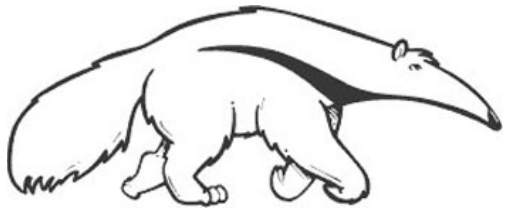
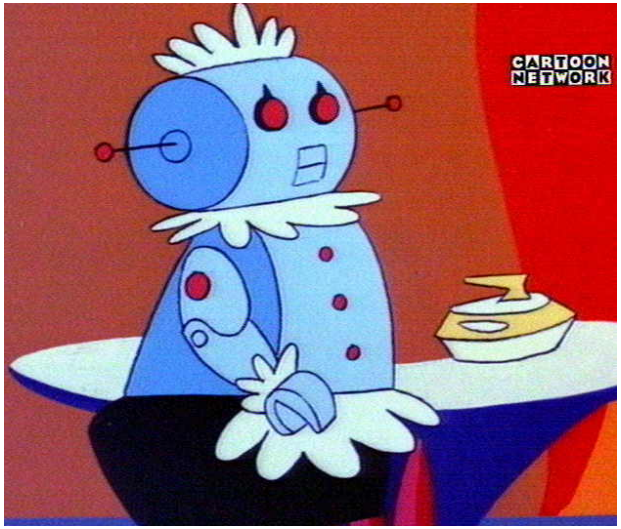
Introduction to Artificial Intelligence

CS171, Winter Quarter, 2019
Introduction to Artificial Intelligence
Prof. Richard Lathrop



Read Beforehand: R&N 1-2, 26.preamble, 26.3-4, 27.4
Optional: R&N 26.1-2, 27.1-3

What is AI?


CARTOON NETWORK

?

=

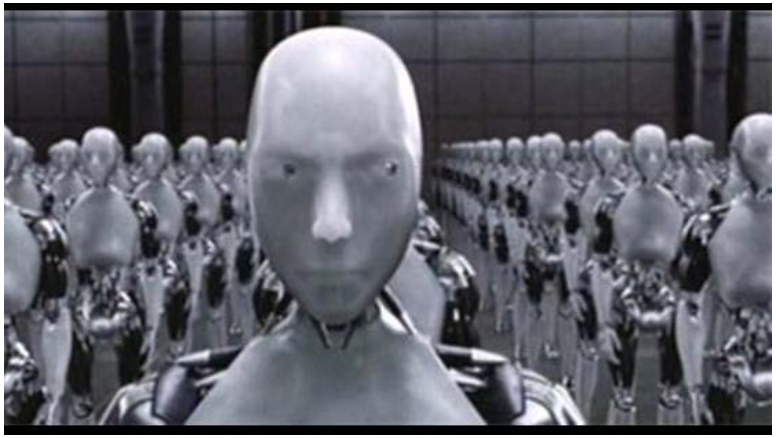

Roomba SCHEDULER
iRobot

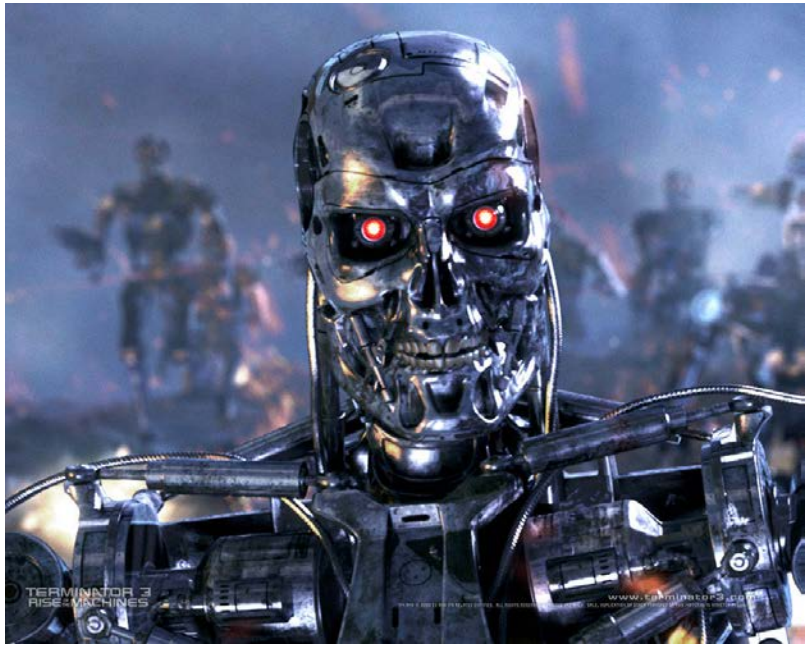
?

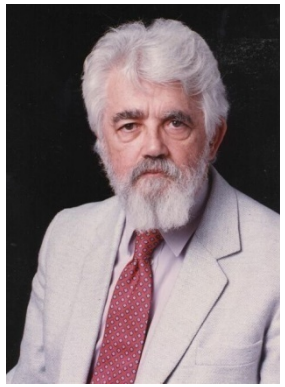
=


JVC
D-VHS
D-THEATER

What is AI?







What is Artificial Intelligence
(John McCarthy , Basic Questions)

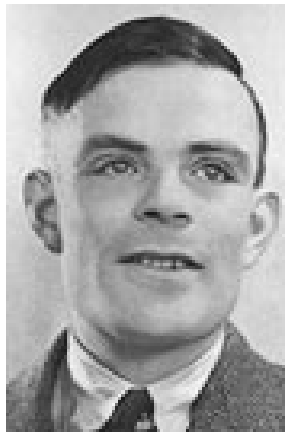
- **What is artificial intelligence?**
- It is the science and engineering of making intelligent machines, especially intelligent computer programs. It is related to the similar task of using computers to understand human intelligence, but AI does not have to confine itself to methods that are biologically observable.

- **Yes, but what is intelligence?**
- Intelligence is the computational part of the ability to achieve goals in the world. Varying kinds and degrees of intelligence occur in people, many animals and some machines.

- **Isn't there a solid definition of intelligence that doesn't depend on relating it to human intelligence?**
- Not yet. The problem is that we cannot yet characterize in general what kinds of computational procedures we want to call intelligent. We understand some of the mechanisms of intelligence and not others.

