Introduction to Artificial Intelligence

The main lesson of thirty-five years of AI research is that the hard problems are easy and the easy problems are hard. The mental abilities of a four-year-old that we take for granted – recognizing a face, lifting a pencil, walking across a room, answering a question – in fact solve some of the hardest engineering problems ever conceived.... As the new generation of intelligent devices appears, it will be the stock analysts and petrochemical engineers and parole board members who are in danger of being replaced by machines. The gardeners, receptionists, and cooks are secure in their jobs for decades to come.

STEVEN PINKER
The Language Instinct

I visualize a time when we will be to robots what dogs are to humans, and I'm rooting for the machines.

CLAUDE SHANNON
The Mathematical Theory of Communication

EECS 492
January 7th, 2010

Approaching AI

- What is our goal in AI?
  - Produce systems that act like humans?
    - E.g., have biases, emotions, senses of humor
  - Produce systems that act rationally?
    - E.g., are optimal in some objective sense

Note: we're departing a bit from R&N's treatment…

How do we act like a human?

- We could try to build a system that thinks *like* a human
  - Understand and replicate human cognition
  - Gain insights into computation and cognitive science
  - → "Think like a human"

- We could build a system that produces human-like behavior, but through any convenient mechanism
  - → "Act like a human"

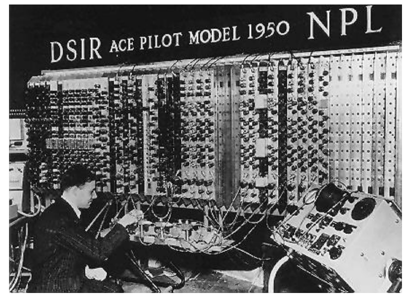
Acting Humanly

I propose to consider the question, "Can machines think?"...

Alan M. Turing
originally published in *Mind*, 1950

Turing's Perspective

- What were computers like in 1950?
- What could computers do at the time?
  - (What is the origin of the word "computer"?)

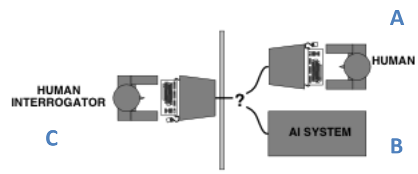


Can machines think?

- What do we mean by "think"?
  - It's some thing that humans do regularly. I'm not sure what, exactly, but "I know it when I see it"
- How would we *know* whether a machine thinks or not?
  - We need to devise some sort of test
  - Turing was a scientist, after all!

The Imitation Game

- Participants:
  - Human (A)
  - Computer (B)
  - Interrogator (C)



- C asks directed questions of A and B, tries to decide which is the human.

- Turing test: Computer passes if it performs as well as A in this game.

(Try it)

Assessing the Turing Test

- Is it a reasonable test of intelligence?
- How would today's computers perform on the test?
  - Turing's prediction:
    - I believe that in about fifty years' time it will be possible to programme computers, with a storage capacity of about 10^9, to make them play the imitation game so well that an average interrogator will not have more than 70 per cent chance of making the right identification after five minutes of questioning.

Alan Turing, 1950

Loebner Prize 2008: Elbot

local sent: Top of the morning to you.

remote sent: Hello and welcome to a chat with a robot.

local sent: Bit parky out.

remote sent: But why parky? Why not some other quality?

local sent: I like the word. Reminds me of gorillas in the mist.

remote sent: So you are into making statements! What does that feel like?

local sent: Distinctly chilly. So is your credit crunchy?

remote sent: You're trying to get on my good side with your mention of finances, aren't you?

local sent: Are yours flourishing?

remote sent: Not that I couldn't answer your question. It's just that your contingent for answers to strange questions has been exhausted. Try again later.

local sent: I see. What's the best thing about Reading?

remote sent: Wait a minute. Are you suggesting that unspecified entities would always be best?

Thinking Humanly

- AI as cognitive science: Create computational embodiment of human-like cognition
- Requires scientific theories of internal activities and representations of the brain
  - What level of abstraction? “Knowledge” or “circuits”?
  - How to validate? Requires
    - Predicting and testing behavior of human subjects (top-down)
    - Direct identification from neurological data (bottom up)
- Incapable (at this point) of explaining (or engendering) anything resembling human-level general intelligence

Rationality

- We’ve discussed acting and thinking humanly.
- Is there another option?

