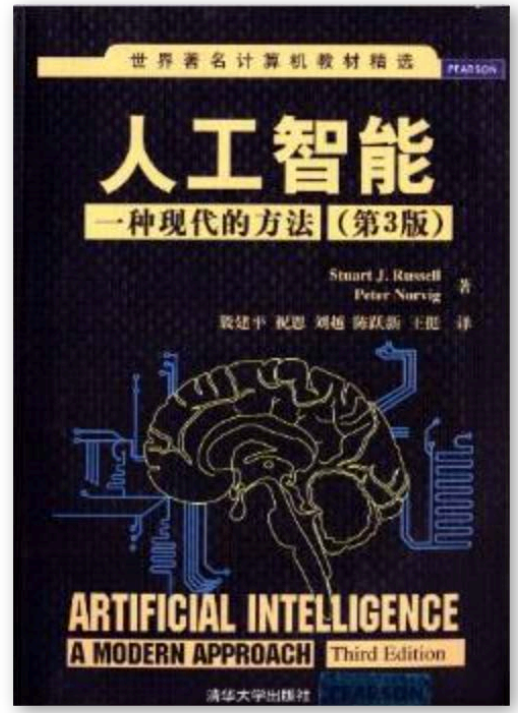
http://cs.nju.edu.cn/yuy/course_ai15.ashx

About this course

课程名称：人工智能

教材：AIMA





http://aima.cs.berkeley.edu/

About this course

时间: 周五3-4节 + 双周周三1-2节

课程主页：cs.nju.edu.cn/yuy/course_ai15.ashx

cs.nju.edu.cn/yuy

Teaching

- Artificial Intelligence. (for undergraduate students. Spring, 2015) >>>Course Page>>>
- Data Mining. (for M.Sc. students. Fall, 2014)
- Digital Image Processing. (for undergraduate students from Dept. Math., Spring, 2014)
- Data Mining. (for M.Sc. students. Fall, 2013)
- Introduction to Data Mining. (for undergraduate students. Spring, 2013)
- Digital Image Processing. (for undergraduate students. Spring, 2013)
- Data Mining. (for M.Sc. students. Fall, 2012)
- Introduction to Data Mining. (for undergraduate students. Spring, 2012)
- (Assistant) Data mining. (for graduate students. Fall, 2007)
- (Assistant) Algorithm Design and Analysis. (for undergraduate students. Fall, 2005)

About this course

人工智能课程主页

Edit this Page

(Back to homepage)

Information

- 授课对象: 计算机系本科生
- 教室: 仙林校区仙2-117
- 时间: 周五3-4节 + 双周周三1-2节
- 教材: Stuart J. Russell, Peter Norvig. Artificial Intelligence: A Modern Approach (3rd edition), Pearson, 2011.
- 助教: 杨敬文
- 总评: 课程作业 + 期末考试

作业

- >>>作业1: 推盒子游戏>>> 截止日期: 3月19日晚上8点整
- 作业2
- 作业3
- 作业4
- 作业5
- 黑白棋比赛

课程材料

1. Introduction

学术资源

- 人工智能领域学术期刊/杂志:
  - Artificial Intelligence
  - AI Magazine

What is intelligence?

Intelligence

From Wikipedia, the free encyclopedia

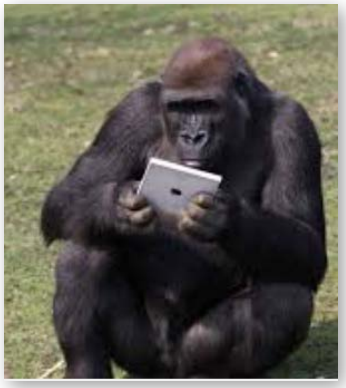
For other uses, see Intelligence (disambiguation).

Intelligence has been defined in many different ways such as in terms of one's capacity for logic, abstract thought, understanding, self-awareness, communication, learning, emotional knowledge, memory, planning, creativity and problem solving. It can also be more generally described as the ability to perceive and/or retain knowledge or information and apply it to itself or other instances of knowledge or information creating referable understanding models of any size, density, or complexity, due to any conscious or subconscious imposed will or instruction to do so.

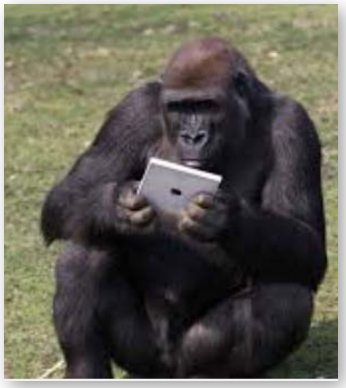
Intelligence is most widely studied in humans, but has also been observed in non-human animals and in plants. Artificial intelligence is the simulation of intelligence in machines.

What is intelligence?

What is intelligence?

