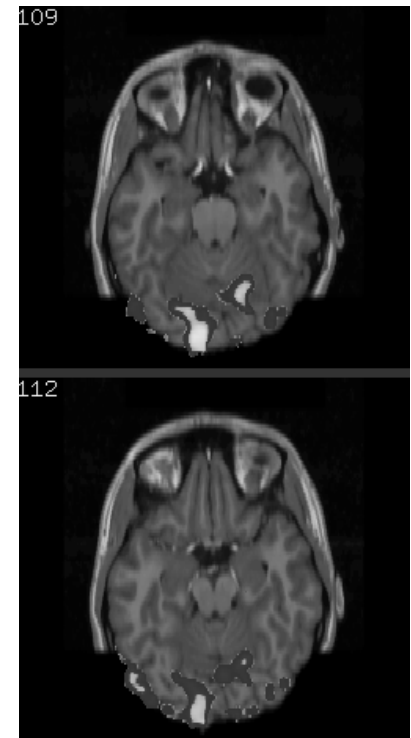# What is AI?

The science of making machines that:

Think like humans

Thinking Like Humans?

- ## The cognitive science approach:
  - 1960s ``cognitive revolution'': information-processing psychology replaced prevailing orthodoxy of behaviorism

- ## Scientific theories of internal activities of the brain
  - What level of abstraction? "Knowledge" or "circuits"?
  - **Cognitive science:** Predicting and testing behavior of human subjects (top-down)
  - **Cognitive neuroscience:** Direct identification from neurological data (bottom-up)
  - Both approaches now distinct from AI
  - Both share with AI the following characteristic:
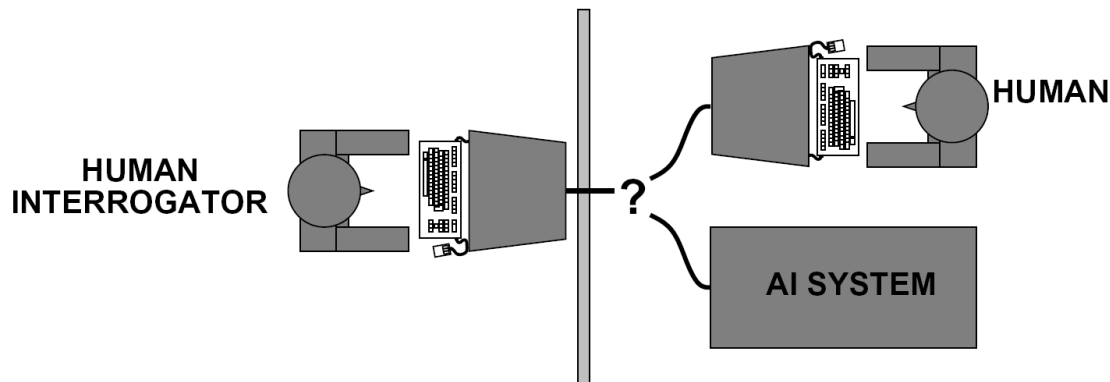   *The available theories do not explain (or engender) anything resembling human-level general intelligence*

Images from Oxford fMRI center

What is AI?

The science of making machines that:

Think like humans
Act like humans

Acting Like Humans?

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

  - Predicted by 2000, 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 or amenable to mathematical analysis**

What is AI?

The science of making machines that:

Think like humans	Think rationally
Act like humans	

Thinking Rationally?

- The "Laws of Thought" approach
  - What does it mean to "think rationally"?
  - Normative / prescriptive rather than descriptive

- Logicist tradition:
  - Logic: notation and rules of derivation for thoughts
  - Aristotle: what are correct arguments/thought processes?
  - Direct line through mathematics, philosophy, to modern AI

- Problems:
  - Not all intelligent behavior is mediated by logical deliberation
  - What is the purpose of thinking? What thoughts should I (bother to) have?
  - Logical systems tend to do the wrong thing in the presence of **uncertainty**

What is AI?