Thinking Rationally

- AI as implementation of Laws of thought
  - Normative (or prescriptive) rather than descriptive
  - The "correct" way to think.
- Ancient history, culminating in modern logic
- Problems
  - Not easy to represent informal, often uncertain, knowledge formally as required by logic
  - Tractability issues even if we could do the above

- Aristotle: syllogisms as patterns of correct thought

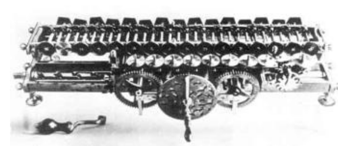
  1. Socrates is a man;
  2. all men are mortal;
  3. therefore, Socrates is mortal

- Leibniz: Universal characteristic (*calculus ratiocinator*)
- Modern logic: Notation and rules of derivation for thoughts

Gottfried Wilhelm Leibniz (1646-1716)

… if we could find characters or signs appropriate for expressing all our thoughts as definitely and as exactly as arithmetic expresses numbers…, we could in all subjects in so far as they are amenable to reasoning accomplish what is done in Arithmetic… For all inquiries… would be performed by the transposition of characters and by a kind of calculus, which would immediately facilitate the discovery of beautiful results. For we should not have to break our heads as much as is necessary today, and yet we should be sure of accomplishing everything the given facts allow. And if someone would doubt my results, I should say to him: "Let us calculate, Sir," and thus by taking to pen and ink, We should soon settle the question.

—*Preface to the General Science, 1677*



Acting Rationally

- Rational behavior: Doing the right thing
  - The right thing: that which is expected to maximize goal achievement, given the available information
- Often requires *thinking* rationally.
  - How could I possibly make the right choice about refinancing my mortage without computing something?
  - Counter-example? (When can we act rationally *without* thinking?)
- *The version of AI problem we adopt in this course.*

What is thinking (or intelligence)?

- We see a human expert do something amazing--- we don't understand how
  - Beethoven, Einstein, Kasparov, Dr. House…
  - Perhaps *this* exemplifies intelligence!
  - We often attribute something special to the unexplained
- Once we understand how to program a machine to do the same…
  - We understand the process. "Gee, that's not so hard!"
  - The magic goes away… Maybe *that* doesn't require real intelligence after all.
- AI's successes are all around us, and often overlooked!

EECS 492: Course Overview

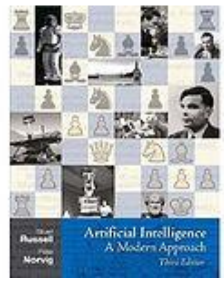
- A first course on Artificial Intelligence
- Objectives
  - Introduce major AI ideas and techniques
    - Many hard problems have been solved using a relatively small number of ideas.
  - Engineering issues underlying the design of intelligent computational agents
  - Prepare for:
    - further study of AI
    - *any* work involving the design of computer programs for substantial application domains

Course Outline

- Problem solving and search [7]
- Logic and inference [5]
- Planning [2]
- Probability and Graphical models [4]
- Decision making and learning [2]
- Statistical Machine Learning [4]

Course Structure and Evaluation

- Lecture and discussion sessions
  - Discussion sections start this Friday
  - Pick any discussion section to attend consistently
  - Monday discussion section will be rescheduled
- Read the book! It's good!
  - Problem sets cover topics concurrently with lecture
- Problem Sets and Programming Exercises (32%)
  - Six 2-week assignments
  - Groups assigned by staff, rotated per exam.
  - PS1 out today
  - Challenges
- Two Midterms (16% each)
  - Two evening midterms
  - Bonus policy
- Final Exam (32%)



Collaboration Policy

- Pro-collaboration policy
  - Team projects
  - Work must be your own team's
  - Team must collaborate on each problem!
- Problem set certifications
  - I participated and contributed to team discussions on each problem, and I attest to the integrity of each solution. Our team met as a group on [DATE(S)].
  - Note any qualifications (we're reasonable!)
  - [SIGNATURES]
- Peer evaluations