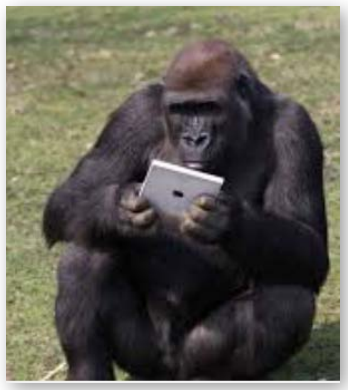


What is intelligence?



What is intelligence?

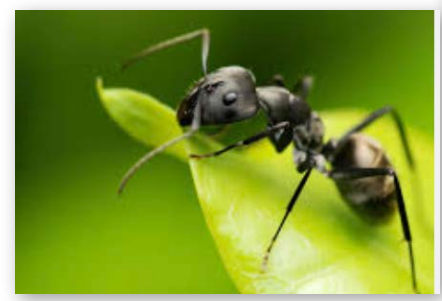
What is intelligence?



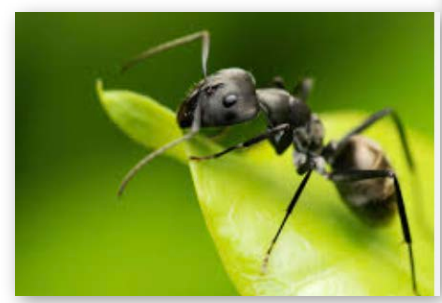
What is intelligence?

What is intelligence?

What is intelligence?



What is intelligence?

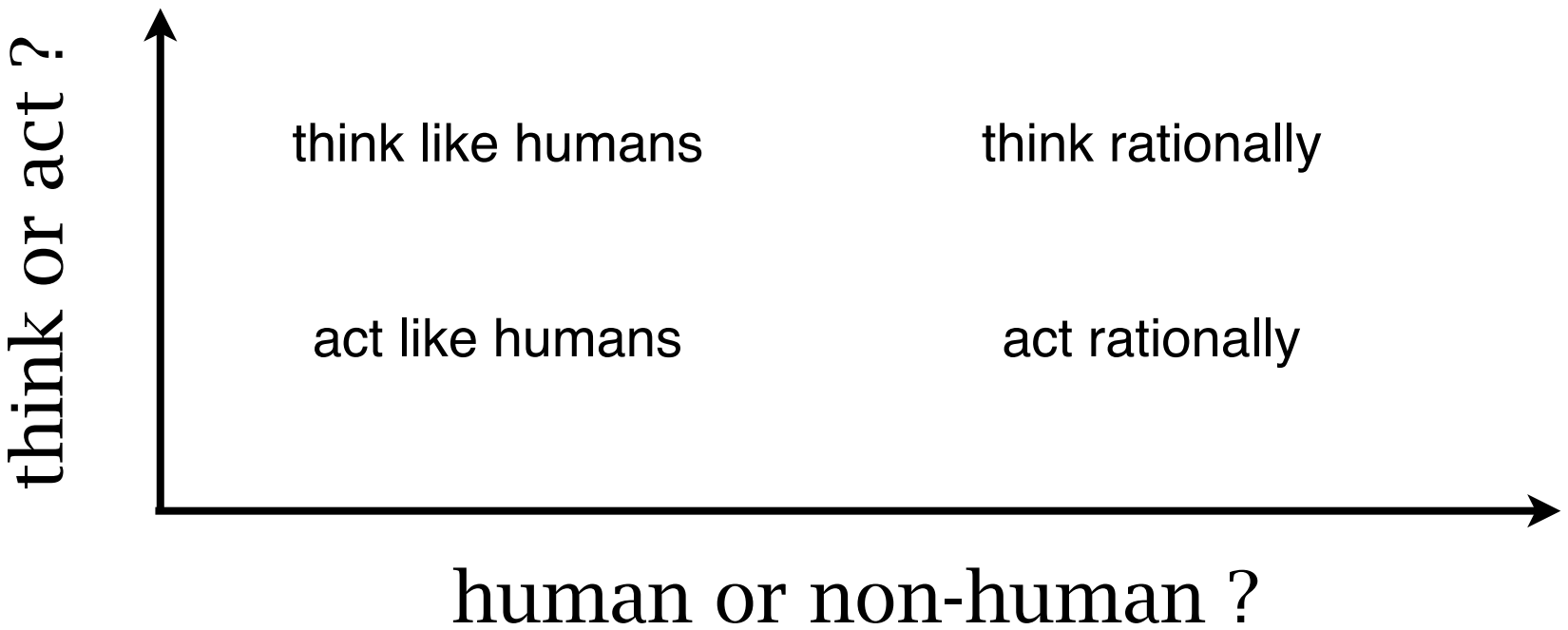


What is intelligence?

The uncertain about intelligence is a fundamental problem of AI

What is AI?

