(version without (potentially) copyrighted images)

CS494/594: Artificial Intelligence

Fall 2009
Tuesday/Thursday, 12:40 – 1:55

Instructor: Dr. Lynne E. Parker

TA: Nick Overfield

"*Artificial Intelligence is the study of how to make real computers act like the ones in the movies*."
--Anonymous

Outline

- Overview syllabus and class policies
- Course Overview
- Introduction to AI (Chapter 1)
  - What is AI?
  - A brief history
  - The state of the art
- Intelligent Agents (Chapter 2.1-2.3)
  - Agents and environments
  - Rationality
  - PEAS (Performance measure, Environment, Actuators, Sensors)
  - Environment types
  - Agent types (next time)

Overview of Syllabus and Class Policies

(See handout)

Course Overview

- Introduction to AI
- Intelligent Agents
- Problem-solving by search
- Logical systems
- Planning systems
- Uncertainty – probability and decision theory
- Learning
- Perception and robotics
- Philosophical issues

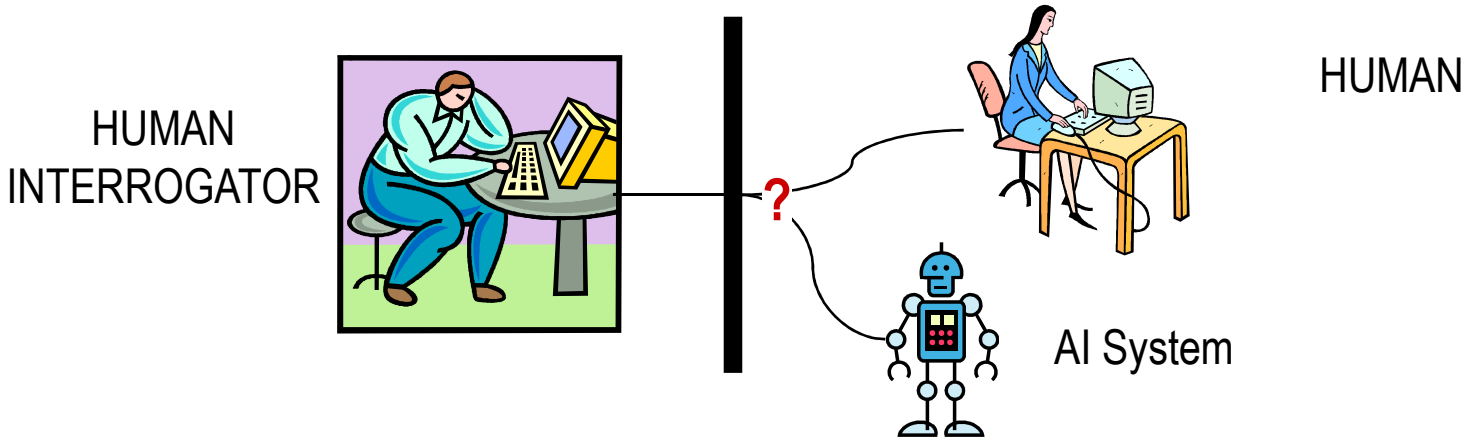
What is AI?

Systems that think like humans	Systems that think rationally
"The automation of activities that we associate with human thinking – activities such as decision-making, problem solving, learning, …" (Bellman, 1978)	"The study of mental faculties through the use of computational models." (Charniak and McDermott, 1985)
Systems that act like humans	**Systems that act rationally**
"The art of creating machines that perform functions that require intelligence when performed by people", (Kurzweil, 1990)	"AI…is concerned with intelligent behavior in artifacts." (Nilsson, 1998)

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



- 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 6 major components of AI: knowledge representation, automated reasoning, natural language understanding, machine learning, computer vision, robotics

Problem: Turing test is not reproducible, constructive, or amenable to mathematical analysis

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)
    2) 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

- 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?

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

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

Foundations of AI

- Philosophy (428 BC – Present)
  - Can formal rules be used to draw valid conclusions?
  - How does the mental mind arise from a physical brain?
  - Where does knowledge come from?
  - How does knowledge lead to action?

- Aristotle, Leonardo da Vinci, Pascal, Descartes, etc.

Foundations of AI (con't.)

- Mathematics (~800 – present)
  - What are the formal rules to draw valid conclusions?
  - What can be computed?
  - How do we reason with uncertain information?

- Logic, computation (algorithms), probability

Foundations of AI (con't.)

- Economics (1776-present)
  - How should we make decisions so as to maximize profit?
  - How should we do this when others may not go along?
  - How should we do this when the payoff may be far in the future?

- Utility, decision theory, game theory, operations research, Markov decision processes

Foundations of AI (con't.)

- Neuroscience (1861-present)
  - How do brains process information?