- **More in:** http://www-formal.stanford.edu/jmc/whatisai/node1.html

The Turing test



Can Machine think? A. M. Turing, 1950

- Test requires computer to "pass itself off" as human
  - Necessary?
  - Sufficient?

- Requires:
  - Natural language
  - Knowledge representation
  - Automated reasoning
  - Machine learning
  - (vision, robotics) for full test

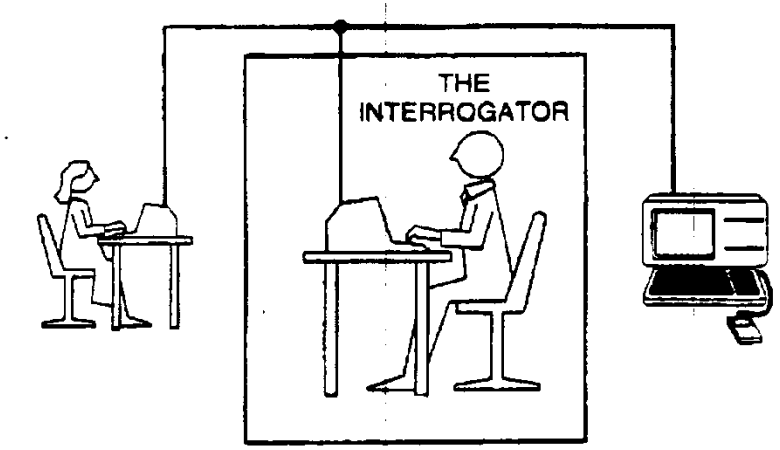


Figure 1.1 The Turing test.

What is Artificial Intelligence?

- Nils J. Nilsson :
  - "Artificial intelligence is that activity devoted to making machines intelligent, and intelligence is that quality that enables an entity to function appropriately and with foresight in its environment."

What is Artificial Intelligence?

- **Thought processes**
  - "The exciting new effort to make computers **think** .. Machines with minds, in the full and literal sense" (Haugeland, 1985)
- **Behavior**
  - "The study of how to make computers **do things** at which, at the moment, people are better." (Rich, and Knight, 1991)
- **Activities**
  - The automation of activities that we associate with human thinking, activities such as decision-making, problem solving, learning… (Bellman)
- **Things we would call "intelligent" if done by a human**

AI as “Raisin Bread”

- Esther Dyson [predicted] AI would [be] embedded in main-stream, strategically important systems, like raisins in a loaf of raisin bread.
  - The “bread” represents any main-stream computer-augmented engineering system.
  - The “raisins” represent nuggets of control, where “smart” control ⇒ better function.
- Time has proven Dyson's prediction correct.
- Emphasis shifts away from replacing expensive human experts toward main-stream computing systems that create strategic advantage.
- Many AI systems connect to large data bases, deal with legacy data, talk to networks, handle noise and data corruption, are implemented in popular languages, and run on standard operating systems.
- Humans usually are important contributors to the total solution.
  - Adapted from Patrick Winston, former Director, MIT AI Laboratory

What is AI?

- Competing axes of definitions:
  - Thinking vs. Acting
  - Human-like vs. Rational

  - Often not the same thing
  - Cognitive science, economics, …

- How to simulate human intellect & behavior by machine
  - Mathematical problems (puzzles, games, theorems)
  - Common-sense reasoning
  - Expert knowledge (law, medicine)
  - Social behavior
  - Web & online intelligence
  - Planning, e.g. operations research

Act/Think Humanly/Rationally

- Act Humanly
  - Turing test
- Think Humanly
  - Introspection; Cognitive science
- Think rationally
  - Logic; representing & reasoning over problems
- Acting rationally
  - Agents; sensing & acting; feedback systems

Current “Hot” areas/applications

- Big Data/knowledge extraction with Machine Learning
  - BD2K = “Big Data to Knowledge”
- Deep Learning/artificial neural systems
- Transportation/logistics/self-driving cars
- Robotics/factory automation/mobility for the disabled
- Vision/scene or video analysis
- Internet/social media
- Biology/medicine/improving healthcare
- Question answering/knowledge retrieval
- Finance/market trading/personal wealth management
- Your favorite area here….

Agents

- An agent is anything that can be viewed as perceiving its environment through sensors and acting upon that environment through actuators

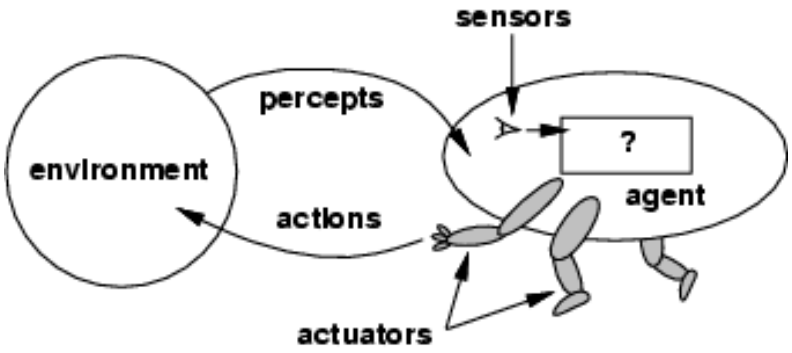
- Human agent:
  - Sensors: eyes, ears, …
  - Actuators: hands, legs, mouth…