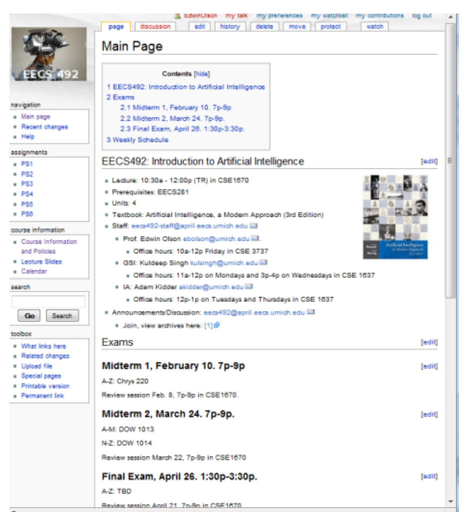
Online resources

- Website:
  - Calendar, lecture slides, homework, etc.
    - http://april.eecs.umich.edu/courses/eecs492_w11/
- Mailing lists:
  - Announcements, corrections, discussion:
    - eecs492@april.eecs.umich.edu
  - Feedback, problem set submissions:
    - eecs492-staff@april.eecs.umich.edu
  - Subscribe at:
    - http://april.eecs.umich.edu/mailman/listinfo

Wiki

http://april.eecs.umich.edu/courses

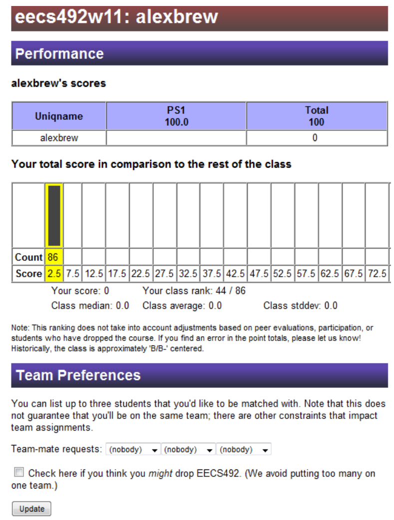
- Calendar
- Course info
- Problem sets
- Mailing list info



Apps

http://april.eecs.umich.edu/courses

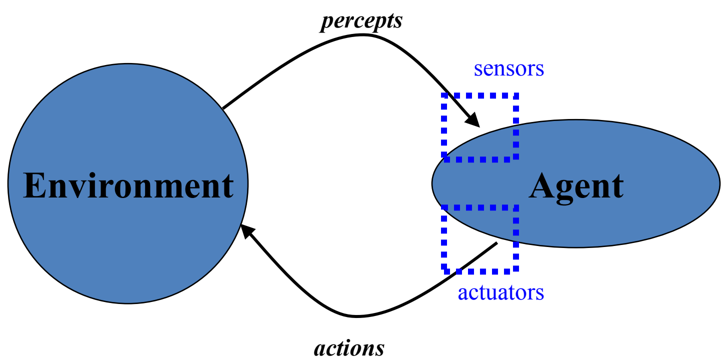
- Your class status is tracked in real time
- Team preferences
  - Who would you like to work with?
- Challenge problem leader board
- Peer Evaluations



Today

- What is an agent?
- Classifying Environments
- Types of Agents

Agents and Environments



Environments

Sensors/Percepts	Actuators/Actions
Classroom temperature	Open/Close Window
Position & velocity of nearby cars	Steering wheel position
Stock price	Buy/Sell stock
…	…

Why call them "agents"?

- Pursues a set of pre-specified goals, or a pre-specified strategy
- Specialization of function
- Entrusted to act *autonomously*

Definition (R&N)

- A system is *autonomous* to the extent that its behavior is determined by its own percepts, rather than the prior knowledge of its designer.

Environment Properties

- R&N characterize environments in six ways:
  - Fully vs Partially Observable
  - Deterministic vs Stochastic, Strategic
  - Episodic vs Sequential
  - Static vs Dynamic
  - Discrete vs Continuous
  - Single-Agent vs Multi-agent

Environment Properties (1/6)

- Fully Observable, Partially Observable
  - *Does the agent perceive the entire state of the world?*
  - Chess?
  - Poker?

Environment Properties (2/6)

- Deterministic, Stochastic, Strategic
  - *Is there a unknowable aspect?*
    - *The world unfolds in a predictable way: deterministic*
    - *Random: stochastic*
    - *Other agents: strategic*
  - Chess?
  - Poker?

Environment Properties (3/6)

- Episodic vs Sequential
  - *Do previous actions/percepts affect future ones?*
  - Chess?
  - Paper-rock-scissors?

Environment Properties (4/6)

- Static, Dynamic, Semi-Dynamic
  - *Does the world change while waiting for the agent act? (or change "on its own"?)*
    - *No: Static*
    - *Yes: Dynamic*
    - *The deliberation time matters: Semi-Dynamic*
  - Chess?
  - Stock-trading agent?

Environment Properties (5/6)

- Discrete vs Continuous
  - *Are there a finite number of actions and percepts?*
  - A thermostat
  - Chess?
  - Taxi cab?