AI is a system that



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:

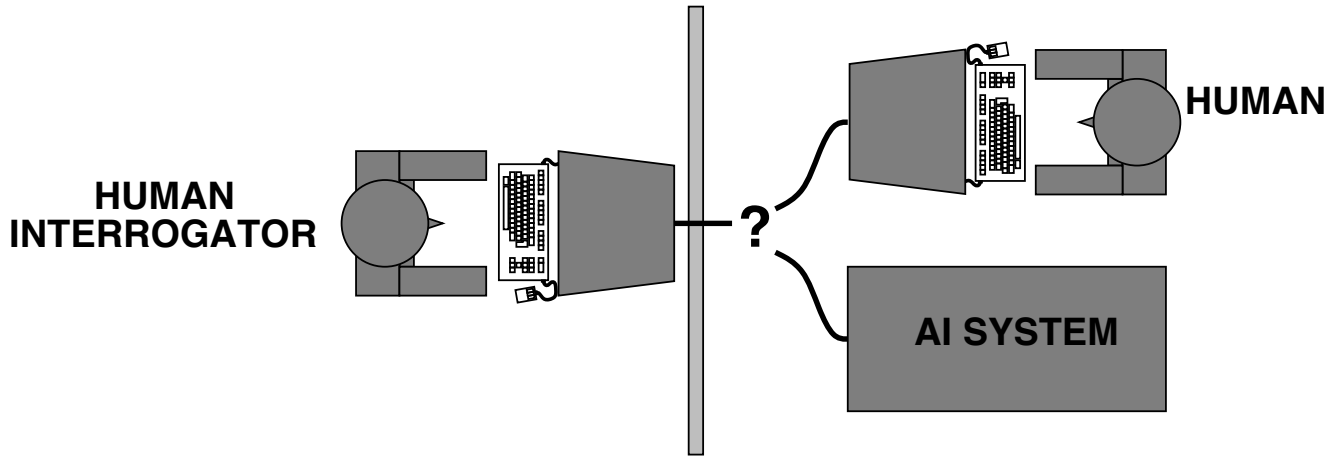
the available theories do not explain (or engender) anything resembling human-level general intelligence

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

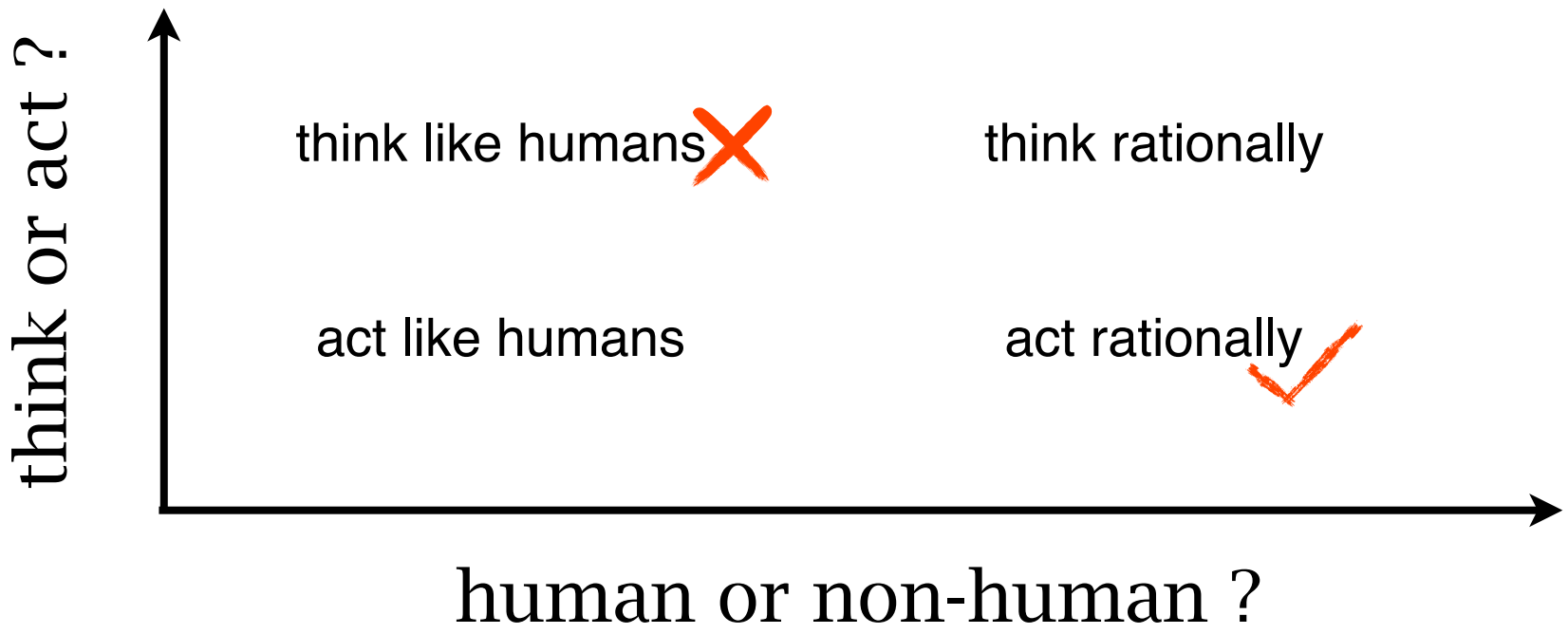
Doesn't necessarily involve thinking—e.g., blinking reflex—but thinking should be in the service of rational action

Aristotle (Nicomachean Ethics):

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

What is AI?

