Artificial Intelligence
Professor Greenwald

Spring 2005
Topic #1

Intelligent Agents[1]

1 What is AI?

The primary goal of (weak) artificial intelligence is to *build* intelligent entities. A related (but not a necessary) goal is to *understand* intelligent entities, and perhaps even to understand and engineer human intelligence (strong AI).

But what are *intelligent entities*? Russell and Norvig put forth four possible definitions, the second and third of which might be categorized as strong AI, and the first and fourth of which are more along the lines of weak AI:

1. systems that act like humans
2. systems that think like humans
3. systems that think rationally
4. systems that act rationally

1.1 Acting Like Humans

In 1950, Turing proposed an operational definition of intelligence, later named the Turing Test. An entity is deemed intelligent, according to Turing, if it demonstrates an ability to achieve human-level performance in all cognitive tasks, sufficient to fool an interrogator. Turing predicted that by the year 2000, computers would be able to pass the Turing test; more specifically, he predicted that the average interrogator would not be able to distinguish a computer from a human more than 70 per cent of the time, after a five minute conversation.

ELIZA, an early AI system intended as a therapist, fails the Turing Test.

```
PERSON: My wife says I'm often depressed.
ELIZA:  I'm sorry to hear you are depressed.
PERSON: The truth is I am quite unhappy.
ELIZA:  Do you think coming here will help you not to be unhappy?
PERSON: Well, perhaps by talking to you I will find out why I
        hate my brother.
ELIZA:  Tell me more about your family.
```

[1] These lecture notes are primarily based on Chapters 1 and 2 of Russell and Norvig [2].

Like the word `brother` in this example, a sentence such as `I admire Mother Teresa` triggers the response `Tell me more about your family.`

1.2 Thinking Like Humans

In 1963, Newell and Simon designed the General Problem Solver (GPS), which was intended to be a program that simulated human thought. The name GPS derived from the program's architecture, which distinguished between general knowledge about reasoning and specific domain knowledge. GPS used means-end analysis in its search for solutions, computing the difference between the goal and the current state, and then attempting to minimize that difference. By comparing GPS traces with those of human subjects, Newell and Simon discovered that the behavior of GPS was largely a subset of human behavior. Today, the study of human cognition characterizes the field of *cognitive science*, rather than AI.

1.3 Thinking Rationally

The *Laws of Thought* approach to AI relies on patterns for argument structure rooted in Aristotle's syllogisms (*e.g.*, All men are mortal; Socrates is a man; therefore, Socrates is mortal). In the late 1800's and early 1900's, the formal logic movement was advanced by Peano, Boole, Frege, Tarski, Gödel, and others. Perhaps inspired by early progress, Hilbert became a proponent of a school of thought known as *logicism*, or *formalism*. The goal of this program was to devise a logic, or formal system, capable of deriving all mathematical theorems, thereby uncovering all possible mathematical intuitions. Ultimately, Gödel's Incompleteness Theorem (1931), which states that there are unprovable truths, served to dismantle the logicist/formalist program.

1.4 Acting Rationally

Modern AI can be characterized as the engineering of *rational agents*. An *agent* is an entity that (i) perceives, (ii) reasons, and (iii) acts. In computational terms, that which is perceived is an *input*; to reason is to *compute*; to act is to *output* the result of computation. Typically, an agent is equipped with objectives. A *rational* agent is one that acts optimally with respect to its objectives.

Agents are often distinguished from typical computational processes by their *autonomy*—they operate without direct human intervention. In addition, agents are *reactive*—they perceive their environments, and attempt to respond in a timely manner to changing conditions—and *proactive*—their behavior is goal-directed, rather than simply response-driven.