- Moore's law predicts that CPU's gate count will equal brain's neuron count around 2020.
- *But, even though a computer is a million times faster in raw switching speed, the brain is actually 100,000 times faster at what it does.*

Foundations of AI (con't.)

- Psychology (1879 – present)
  - How do humans and animals think and act?

- Behaviorism, cognitive psychology, cognitive science

Foundations of AI (con't.)

- Computer engineering (1940 – present)
  - How can we build an efficient computer?

- AI requires: (1) intelligence, (2) an artifact (i.e., a computer upon which the intelligence is generated)

Foundations of AI (con't.)

- Control theory and Cybernetics (1948 – present)
  - How can artifacts operate under their own control?

- Control theory: Maximizing an objective function over time
  - Uses calculus and matrix algebra, which lend themselves to systems that are describable by fixed sets of continuous variables;
    - Exact analysis typically feasible only for linear systems
- AI: Designing systems that behave optimally
  - Founded as a way to "escape" from limitations of the mathematics of control theory
    - Use of logical inference and computation allows AI to consider problems such as language, vision, and planning, which are outside the field of control theory

Foundations of AI (con't.)

- Linguistics (1957-present)
  - How does language relate to thought?

- Computational linguistics, natural language processing, knowledge representation

Summary of 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
Economics	-- formal theory of rational decisions
Neuroscience	-- plastic physical substrate for mental activity
Psychology	-- adaptation
	-- phenomena of perception and motor control
	-- experimental techniques (psychophysics, etc.)
Control theory	-- homeostatic systems, stability
	-- simple optimal agent designs
Linguistics	-- knowledge representation
	-- grammar

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
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": Artificial life, GAs, soft computing
1995-	Agents, agents everywhere …

What can AI do today?

- Autonomous planning and scheduling:
  - NASA's Remote Agent program became 1st onboard autonomous planning program to control the scheduling of operations for a spacecraft

What can AI do today?

- Game playing:
  - IBM's Deep Blue became 1st computer to defeat world champion in a chess match

What can AI do today?

- Autonomous control:
  - ALVINN computer vision system was trained to steer a car and keep it following in a lane; used to drive the CMU NavLab minivan across U.S. (98% of the time)

What can AI do today?

- Diagnosis:
  - Medical diagnosis programs based on probabilistic analysis have been able to perform at the level of an expert physician in several areas of medicine

What can AI do today? (con't.)

- Logistics Planning:
  - U.S. military deployed a Dynamic Analysis and Replanning Tool (DART) in 1991, for automated logistics planning and scheduling, generating plans in hours that previously would have taken weeks

What can AI do today? (con't.)

- Robotics:
  - Many surgeons now use robotic devices in surgery (e.g., da Vinci robot)

What can AI do today? (con't.)

- Language understanding and problem solving:
  - PROVERB (1999) is a computer program that can solve crossword puzzles better than most humans

State of the art

"Thought Discussion" for next class:

Which of the following can currently be done autonomously (by intelligent machine or agent)?

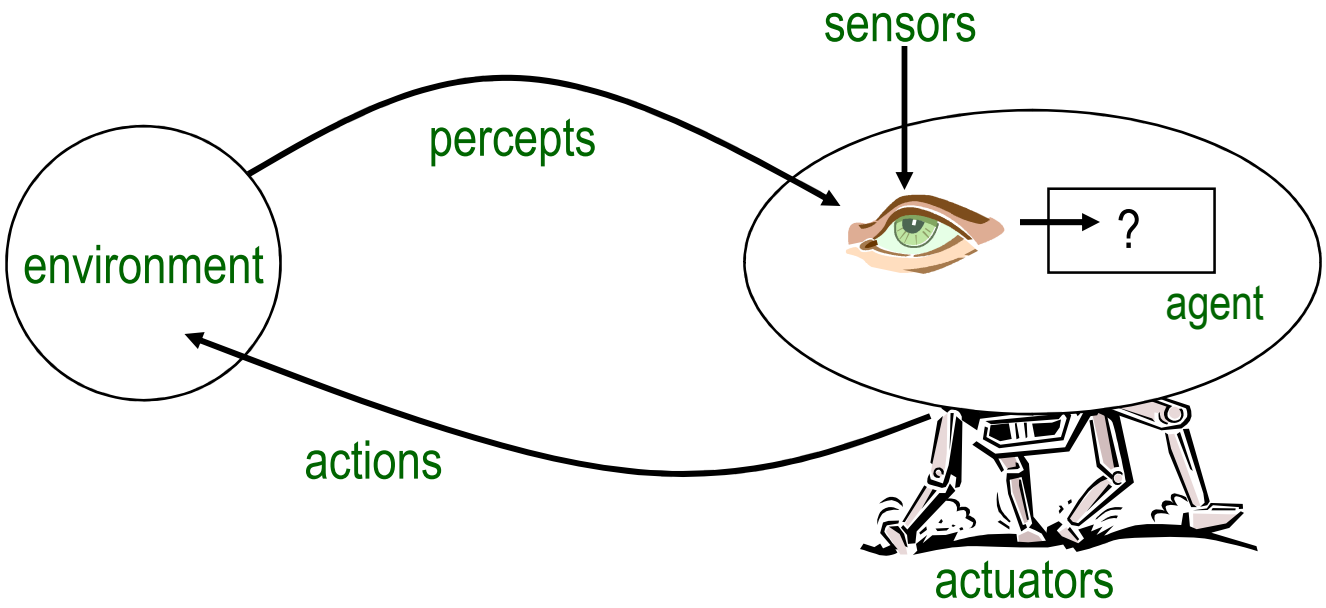
- Play a decent game of table tennis
- Drive along a curving mountain road
- Drive in the center of Cairo
- Buy a week's worth of groceries at Kroger
- Buy a week's worth of groceries on the web
- Play a decent game of bridge
- Discover and prove a new mathematical theorem
- Write an intentionally funny story
- Give a competent legal advice in a specialized area of law
- Translate spoken English into spoken Swedish in real time
- Perform a complex surgical operation