- Robotic agent
  - Sensors: cameras, range finders, …
  - Actuators: motors

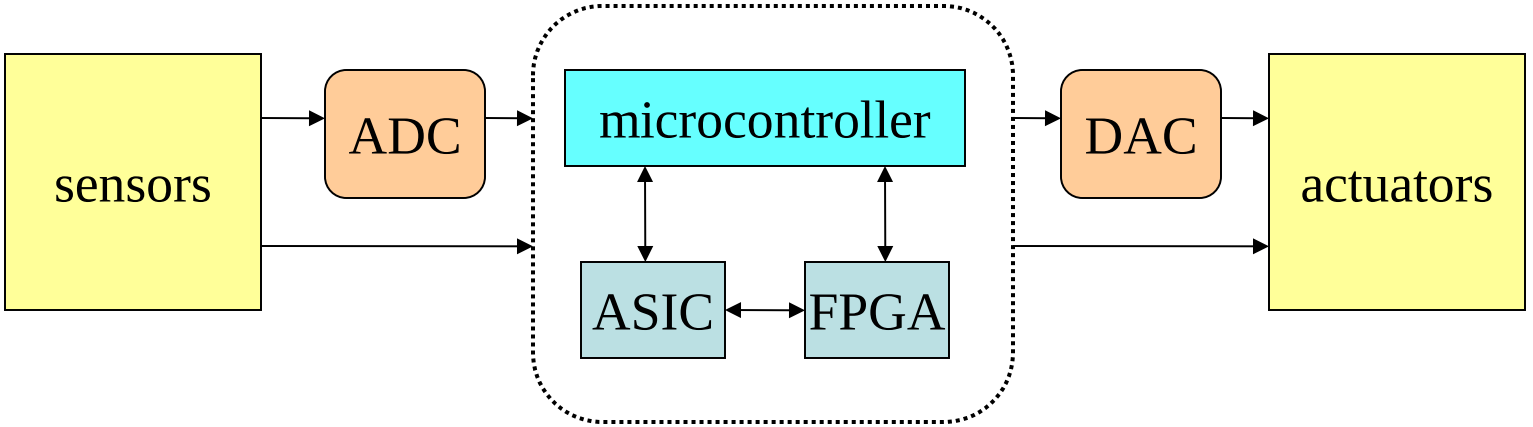
Agents

- **Craik (1943; R&N p. 13) specified the three key steps of a knowledge-based agent:**
  - (1) the stimulus must be translated into an internal representation;
  - (2) the representation is manipulated by cognitive processes to derive new internal representations;
  - and (3) these representations are in turn retranslated back into action.

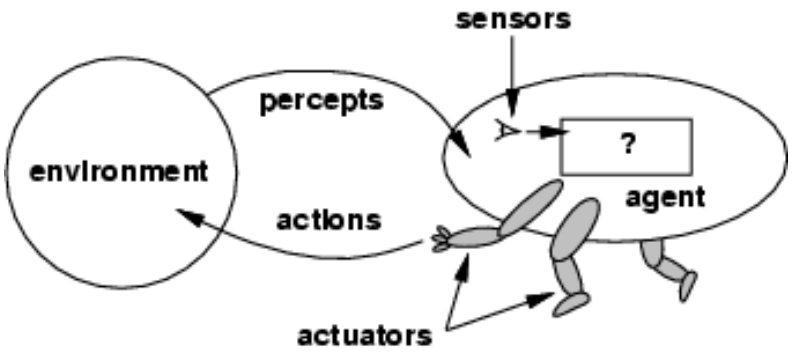
Agents and environments



Compare: Standard Embedded System Structure



Agents and environments



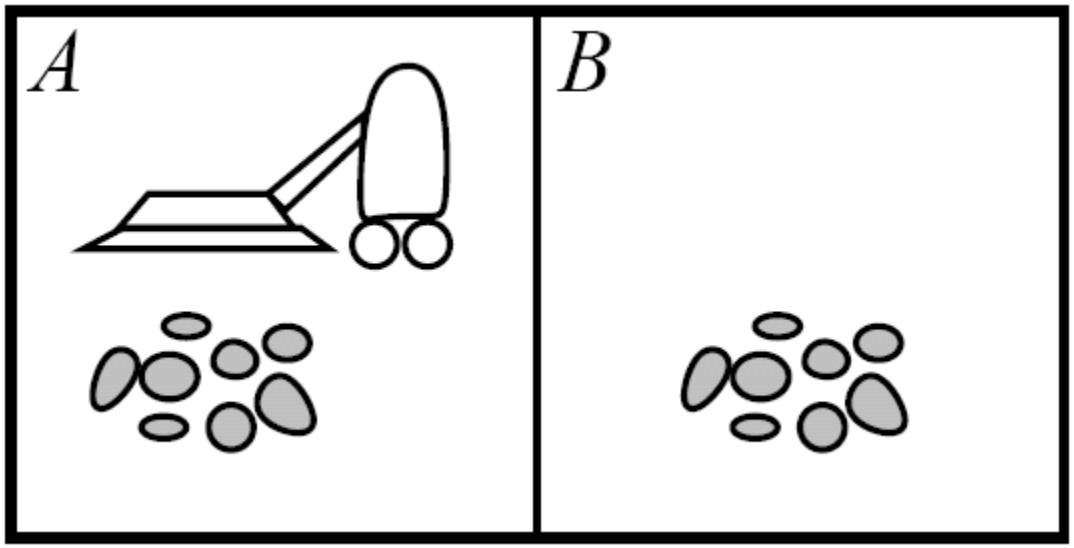
- The agent function maps from percept histories to actions:

$$[f: \mathcal{P}^{\star} \rightarrow \mathcal{A}]$$

- The agent program runs on the physical architecture to produce f

- agent = architecture + program

Vacuum World



- Percepts: location, contents
  - e.g., [A, dirty]
- Actions: {left, right, vacuum,…}

Rational agents

- Rational Agent: For each possible percept sequence, a rational agent should select an action that is *expected* to maximize its performance measure, based on the evidence provided by the percept sequence and whatever built-in knowledge the agent has.

- Performance measure: An objective criterion for success of an agent's behavior ("cost", "reward", "utility")

- E.g., performance measure of a vacuum-cleaner agent could be amount of dirt cleaned up, amount of time taken, amount of electricity consumed, amount of noise generated, etc.