The science of making machines that:

Think like humans	Think rationally
Act like humans	Act rationally

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
  - Thinking can be in the service of rational action
  - Entirely dependent on goals!
  - Irrational ≠ insane, irrationality is sub-optimal action
  - Rational ≠ successful

- **Our focus here: rational agents**
  - Systems which make the best possible decisions given goals, evidence, and constraints
  - In the real world, usually lots of uncertainty
    - … and lots of complexity
  - Usually, we're just approximating rationality

- **"Computational rationality"**

Rational Decisions

We'll use the term **rational** in a particular way:

- Rational: maximally achieving pre-defined goals
- Rational only concerns what decisions are made

  (not the thought process behind them)
- Goals are expressed in terms of the **utility** of outcomes
- Being rational means **maximizing your expected utility**
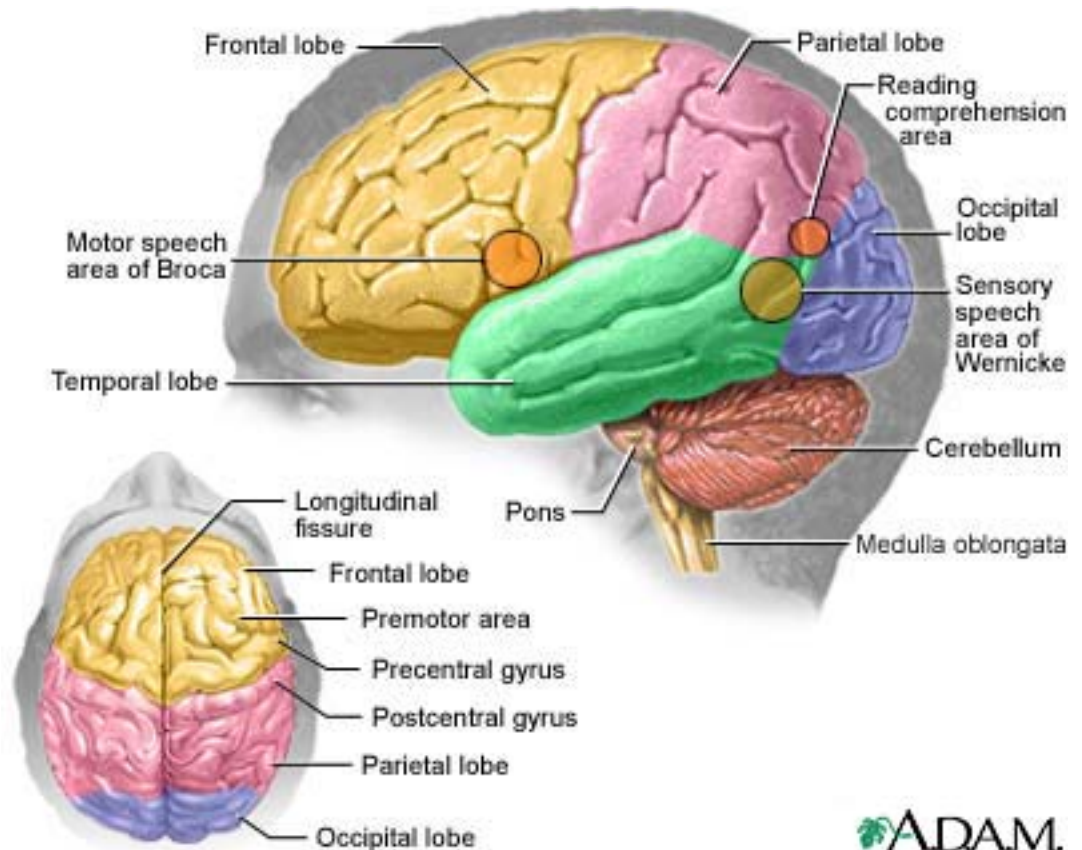
A better title for this course would be:

Computational Rationality

Acting Rationally

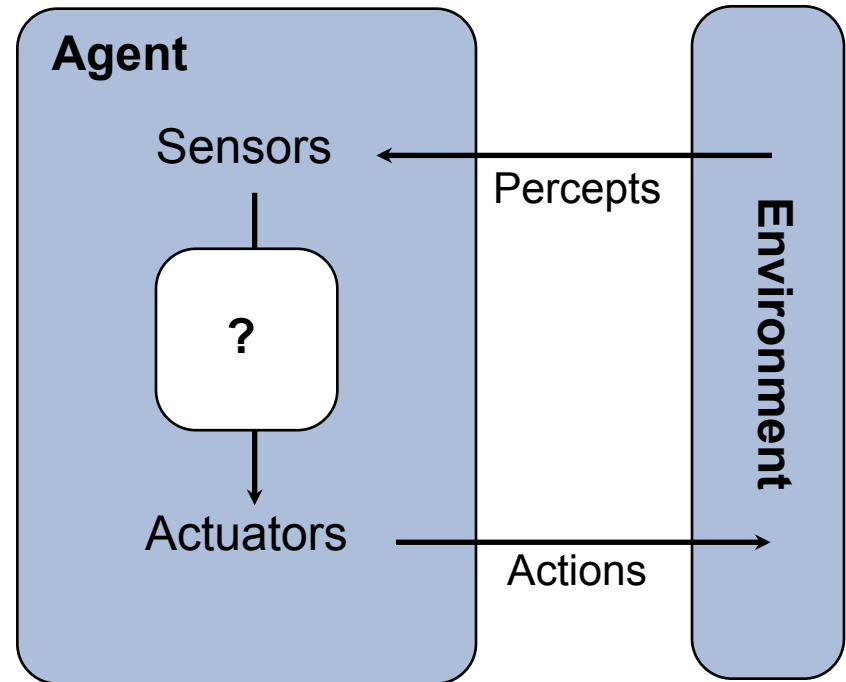Maximize your expected utility.

What about the brain?

- Brains (human minds) are very good at making rational decisions (but not perfect)

- Brains aren't as modular as software

- "Brains are to intelligence as wings are to flight"

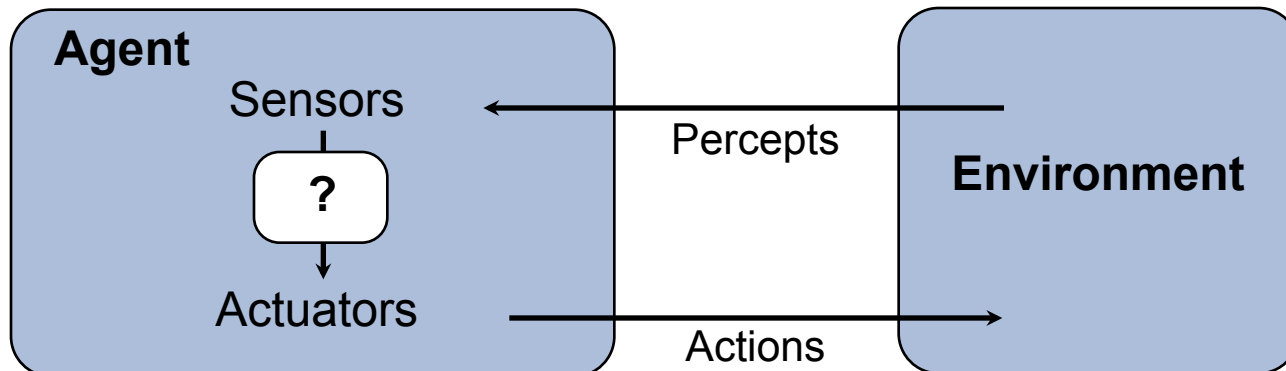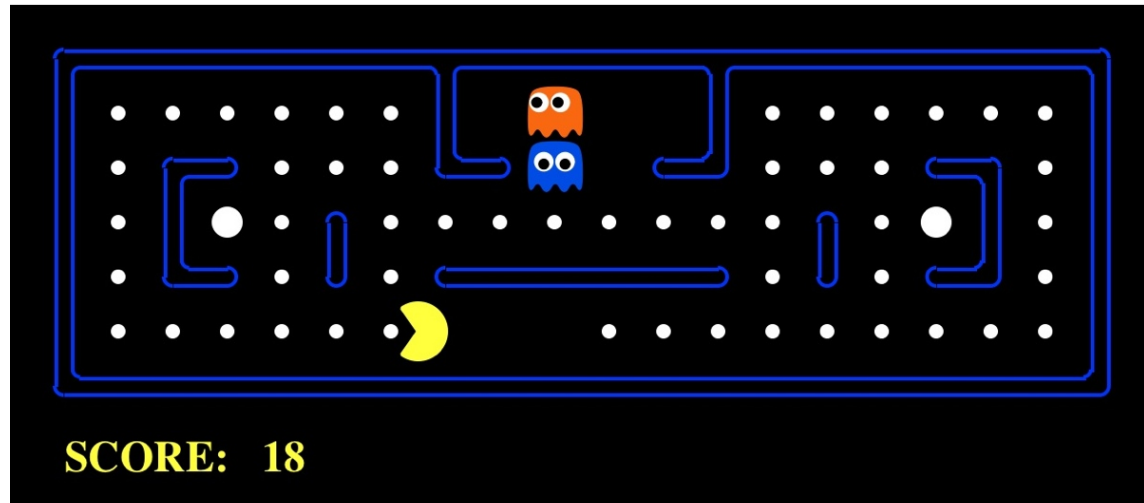- Lessons learned: **prediction** and **simulation** are key to decision making

