CSCI 561
Foundations of
Artificial Intelligence
Lecture 1: Overview and Background

(Chapter 1 & 2)

SUMMER 2017

INSTRUCTOR: PROF. SHEILA TEJADA

What is Artificial Intelligence (AI)?

Some problematic definitions

- “The study of how to make computers do things at which, at the moment, people are better.” *(E. Rich & K. Knight)*
- “The concept of making computers do tasks once considered to require thinking.”
- The use of logic or rules or … to solve problems on computers.
- “An algorithm by which the computer gives the illusion of thinking like a human.” *(D. Gruber)*
- “Making computers behave like humans.” *(Webopedia)*

What is Artificial Intelligence? Four views

	Like People	Rationally
Think	**Thinking Humanly** "The exciting new effort to make computers think . . . *machines with minds*, in the full and literal sense." (Haugeland, 1985) "[The automation of] activities that we associate with human thinking, activities such as decision-making, problem solving, learning . . ." (Bellman, 1978)	**Thinking Rationally** "The study of mental faculties through the use of computational models." (Charniak and McDermott, 1985) "The study of the computations that make it possible to perceive, reason, and act." (Winston, 1992)
Act	**Acting Humanly** "The art of creating machines that perform functions that require intelligence when performed by people." (Kurzweil, 1990) "The study of how to make computers do things at which, at the moment, people are better." (Rich and Knight, 1991)	**Acting Rationally** "Computational Intelligence is the study of the design of intelligent agents." (Poole *et al.*, 1998) "AI . . . is concerned with intelligent behavior in artifacts." (Nilsson, 1998)

A Synthesis Definition

"The scientific understanding of the mechanisms underlying thought and intelligent behavior and their embodiment in machines." ***(AAAI)***

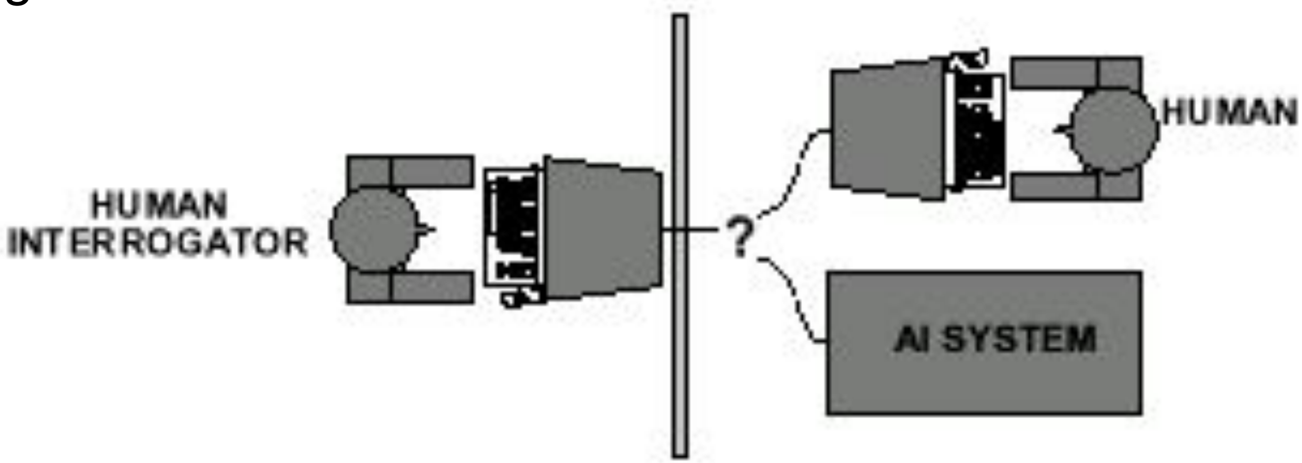
AI overlaps strongly with Cognitive Science and its various subdisciplines, but also relates to:

- **Mathematics**: Formalizations and analysis
- **Economics**: Decision making
- **Operations research**: Optimization and search
- **Engineering**: Robotics