Rational agents

- Rationality is distinct from omniscience (all-knowing with infinite knowledge)
- Agents can perform actions in order to modify future percepts so as to obtain useful information (information gathering, exploration)
- An agent is autonomous if its behavior is determined by its own percepts & experience (with ability to learn and adapt) without depending solely on build-in knowledge

Task environment

- To design a rational agent, we must specify the task environment — **"PEAS"**
- Example: automated taxi system
  - Performance measure
    - Safety, destination, profits, legality, comfort, …
  - Environment
    - City streets, freeways; traffic, pedestrians, weather, …
  - Actuators
    - Steering, brakes, accelerator, horn, …
  - Sensors
    - Video, sonar, radar, GPS / navigation, keyboard, …

PEAS

- Example: Agent = Medical diagnosis system

Performance measure: Healthy patient, minimize costs, lawsuits

Environment: Patient, hospital, staff

Actuators: Screen display (questions, tests, diagnoses, treatments, referrals)

Sensors: Keyboard (entry of symptoms, findings, patient's answers)

PEAS

- Example: Agent = Part-picking robot (a robot that picks up parts or tools and places them in a new location)
- Performance measure: Percentage of parts in correct bins
- Environment: Conveyor belt with parts, bins
- Actuators: Jointed arm and hand
- Sensors: Camera, joint angle sensors

Environment types

- **Fully observable (vs. partially observable):** An agent's sensors give it access to the complete state of the environment at each point in time.
- **Deterministic (vs. stochastic):** The next state of the environment is completely determined by the current state and the action executed by the agent. (If the environment is deterministic except for the actions of other agents, then the environment is **strategic**)
- **Episodic (vs. sequential):** An agent's action is divided into atomic episodes. Decisions do not depend on previous decisions/actions.
- **Known (vs. unknown):** An environment is considered to be "known" if the agent understands the laws that govern the environment's behavior.

Environment types

- **Static (vs. dynamic):** The environment is unchanged while an agent is deliberating. (The environment is **semidynamic** if the environment itself does not change with the passage of time but the agent's performance score does)

- **Discrete (vs. continuous)**: A limited number of distinct, clearly defined percepts and actions.
  - How do we **represent** or **abstract** or **model** the world?

- **Single agent (vs. multi-agent):** An agent operating by itself in an environment. Does the other agent interfere with my performance measure?

task environm.	observable	determ./ stochastic	episodic/ sequential	static/ dynamic	discrete/ continuous	agents
crossword puzzle	fully	determ.	sequential	static	discrete	single
chess with clock	fully	strategic	sequential	semi	discrete	multi
poker						
back gammon						
taxi driving	partial	stochastic	sequential	dynamic	continuous	multi
medical diagnosis	partial	stochastic	sequential	dynamic	continuous	single
image analysis	fully	determ.	episodic	semi	continuous	single
partpicking robot	partial	stochastic	episodic	dynamic	continuous	single
refinery controller	partial	stochastic	sequential	dynamic	continuous	single
interact. Eng. tutor	partial	stochastic	sequential	dynamic	discrete	multi

task environm.	observable	determ./ stochastic	episodic/ sequential	static/ dynamic	discrete/ continuous	agents
crossword puzzle	fully	determ.	sequential	static	discrete	single
chess with clock	fully	strategic	sequential	semi	discrete	multi
poker	partial	stochastic	sequential	static	discrete	multi
back gammon						
taxi driving	partial	stochastic	sequential	dynamic	continuous	multi
medical diagnosis	partial	stochastic	sequential	dynamic	continuous	single
image analysis	fully	determ.	episodic	semi	continuous	single
partpicking robot	partial	stochastic	episodic	dynamic	continuous	single
refinery controller	partial	stochastic	sequential	dynamic	continuous	single
interact. Eng. tutor	partial	stochastic	sequential	dynamic	discrete	multi

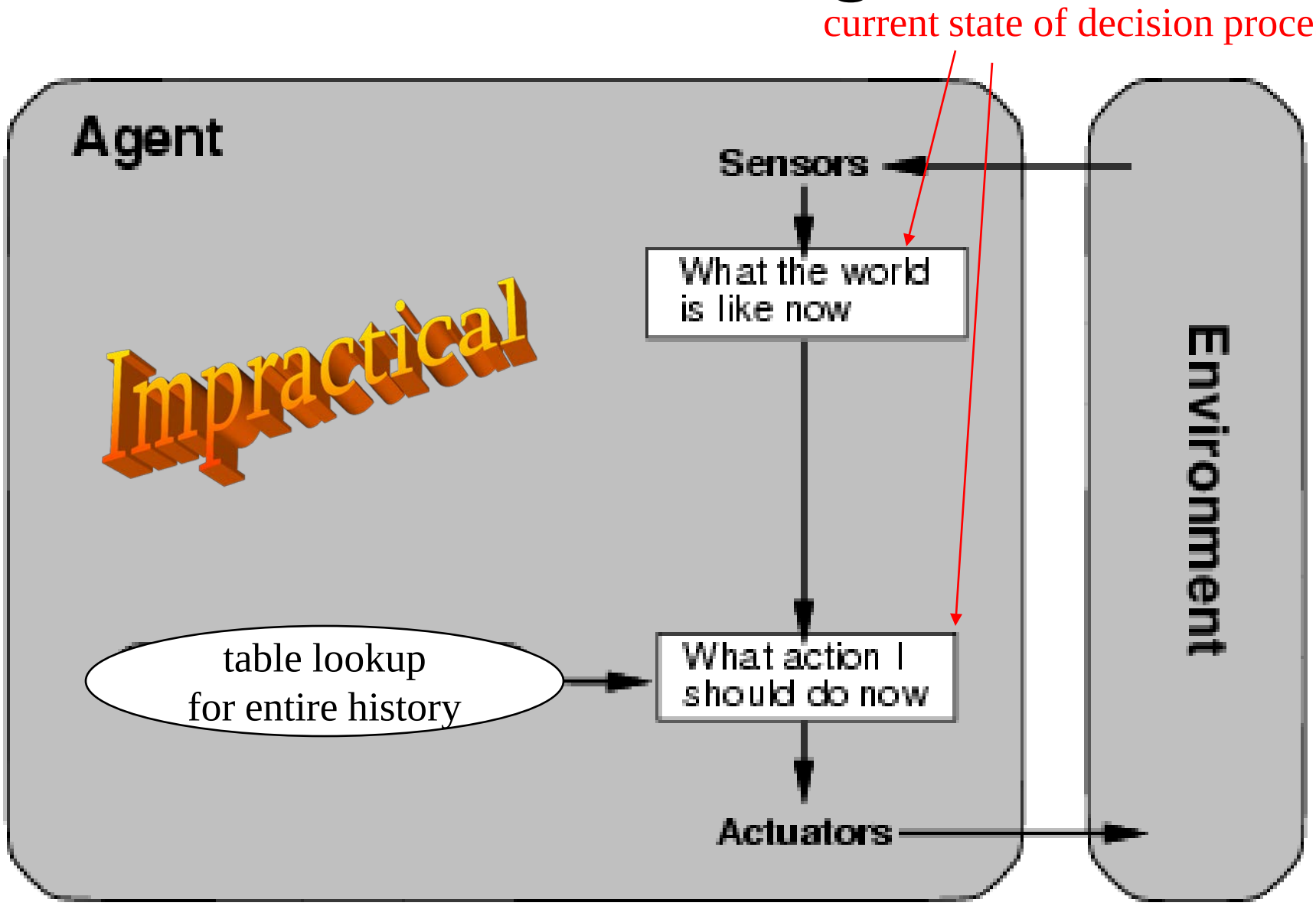
task environm.	observable	determ./ stochastic	episodic/ sequential	static/ dynamic	discrete/ continuous	agents
crossword puzzle	fully	determ.	sequential	static	discrete	single
chess with clock	fully	strategic	sequential	semi	discrete	multi
poker	partial	stochastic	sequential	static	discrete	multi
back gammon	fully	stochastic	sequential	static	discrete	multi
taxi driving	partial	stochastic	sequential	dynamic	continuous	multi
medical diagnosis	partial	stochastic	sequential	dynamic	continuous	single
image analysis	fully	determ.	episodic	semi	continuous	single
partpicking robot	partial	stochastic	episodic	dynamic	continuous	single
refinery controller	partial	stochastic	sequential	dynamic	continuous	single
interact. Eng. tutor	partial	stochastic	sequential	dynamic	discrete	multi

