Plan for the 2nd hour

- What is an agent?
- PEAS (Performance measure, Environment, Actuators, Sensors)
- Agent architectures.
- Environments
- Multi-agent systems.

Acting humanly: The Turing test

Turing (1950) “Computing machinery and intelligence”:
- Can machines think? → Can machines behave intelligently?
- Operational test for intelligent behavior: the *Imitation Game*

<table>
<thead>
<tr>
<th>Systems that think like humans</th>
<th>Systems that act like humans</th>
</tr>
</thead>
</table>

- Loebner prize
- Anticipated all major arguments against AI in last 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 humanly: cognitive science

1960s “cognitive revolution”: information-processing psychology replaced the then prevailing orthodoxy of behaviorism.
Requires scientific theories of internal activities of the brain:
- What level of abstraction? “Knowledge” or “circuits”?
- How to validate? Requires
  - Predicting and testing behavior of human subjects (top-down),
  - or 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!

Thinking rationally: laws of thought

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:
- Not all intelligent behavior is mediated by logical deliberation
- What is the purpose of thinking? What thoughts should I have out of all the thoughts (logical or otherwise) that I could have?

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.
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:

\[ f : P^* \rightarrow A \]

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. → design best program for given machine resources.

The vacuum-cleaning world

Percepts: location and contents, e.g. \(< A, Dirty >\)
Actions: Left, Right, Suck, NoOp

A vacuum-cleaning 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
\(\ldots\) | \(\ldots\)
A vacuum-cleaning agent

<table>
<thead>
<tr>
<th>Percept sequence</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; A, Clean &gt;</td>
<td>Right</td>
</tr>
<tr>
<td>&lt; A, Dirty &gt;</td>
<td>Suck</td>
</tr>
<tr>
<td>&lt; B, Clean &gt;</td>
<td>Left</td>
</tr>
<tr>
<td>&lt; B, Dirty &gt;</td>
<td>Suck</td>
</tr>
<tr>
<td>&lt; A, Clean &gt;, &lt; A, Clean &gt;</td>
<td>Right</td>
</tr>
<tr>
<td>&lt; A, Clean &gt;, &lt; A, Dirty &gt;</td>
<td>Suck</td>
</tr>
</tbody>
</table>

function Reflex_Vacuum_Agent (location, status)
if status == Dirty then return Suck
if location == A then return Right
if location == B then return Left

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?

What is the RIGHT function?
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 is not omniscient as percepts may not supply all relevant information.

A rational agent

[Wooldridge, 2000]

An agent is said to be rational if it chooses to perform actions that are in its own best interests, given the beliefs it has about the world.

Properties of rational agents:
- Autonomy (they decide);
- Proactiveness (they try to achieve their goals);
- Reactivity (they react to changes in the environment);
- Social ability (they negotiate and cooperate with other agents).
**PEAS**

- PEAS: Performance measure, Environment, Actuators, Sensors
- Must first specify the setting for intelligent agent design
- Consider, e.g., the task of designing an automated taxi driver:
  - Performance measure
  - Environment
  - Actuators
  - Sensors

---

**AUTOMATED TAXI DRIVER:**

- Performance measure: Safe, fast, legal, comfortable trip, maximize profits
- Environment: Roads, other traffic, pedestrians, customers
- Actuators: Steering, accelerator, brake, signal, horn
- Sensors: Cameras, radars, speedometer, GPS, odometer, engine sensors, car-human interface

---

**Autonomous agents**

Can make decisions on their own.

Why do they need to? Because of the following properties of real environments (cf. Russell and Norvig):

- the real world is inaccessible (partially observable);
- the real world is nondeterministic (stochastic, sometimes strategic);
- the real world is nonepisodic (sequential);
- the real world is dynamic (non-static);
- the real world is continuous (non-discrete).

---

**Agent taxonomy**

- simple reflex agents
- reflex agents with state
- goal-based agents
- utility-based agents
- learning agents - independent property from the list above
Agents

Simple reflex agent

Agent
<table>
<thead>
<tr>
<th>Sensors</th>
<th>Environment</th>
</tr>
</thead>
<tbody>
<tr>
<td>What the world is like now</td>
<td></td>
</tr>
<tr>
<td>Condition-action rules</td>
<td></td>
</tr>
<tr>
<td>What action I should do now</td>
<td></td>
</tr>
<tr>
<td>Actuators</td>
<td></td>
</tr>
</tbody>
</table>

Reflex agent with state

Agent
<table>
<thead>
<tr>
<th>State</th>
</tr>
</thead>
<tbody>
<tr>
<td>How the world evolves</td>
</tr>
<tr>
<td>What the world is like now</td>
</tr>
<tr>
<td>Condition-action rules</td>
</tr>
<tr>
<td>What action I should do now</td>
</tr>
<tr>
<td>Actuators</td>
</tr>
</tbody>
</table>

Goal-based agent

Agent
<table>
<thead>
<tr>
<th>State</th>
</tr>
</thead>
<tbody>
<tr>
<td>How the world evolves</td>
</tr>
<tr>
<td>What the world is like now</td>
</tr>
<tr>
<td>What my actions do</td>
</tr>
<tr>
<td>What it will be like if I do action A</td>
</tr>
<tr>
<td>Goals</td>
</tr>
<tr>
<td>What action I should do now</td>
</tr>
<tr>
<td>Actuators</td>
</tr>
</tbody>
</table>

Utility-based agent

Agent
<table>
<thead>
<tr>
<th>State</th>
</tr>
</thead>
<tbody>
<tr>
<td>How the world evolves</td>
</tr>
<tr>
<td>What the world is like now</td>
</tr>
<tr>
<td>What my actions do</td>
</tr>
<tr>
<td>What it will be like if I do action A</td>
</tr>
<tr>
<td>Utility</td>
</tr>
<tr>
<td>How happy I will be in such a state</td>
</tr>
<tr>
<td>What action I should do now</td>
</tr>
<tr>
<td>Actuators</td>
</tr>
</tbody>
</table>
Rationality is a very powerful assumption.

It allows us to compute things we wouldn’t otherwise be able to dream of!

30+ first years of AI were based solely on this assumption.

Subsumption: Rodney Brooks, 1985

- situatedness
  “the world is its own best model”
- embodiment
- intelligence
  “intelligence is determined by the dynamics of interaction with the world”
- emergence
  “intelligence is in the eye of the observer”
Agents interact with environments through actuators and sensors

The agent function describes what the agent does in all circumstances.

The performance measure evaluates the environment sequence.

A perfectly rational agent maximizes expected performance.
Agents interact with environments through actuators and sensors. The agent function describes what the agent does in all circumstances. The performance measure evaluates the environment sequence. A perfectly rational agent maximizes expected performance. Agent programs implement (some) agent functions.

Environments are categorized along several dimensions: observable? deterministic? episodic? static? discrete? single-agent?

Several basic agent architectures exist: reflex, reflex with state, goal-based, utility-based.

PEAS descriptions define task environments.