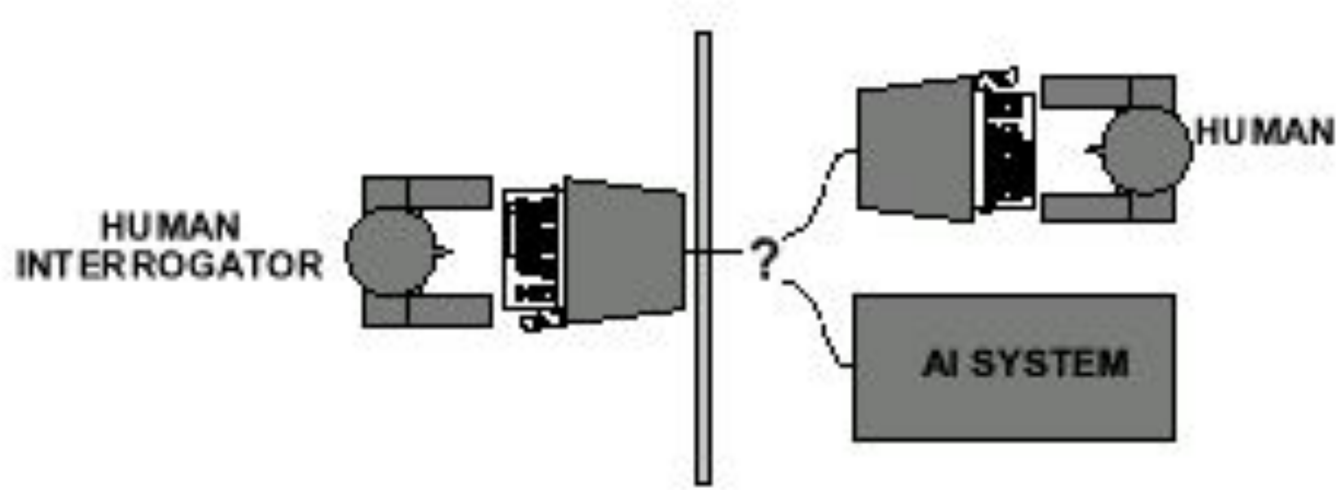
The Turing Test

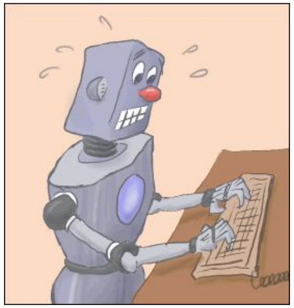
Alan Turing's 1950 article *Computing Machinery and Intelligence* discussed conditions for considering a machine to be intelligent

- "Can machines think?" ← → "Can machines behave intelligently?"
- The Turing test (The Imitation Game): Operational definition of intelligence.



The Turing Test





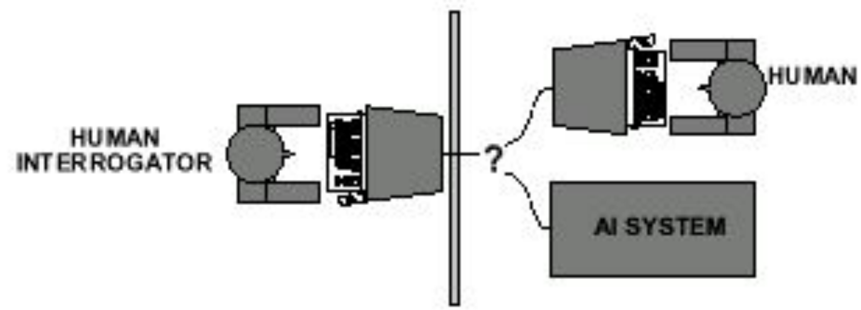
- Computer needs to possess: Natural language processing, Knowledge representation, Automated reasoning, and Machine learning
- Are there any problems/limitations to the Turing Test?

The Turing Test (Problems)

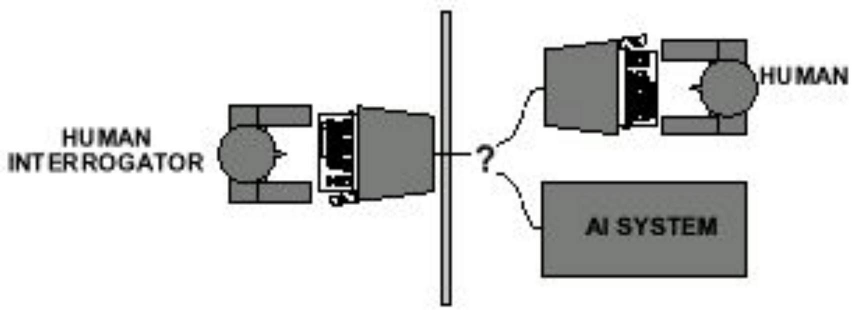
Problems:

1) Turing test is not reproducible, constructive, or amenable to mathematic analysis.