Agent types

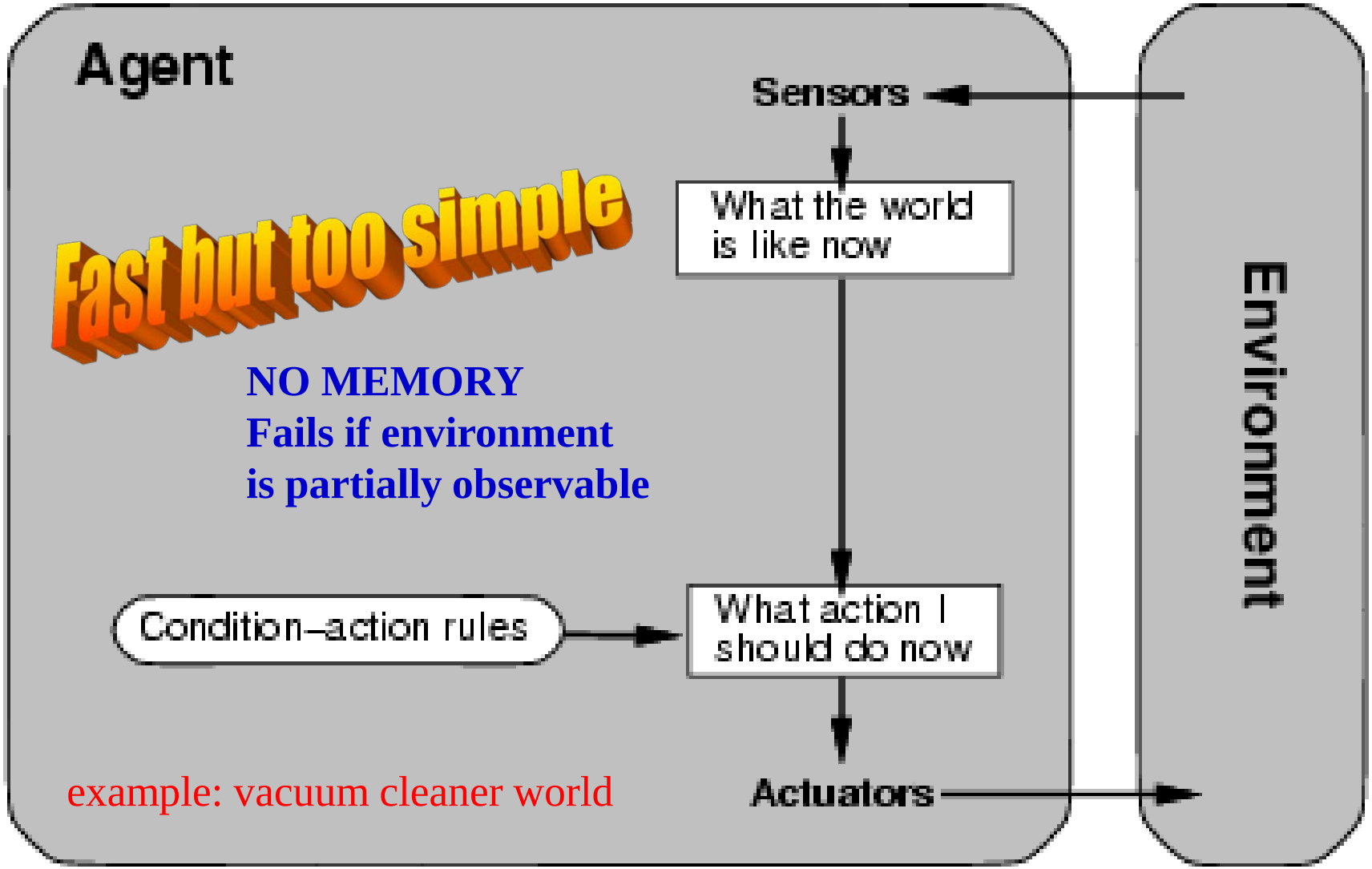
Six basic types, in order of increasing generality:

- Table Driven agents
- Simple reflex agents
- Model-based reflex agents
- Goal-based agents
- Utility-based agents
- Learning agents

Table Driven Agent.