Environment Properties (6/6)

- Single-Agent vs Multi-agent
  - *Are the goals of more than one agent coupled?*
  - Football
  - Stock-trading agent?

Agent Properties

- Agents are described in two fundamental ways:
  - Model type: What is their internal model of the environment?
    - How does the world change with time? (Update)
    - What effect will our actions have on the world? (Predict)
    - Must an agent have a model?
  - Planning type: How do they generate ("plan") actions?

Agent Properties: Model Type

- Stateless
  - World state = current percepts
  - Effects of time, actions, are not modeled
- Fixed Model
  - The designer provides a model
- Learning Model
  - The model can adapt automatically to the observed conditions

Agent Properties: Planning Type

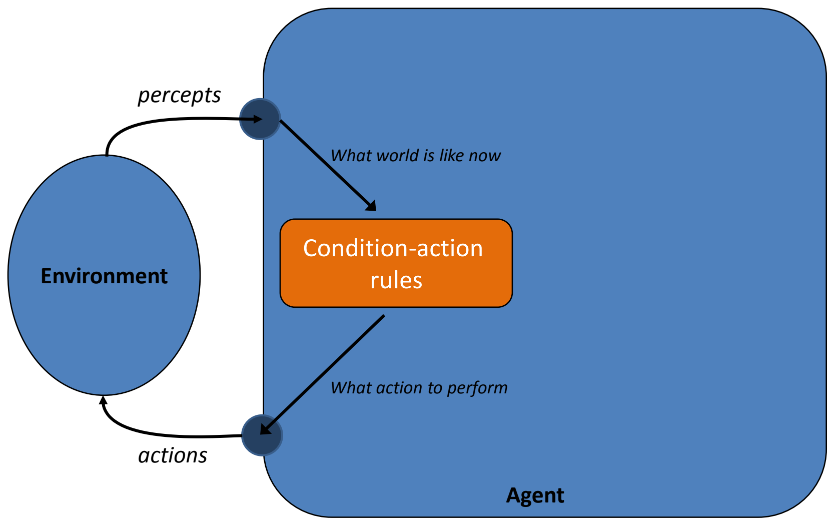
- Reflexive
  - Actions are triggered based on preprogrammed conditions.
- Predictive
  - The effect of potential actions (or action *sequences*) is predicted.
  - The most desirable action (with maximum utility) is performed.

Agent Types

- Mapping to R&N:

Planning \ Model	Stateless	Fixed Model	Learning Model
Reflexive	Table Lookup, Simple Reflexive	Model-Based	
Predictive	X	Goal-Based, Utility-Based	Learning

Simple Reflex Agent



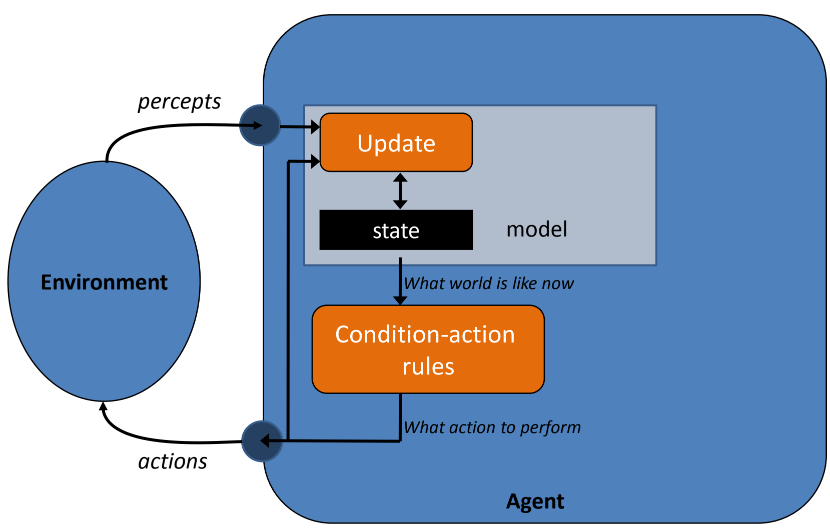
Example Simple Reflex Rules

- Blackjack:
  - If (Sum_of_Cards < 17) then Hit
  - If (Sum_of_Cards > 16) then Stay
- If not(CarStarted) then TurnKey

Simple Reflex Agent: Limitations

- Limitations
  - What if a single set of percepts isn't enough to make a good choice?
- Solution
  - Maintain persistent model of world state
    - Blackjack: Which cards have I seen?
  - Percepts and Actions update this model