AI is a system that



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**

John McCarthy	Turing Award (1971)
Marvin Minsky	Turing Award (1969)
Nathaniel Rochester	
Claude Shannon	the father of information theory
Ray Solomonoff	
Oliver Selfridge	Father of Machine Perception
Trenchard More	
Arthur Samuel	
Herbert A. Simon	Turing Award (1975), Nobel Prize in Economics (1978)
Allen Newell	Turing Award (1975)

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 . . .
1990–	Machine learning quickly develops
2003–	Human-level AI back on the agenda
2006–	Machine learning industry booms

What we call AI in markets

人脸检测、识别

S.I.R.I.

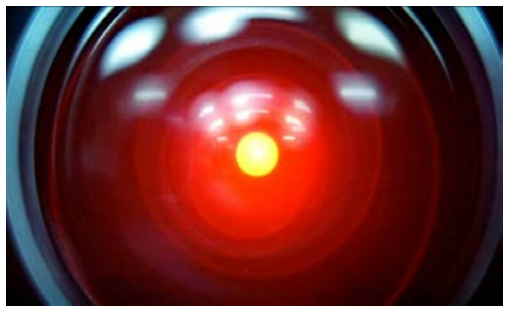
自动驾驶

推荐系统

下棋

BigDog

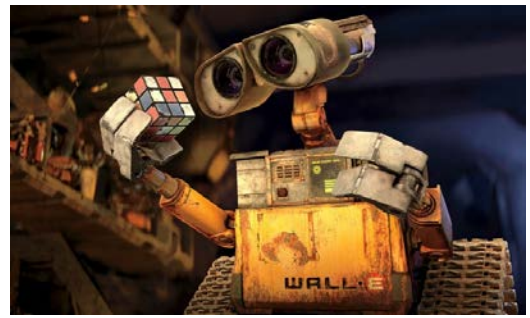
What we call AI in movies



2001: A Space Odyssey
1968

The Matrix
1999

A.I. Artificial Intelligence
2001



Wall-E
2008

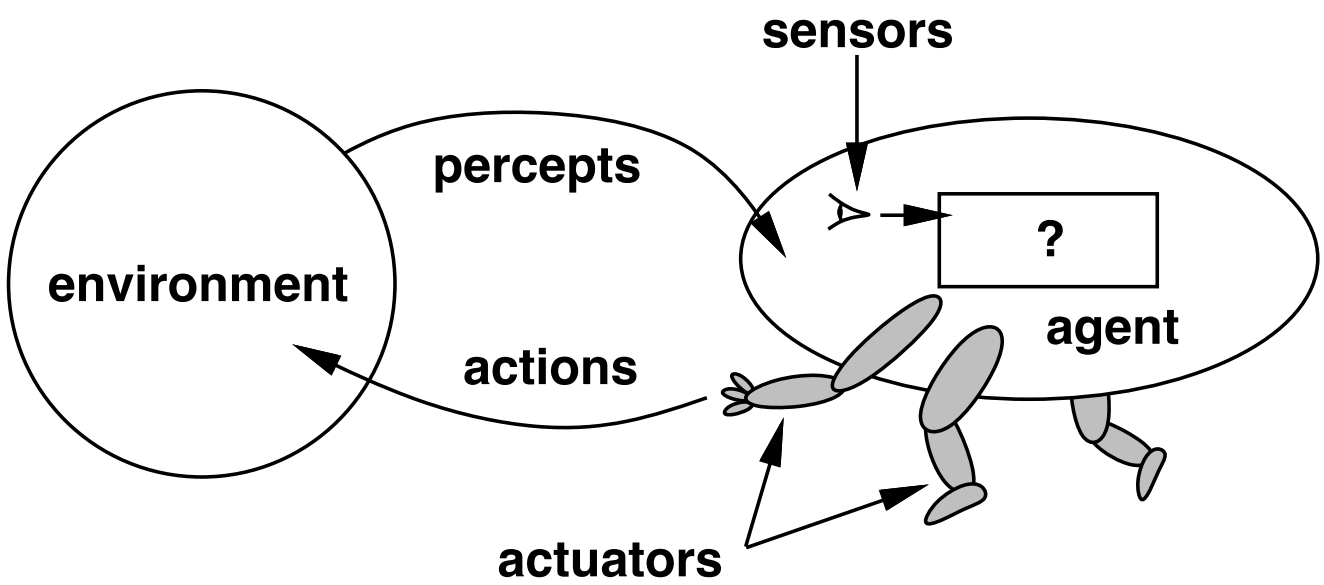
I, Robot
2004

Interstellar
2014

The Terminator
1984

Agent

This course is about designing rational agents



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

The agent function maps from percept histories to actions:

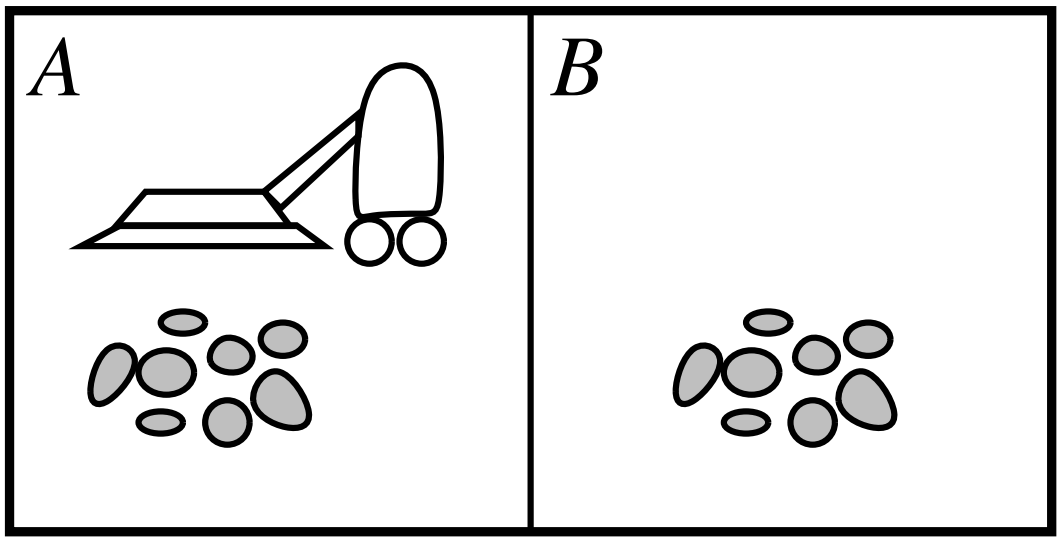
$$f : \mathcal{P}^* \rightarrow \mathcal{A}$$

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

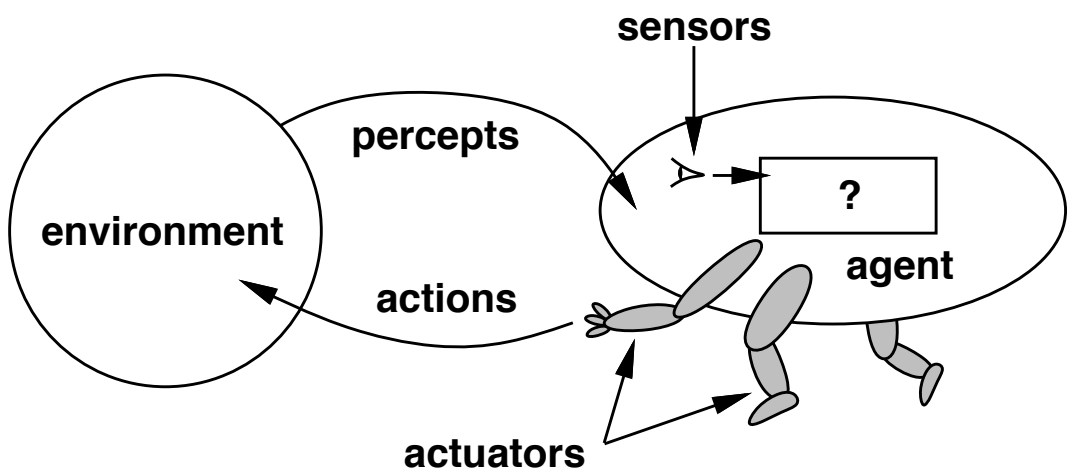
Hence, rational $\neq$ successful

Rational $\Rightarrow$ exploration, learning, autonomy

PEAS

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

Performance measure??
Environment??
Actuators??
Sensors??