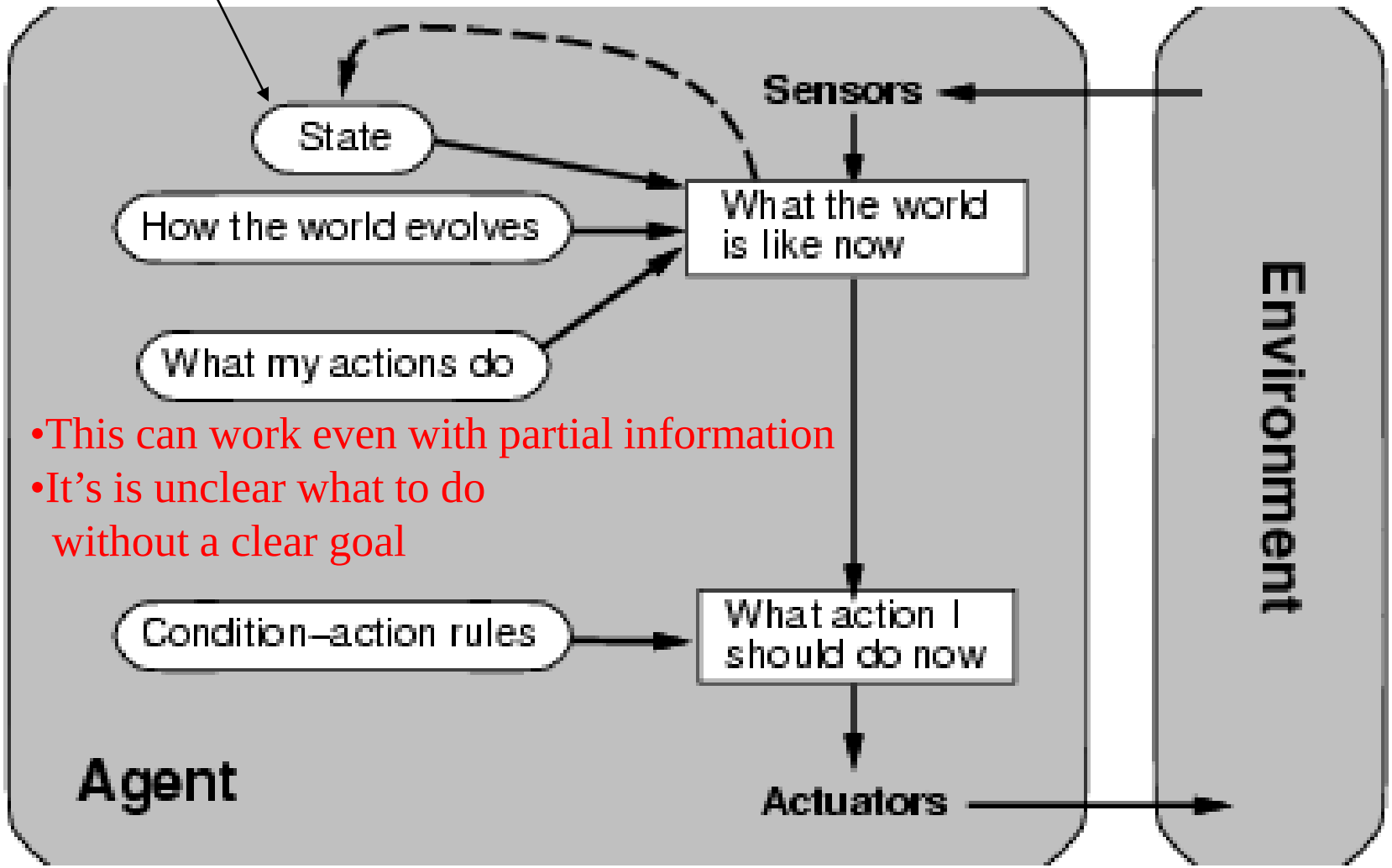
Simple reflex agents



Model-based reflex agents

description of
current world state

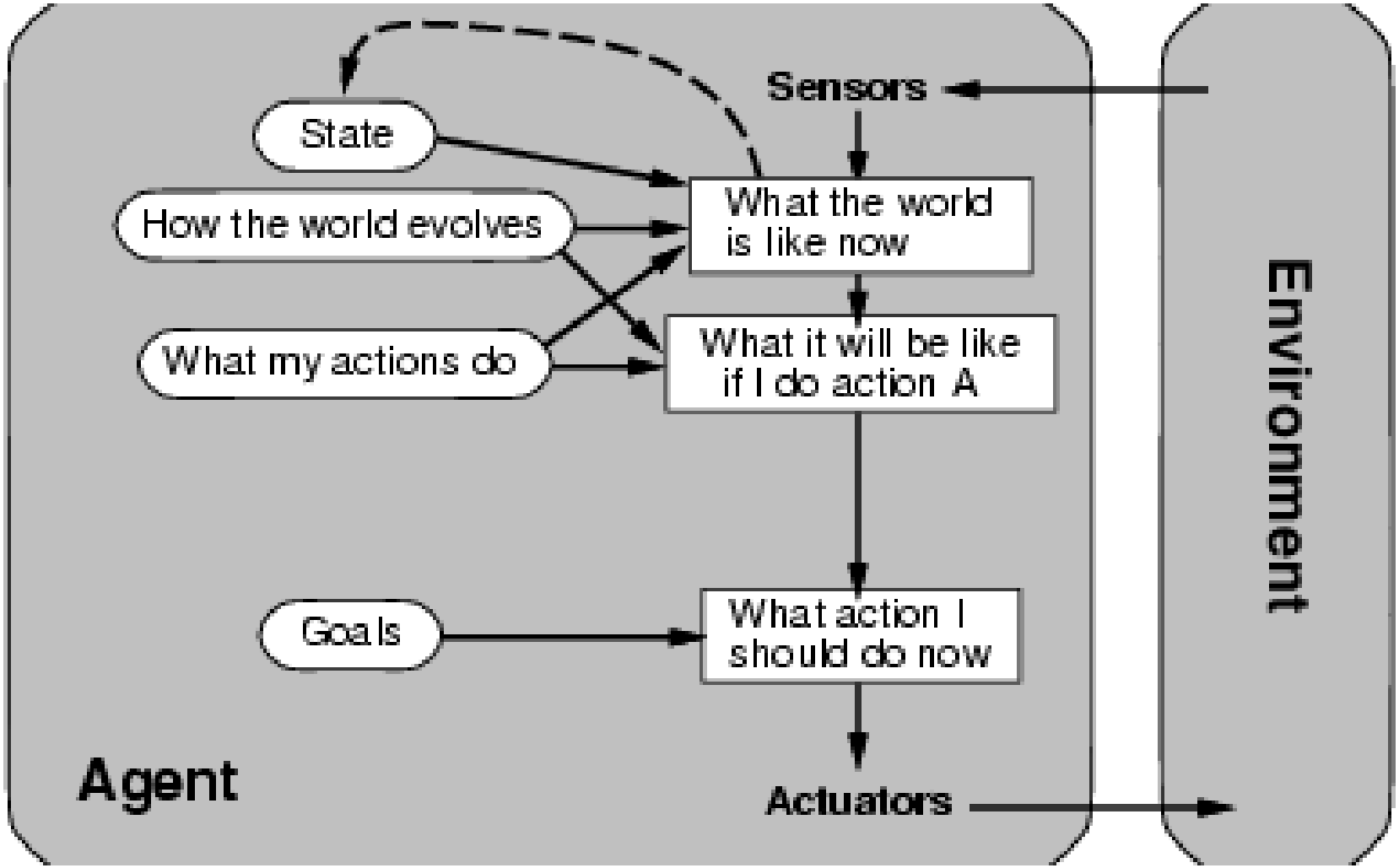
Model the state of the world by:
modeling how the world changes
how its actions change the world