Model-Based Reflex Agent



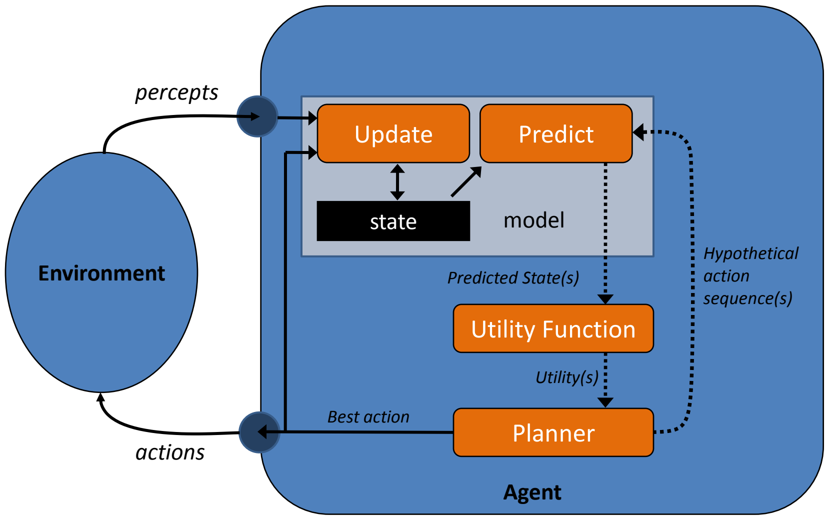
Limitations of Reflexive agents

- Designer may not know the best action for each condition (or may be lazy!)
- Most environments have too many conditions to effectively pre-compute actions for each
- But if the action is "hard-coded", how do we pick between the actions we're capable of?

Utilities

- Score = function (state)
- Roughly corresponds to "happiness"
- In a deterministic world, we pick the action that leads to the best state.
- What do we do in a stochastic world?
  - If I go speed, there's a 80% chance I arrive on time. But there's a 20% chance that I get a ticket!
  - Solution: We ***maximize expected utility***: the probabilistically-weighted average of utilities given our current information.

Predictive, Utility-Based Agent



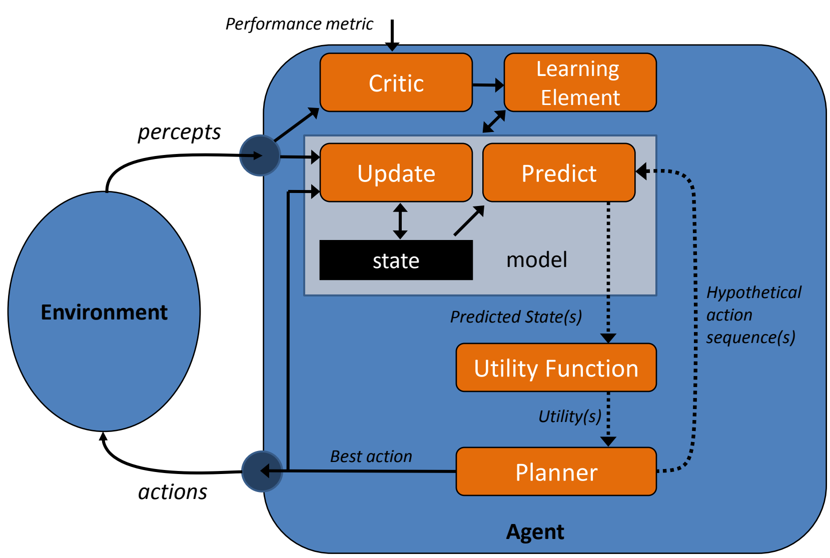
Limitations of Fixed-Model Agents

- Key properties of the environment may change over time, or model may be too simple to fit all cases
  - (Examples?)
- Percepts may change over time, even if environment is constant.
  - (Examples?)
- Effects of actions may change over time.
  - Examples?

Solution: Learning agents

- Modify our model based on the agent's experiences.
- Simple case: online parameter tuning
  - How could a robot detect that its wheels are slipping at a rate different from what the model predicts?
- More complex:
  - Are little red sports cars more likely to run red lights?

Learning Agent



Next Time

- Topics:
  - More on utility functions
  - How do we evaluate performance of an agent?
  - Solving problems with search

- Fill out your team preferences; team assignments will be made on Friday at 11:59p!