Examples

Automatic taxi agent

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

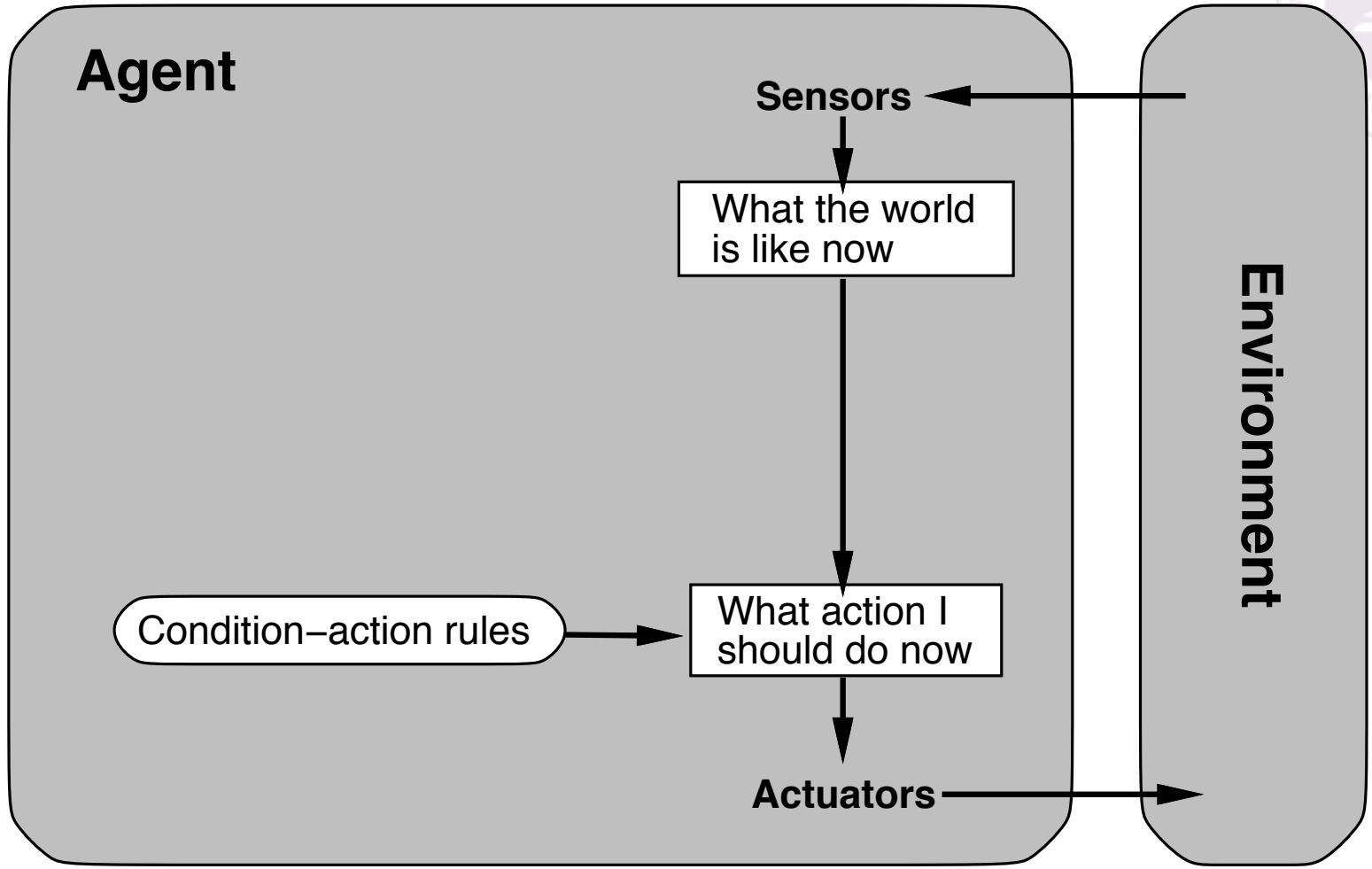
Agent types

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

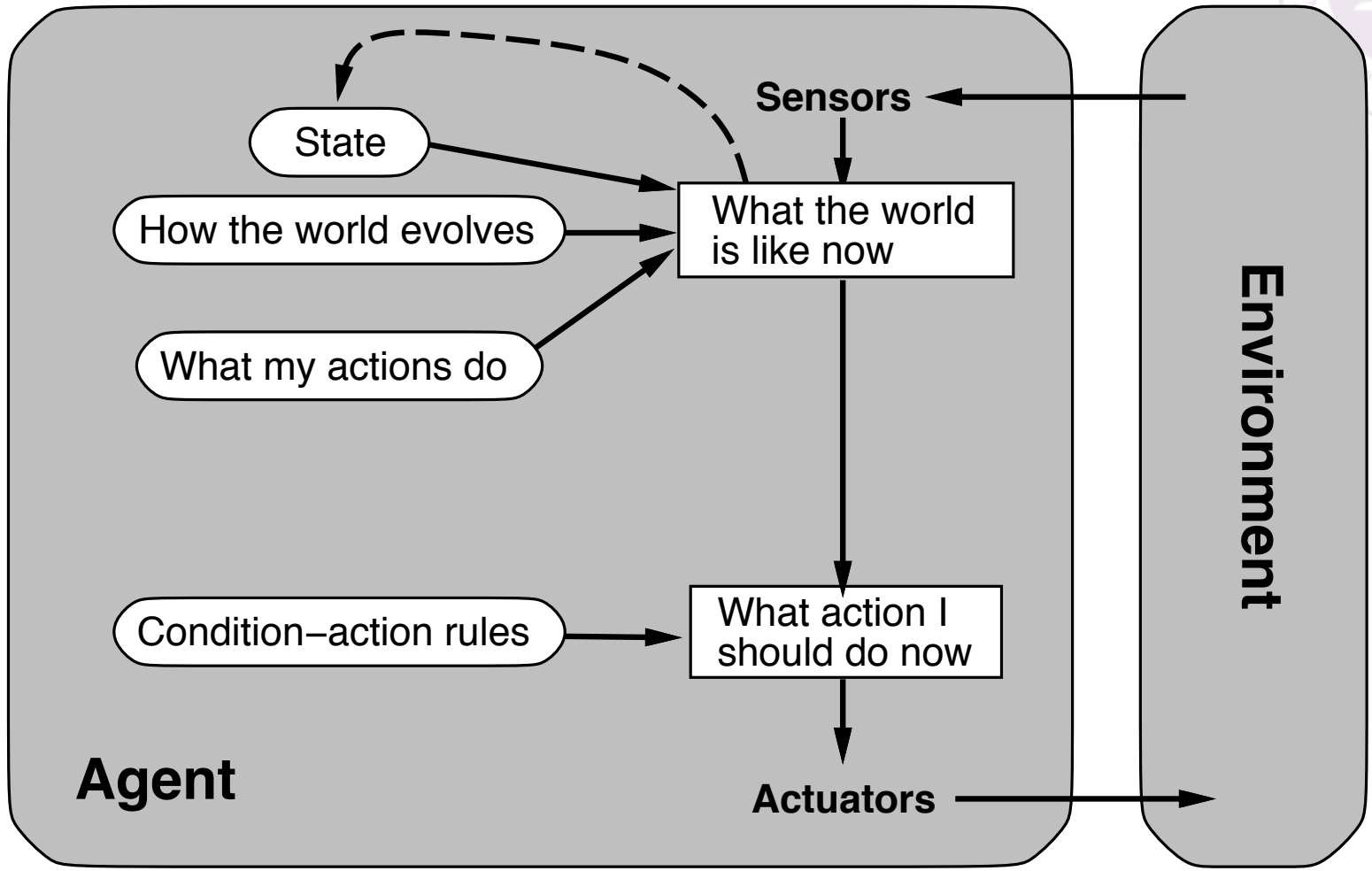