- This can work even with partial information
- It's is unclear what to do without a clear goal

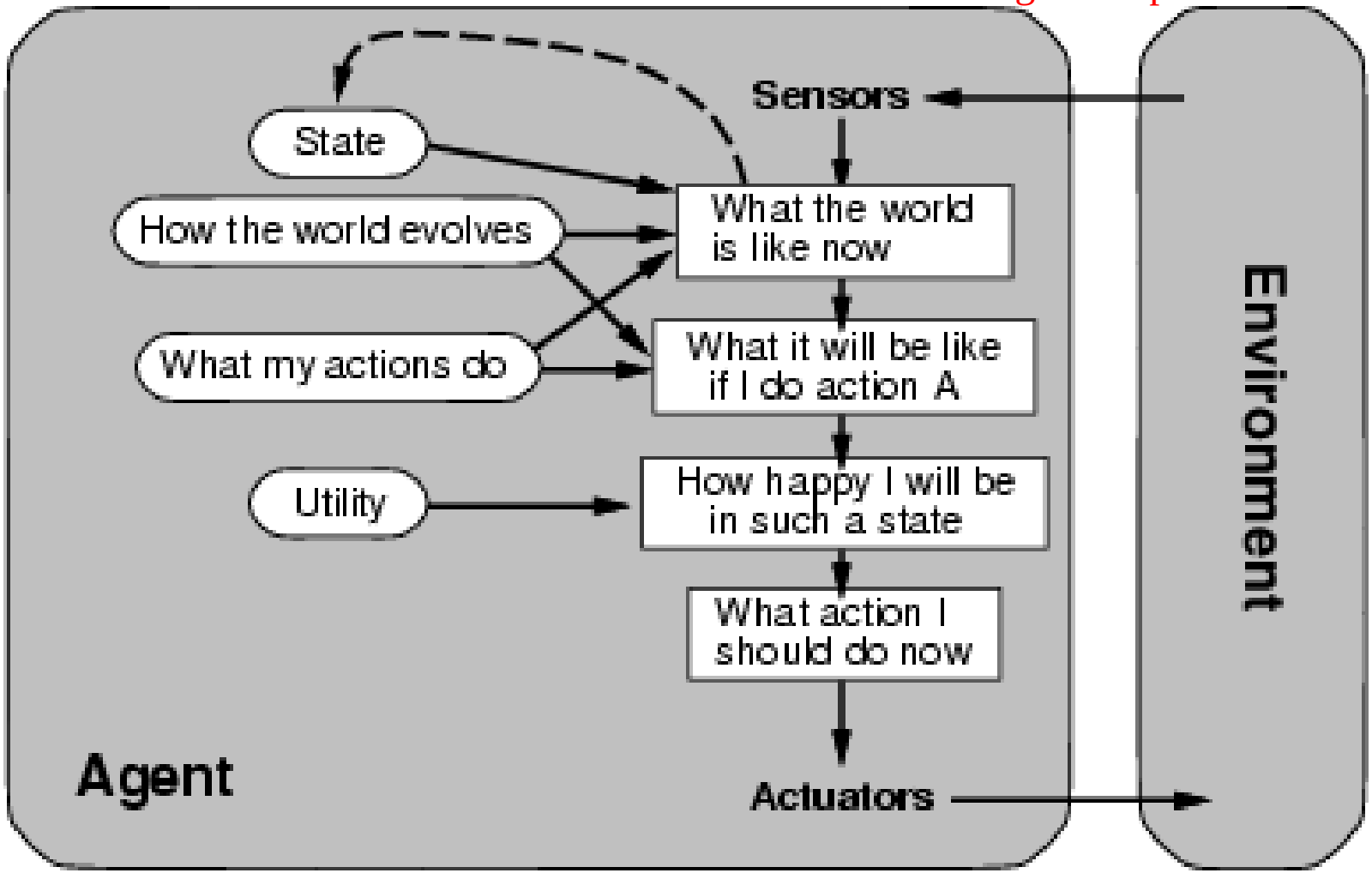
Goal-based agents

Goals provide reason to prefer one action over the other.
We need to predict the future: we need to plan & search