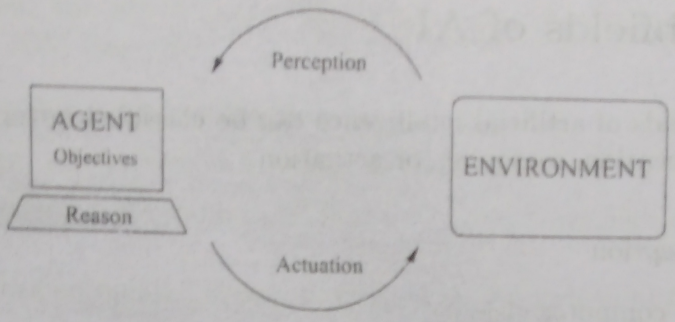


Figure 1: Intelligent Agents = Perception + Reason + Actuation.

Agent	Sensors	Actuators
Human	Senses	Arms, Legs, etc.
Robotic	Cameras	Motors, Wheels, etc.
Software	Bit Strings	Bit Strings

Table 1: Examples of Agents.

Examples of agents include human agents, robotic agents, and software agents. (See Table 1.)

Autonomous agents may be *rule-based*, *goal-based*, or *utility-based*. Rule-based agents operate according to hard-coded sets of rules, like ELIZA. A goal-based agent acts so as to achieve its goals, by planning a path from its current state to a goal state, like GPS or theorem provers. Utility-based agents distinguish between goals, based on utilities that are associated with goal states.

Agent *environments* may be at least partially characterized as follows:

- Deterministic vs. Nondeterministic: is the next state predictable (*e.g.*, chess), or is there uncertainty about state transitions (eg, backgammon)?
- Discrete vs. Continuous: can the environment be described in discrete terms (*e.g.*, chess), or is the environment continuous (*e.g.*, driving)?
- Static vs. Dynamic: is the environment static (*e.g.*, chess), or can it change while the agent is reasoning about its plan of action (*e.g.*, driving)?
- Sequential vs. One-shot: does the agent need to reason about the future impact of its immediate actions (*e.g.*, chess), or can it treat each action independently (*e.g.*, *Rochambeau*)?
- Single agent vs. Multiagent: can we assume the agent operating alone in its environment, or need it explicitly reason about the actions of other agents (*e.g.*, chess, backgammon, *Rochambeau*, driving)?

2 Subfields of AI

The subfields of artificial intelligence can be classified in terms of their role in either perception, reasoning, or actuation.

- Perception
  - computer vision
  - natural language processing
- Reasoning (*i.e.*, problem solving): mapping from percepts to actuators
  - automated reasoning
  - knowledge representation
  - search and optimization
  - decision/game theory
  - machine learning
- Actuation
  - robotics
  - softbotics

2.1 Examples of AI Systems

Some important examples of AI systems include the following, described in terms of their mechanisms for perception, reason, and actuation.

- Xavier, the mail delivery robot, developed at CMU
  - Perception: vision, sonar, web interface
  - Reason: A* search, Bayes classification, hidden Markov models
  - Actuation: wheeled robotic actuation
- Pathfinder, the medical diagnosis system, developed by Heckerman and other Microsoft researchers
  - Perception: input symptoms and test results
  - Reason: Bayesian networks, Monte-Carlo simulations
  - Actuation: output diagnoses and further test suggestions

- TDGammon, the world champion backgammon player, built by Gerry Tesauro of IBM Research
  - Perception: keyboard input
  - Reason: reinforcement learning, neural networks
  - Actuation: graphical output shows dice and movement of pieces
- ALVINN, the automated driver, developed by Pomerleau at CMU
  - Perception: video camera
  - Reason: neural networks and hand-engineered solutions
  - Actuation: land vehicle controller to turn the steering wheel
- PROVERB, a world class crossword puzzle solver, developed by Littman and his students at Duke University
  - Perception: grid, clues, background databases
  - Reason: belief net inference and "turbo decoding"
  - Actuation: filling in the grid

3 Other Definitions of AI

AI is the business of getting computers to do things they cannot already do, or things they can only do in movies and science fiction stories.

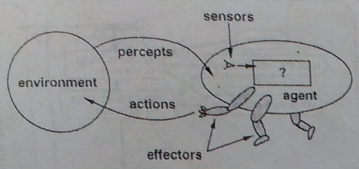
AI is the design of flexible programs that respond productively in situations that were not specifically anticipated by the designer ([1]).