Simple reflex agents



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*

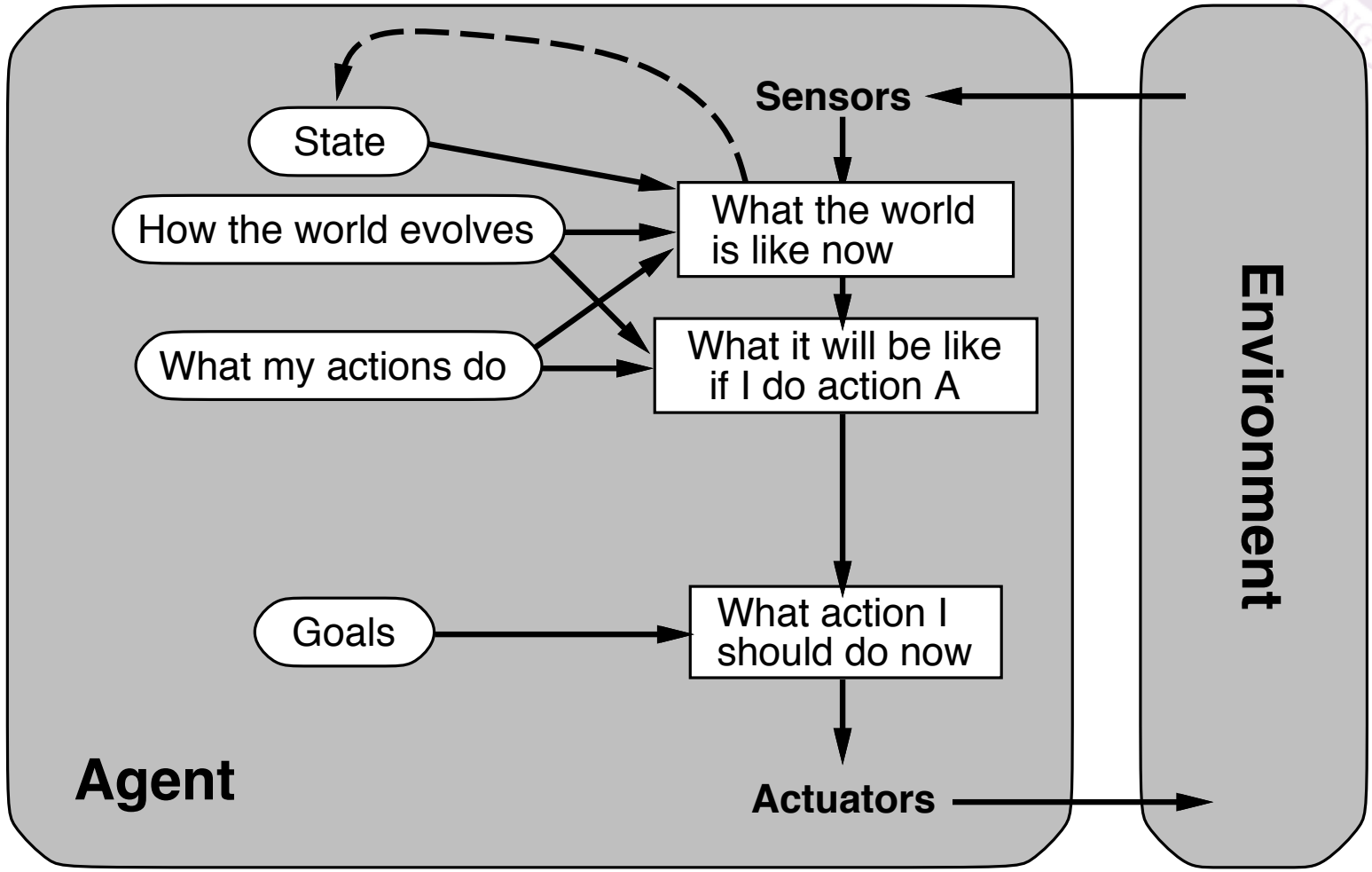
Reflex agents with state



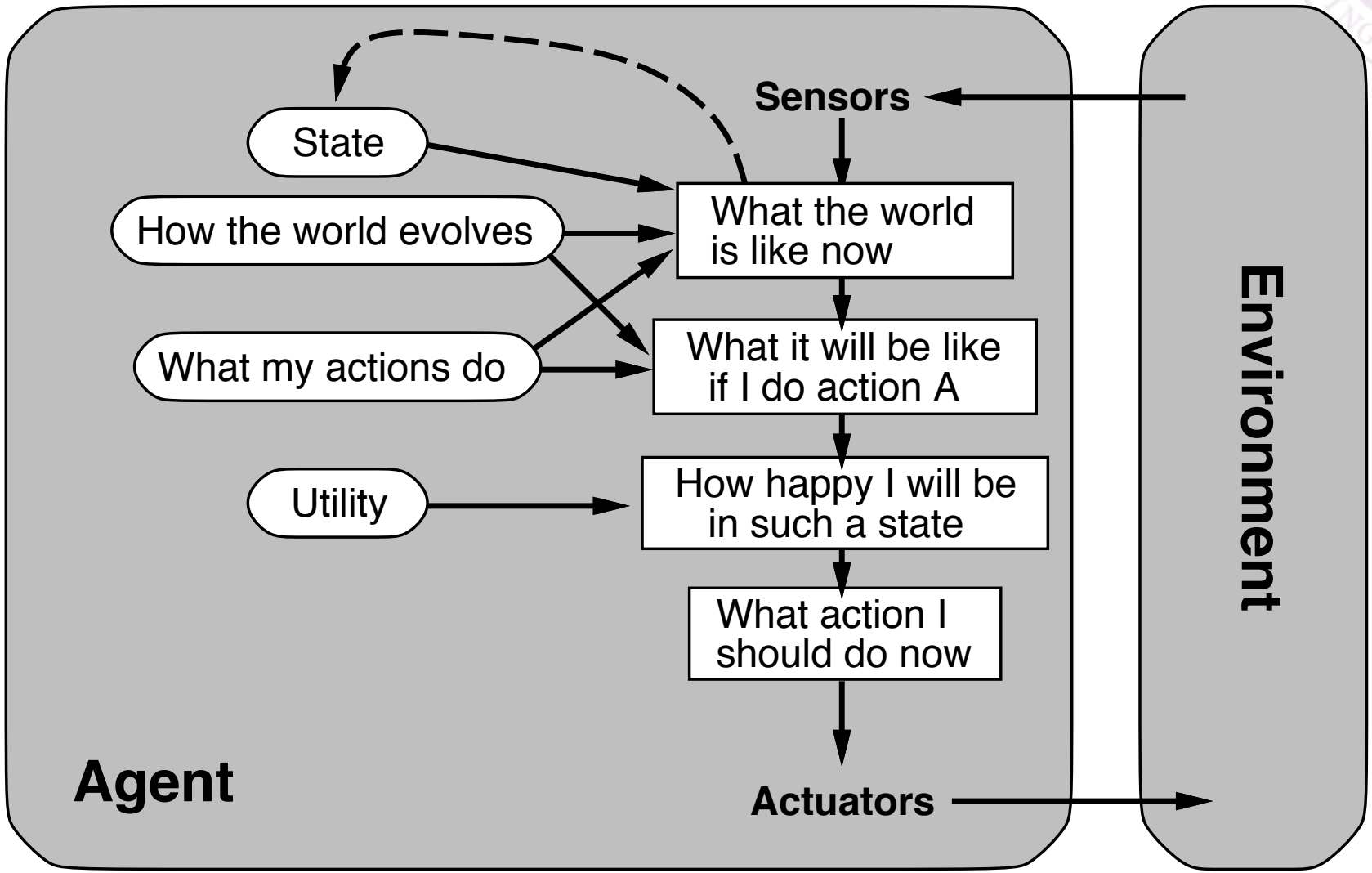
function REFLEX-VACUUM-AGENT([*location*,*status*]) **returns** an action
static: *last_A*, *last_B*, numbers, initially ∞

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

Goal-based agents



Utility-based agents



Learning agents