Designing Rational Agents

- An **agent** is an entity that *perceives* and *acts*.

- A **rational agent** selects actions that maximize its **utility function**.

- Characteristics of the **percepts, environment,** and **action space** dictate techniques for selecting rational actions.

Agent

Sensors

?

Actuators

Percepts

Actions

Environment

- This course is about:
  - General AI techniques for a variety of problem types
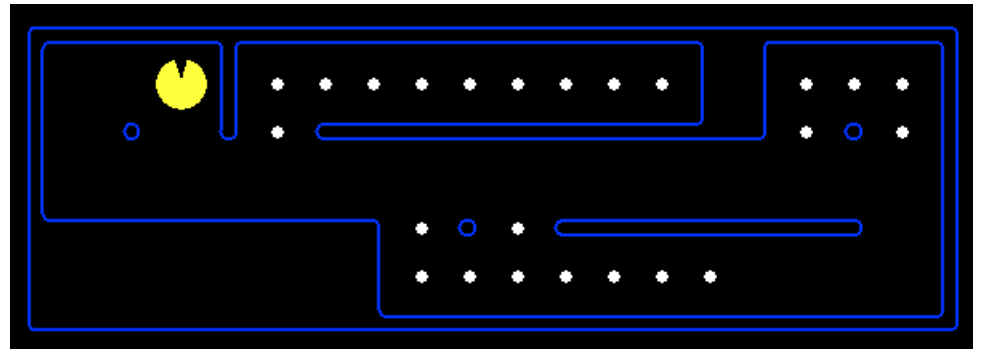 - Learning to recognize when and how a new problem can be solved with an existing technique
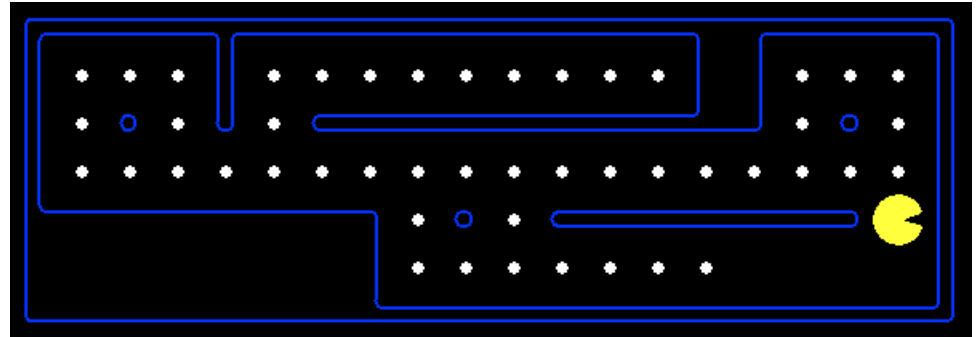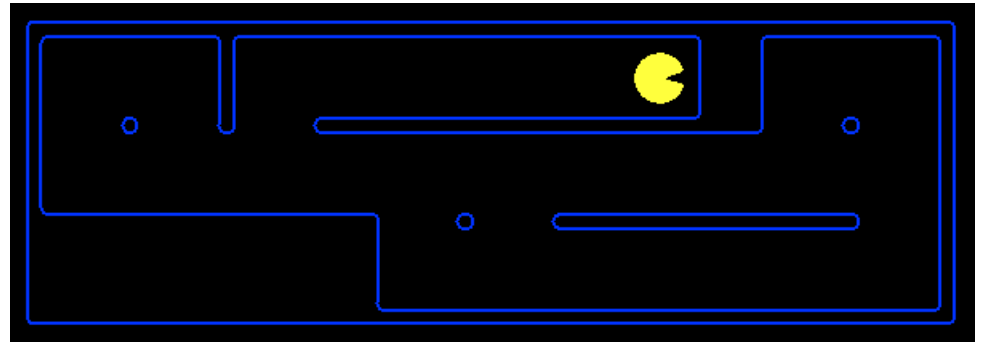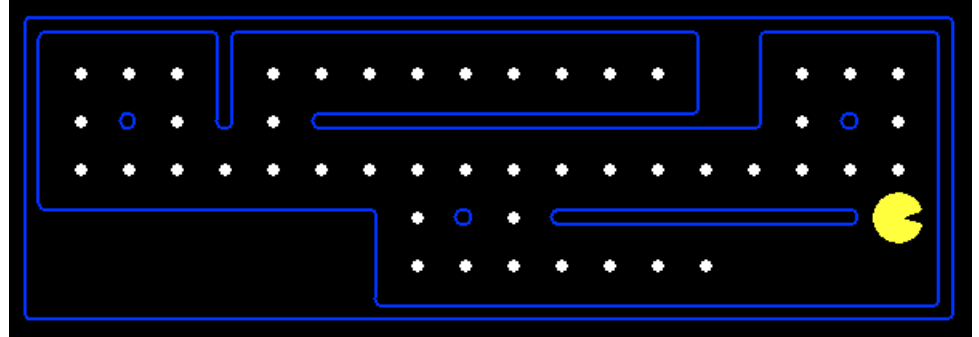
Pacman as an Agent

SCORE: 18

Agent		Environment
Sensors ←	Percepts	
?		
Actuators →	Actions	

Reflex Agents

- **Reflex agents:**
  - Choose action based on current percept (and maybe memory)
  - May have memory or a model of the world's current state
  - Do not consider the future consequences of their actions
  - Consider how the world IS

- **Can a reflex agent be rational?**

Planning Agents

- Plan ahead
- Ask "what if"
- Decisions based on (hypothesized) consequences of actions
- Must have a model of how the world evolves in response to actions
- Consider how the world WOULD BE

Quiz: Reflex or Planning?

Select which type of agent is described:

1. Pacman, where Pacman is programmed to move in the direction of the closest food pellet

2. Pacman, where Pacman is programmed to move in the direction of the closest food pellet, unless there is a ghost in that direction that is less than 3 steps away.

3. A navigation system that first considers all possible routes to the destination, then selects the shortest route.

AI Adjacent Fields

- 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 and statistics
- Psychology
    - Adaptation
    - Phenomena of perception and motor control
    - Experimental techniques (psychophysics, etc.)
- Economics: formal theory of rational decisions
- Linguistics: knowledge representation, grammar
- Neuroscience: physical substrate for mental activity
- Control theory:
    - homeostatic systems, stability
    - simple optimal agent designs