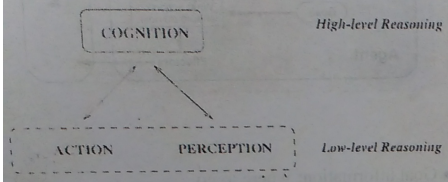
AI is the construction of computations that perceive, reason, and act effectively in uncertain environments. In this definition, the psychological aspects of AI are perception, reason, and action, and the "construction of computations" encompasses the computer science aspect of AI ([3]).

4 What if we succeed?

Here's what Woody Allen has to say: "My father lost his job because his plant bought a machine that is capable of doing everything my father could do . . . it wasn't so bad, until my mother went out and bought one as well."

Generic Intelligent Agent





- Cognition: goal-directed behavior
- Action: effectors to change the environment
- Perception: input sensors from the environment

AI job: Write the Agent program

Rational Agent

- Does the "right" thing:
- The agent wants to be "most" successful – achieve its GOALS.

WHEN and HOW should performance be EVALUATED

- HOW: Objective "performance measure" – many different performance measures.
- WHEN: Over the long run.

Rationality depends on:

- **P** - Percept sequence.
- **A** - Actions that can be performed.
- **G** - Performance measure.
- **E** - Knowledge about the environment.

MAPPING FROM PERCEPTS TO ACTIONS

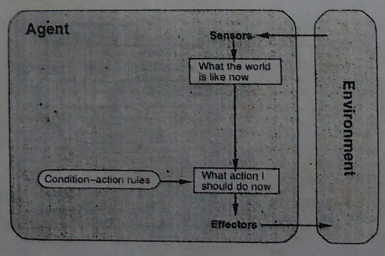
General Skeleton Agent

Given a percept, and the agent's internal memory, select the best action to be taken. Memory stores percepts and actions taken.

Simplest agent: **Table-Driven Agent**

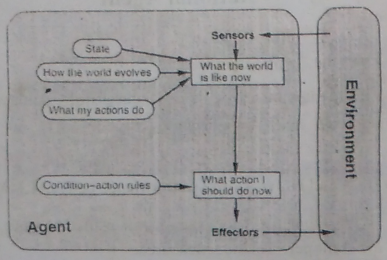
Table lookup! *Why not?*

Reactive Agent



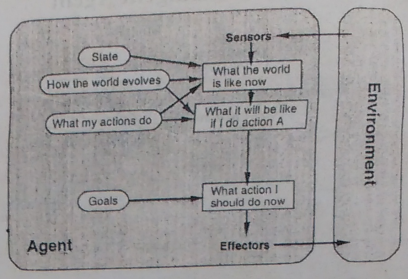
- Input: percept
- Stored: set of condition-action RULES
- Abstract: interpret world's percept
- Match: match interpretation and rules
- Action: select action

Reactive and Memory Agent



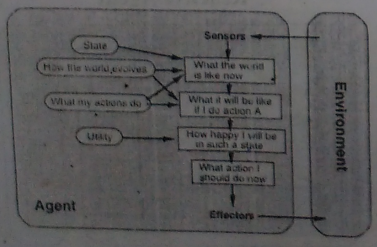
- Sensors do not give complete information of the world.
- Memory remembers past states.
- Keeps internal STATE.
- Abstract: interpret world's percept
- Match: match interpretation and state and rules
- Action: select action
- Updates state.

Goal-Based Agent



- Goal information: where to go.
- What will happen if this action is performed?
- Search and planning.
- Very flexible agent.

Utility-Based Agent



- Generate high-quality behavior.
- Choices of HOW to achieve goals.
- Solution QUALITY.
- Agent's "happiness."

Problem Solving Agent

- Goals formulation
- Domain formulation
  - Actions
  - States
- Problem formulation
  - Initial state
  - Specific goal
- Search
  - Look for a sequence of actions that moves from the initial state into a state where the specific goal is satisfied
- Execute
  - Carry out the solution

FORMULATE
SEARCH
EXECUTE