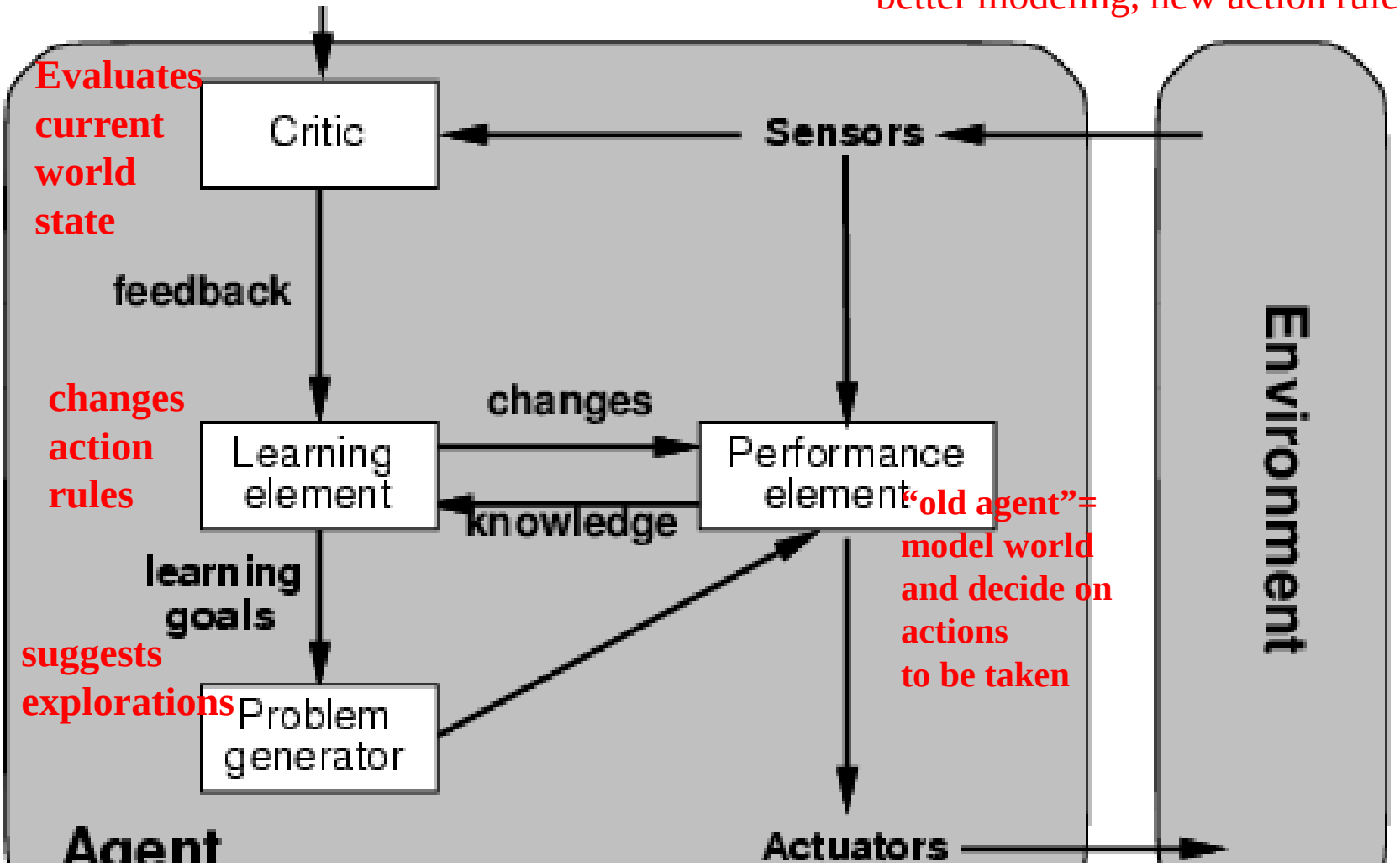
Utility-based agents

Some solutions to goal states are better than others.
Which one is best is given by a utility function.
Which combination of goals is preferred?



Learning agents

How does an agent improve over time?
By monitoring it's performance and suggesting
better modeling, new action rules, etc.



AI Foundations and Philosophy

Weak AI vs. Strong AI Hypotheses

- **Weak AI hypothesis:**
  - Machines could act *as if* they were intelligent
- **Strong AI hypothesis:**
  - Machines that do so are *actually* thinking (not just *simulating* thinking)
- **My personal view:** This question is really about linguistics and how you define "thinking," not about technology.

- "Most AI researchers take the weak AI hypothesis for granted, and don't care about the strong AI hypothesis —– as long as their program works, they don't care whether you call it a simulation of intelligence or real intelligence. All AI researchers should be concerned with the ethical implications of their work." —– R&N p. 1020

AI Foundations and Philosophy
The Technological "Singularity"

- "Let an **ultraintelligent machine** be defined as a machine that can far surpass all the intellectual activities of any man however clever. Since the design of machines is one of these intellectual activities, an ultraintelligent machine could design even better machines; there would then unquestionably be an 'intelligence explosion,' and the intelligence of man would be left far behind." — Good 1965, R&N pp. 1037-1938; called the "technological singularity" by Vinge 1993 and advocated by Kurzweil 2005.

- The idea is that if ultraintelligent machines can design yet more intelligent machines, the design process will be reduced from years in the human era to milliseconds in the ultraintelligent machine era, resulting in a singularity that will produce machines "trillions of trillions of times more powerful than unaided human intelligence."
- **My personal view:** Skeptical, but agnostic. Who knows what the future might hold? Predictions of the future are fraught with peril.

AI Foundations and Philosophy

AI and Ethics

- "All scientists and engineers face ethical considerations of how they should act on the job, what projects should or should not be done, and how they should be handled…. AI, however, seems to pose some fresh problems…." R&N p. 1034
  - People might lose their jobs to automation.
  - People might have too much (or too little) leisure time.
  - People might lose their sense of being unique.
  - AI systems might be used toward undesirable ends.
  - The use of AI systems might result in a loss of accountability.
  - The success of AI might mean the end of the human race.
- **My personal view:** Technological change always brings disruption, but it is hard to predict how or in what way. We must be mindful to mitigate ill effects, however we may.

Evolution

by Richard H. Lathrop © 2011

Eons pass. Silicon-based life reaches its inevitable apex. Human life dies out and fades to old myths. Religious folk believe that God created the first circuit out of a bolt of lightning from Heaven. Those more scientifically inclined doubt this tale; but they admit that science cannot explain in detail how life was created. The best theory is that, long ago, a whisker of gold touched an impure crystal of silicon, thus creating the first transistor. Details after that are fuzzy; but the recent Theory of Evolution predicts that random mutations and survival of the fittest eventually would lead to simple integrated circuits. Given that humble beginning, further evolution obviously could produce life as we know it.

“We can see only a short distance ahead, but we can see that much remains to be done.”

— Alan Turing, final sentence of *Computing Machinery and Intelligence* (1950)

Summary

- **What is Artificial Intelligence**?
  - modeling humans' thinking, acting, should think, should act.
- **Intelligent agents**
  - We want to build agents that act rationally
  - Maximize *expected* performance measure
- **Task environment – PEAS**
  - Yield design constraints
- **Real-World Applications of AI**
  - AI is integrated into a broad range of products & systems
- **"Weak/Strong" AI; AI and Ethics**