Your assignment for next time: Research these topics for discussion! What are the difficulties? When do you predict they will be overcome?

Intelligent Agents

- Outline:
  - Agents and environments
  - Rationality
  - PEAS (Performance measure, Environment, Actuators, Sensors)
  - Environment types
  - Agent types

Agents and environments

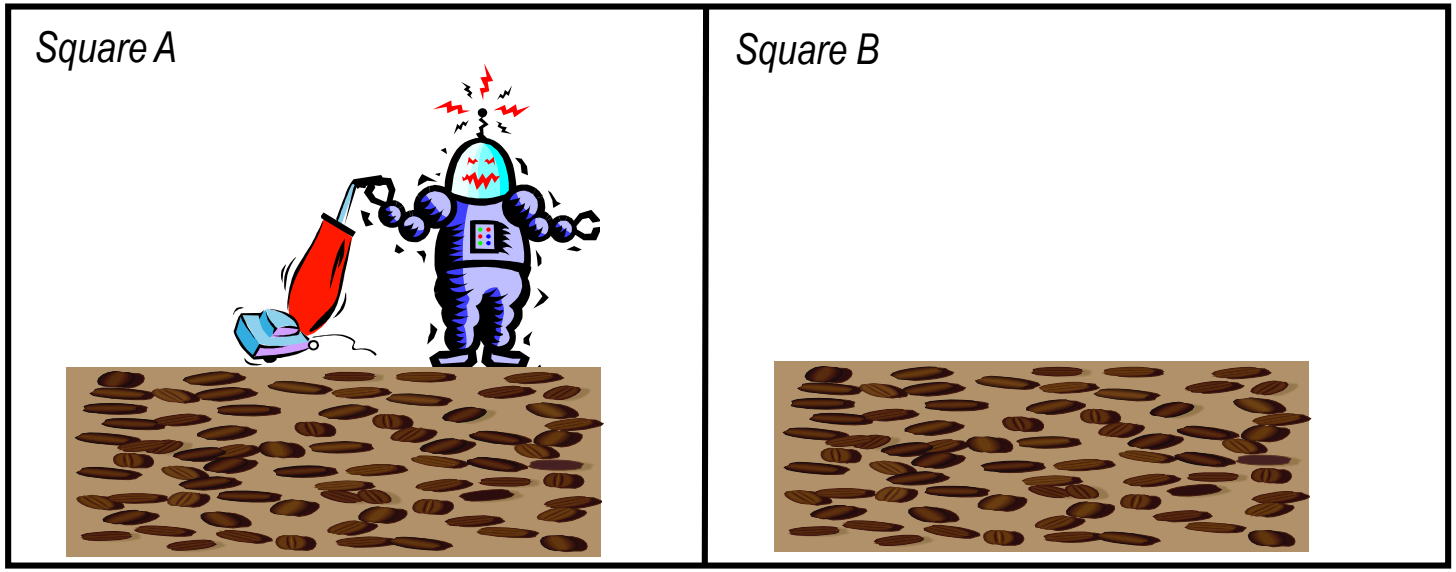


- Agents include humans, robots, softbots, thermostats, etc.
- The agent function maps from percept histories to actions

$$f : P^* \rightarrow A$$

- 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 correct function?

Can it be implemented in a small agent program?

Rationality

- Fixed performance measure evaluates the environment sequence
  - Most dirt cleaned up in time T?
  - 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 and its prior knowledge

- Rational $\neq$ omniscient
- Rational $\neq$ clairvoyant
- Rational $\neq$ successful

- Rational $\Rightarrow$ exploration, learning, autonomy

Next time…

- Agent types

- And remember “Thought Discussion” for next time:

  State of the Art in AI – what currently can, and can’t, be done.