2) Subject to “tricks”



The Turing Test (Problems)



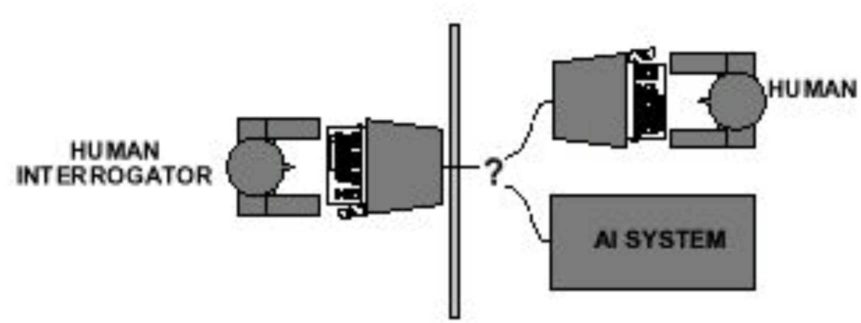
Problems:

1) Turing test is not reproducible, constructive, or amenable to mathematic analysis.

2) Subject to “tricks”

3) What about physical interaction with interrogator and environment?

The Full Turing Test



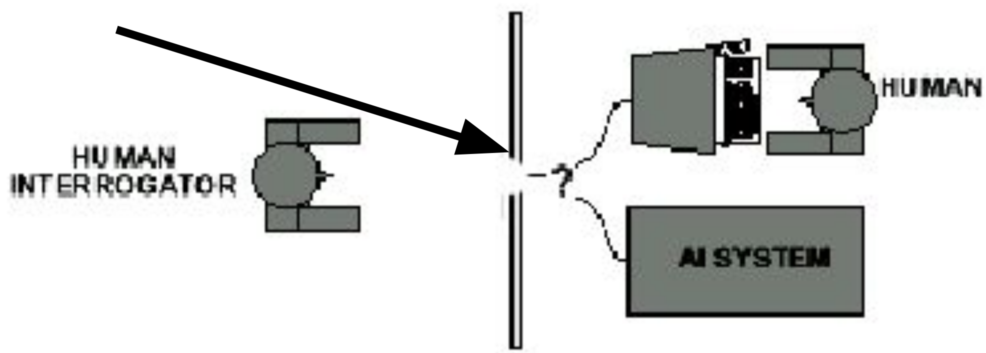
Problems:

1) Turing test is not reproducible, constructive, or amenable to mathematic analysis.

2) Subject to “tricks”

3) What about physical interaction with interrogator and environment?

Trap door



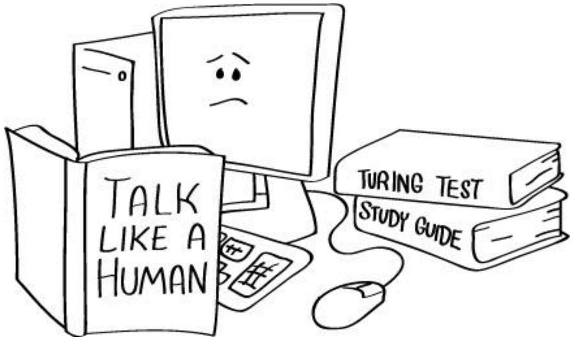
What would a computer need to pass the Turing test?

Natural language processing: to communicate with examiner.

Knowledge representation: to store and retrieve information provided before or during interrogation.

Automated reasoning: to use the stored information to answer questions and to draw new conclusions.

Machine learning: to adapt to new circumstances and to detect and extrapolate patterns.



What would a computer need to pass the Turing test?

Vision (for Total Turing test): to recognize the examiner's actions and various objects presented by the examiner.

Motor control (total test): to act upon objects as requested.

Other senses (total test): such as hearing, smell, touch, etc.

Some Examples

DARPA Urban Challenge (2007)

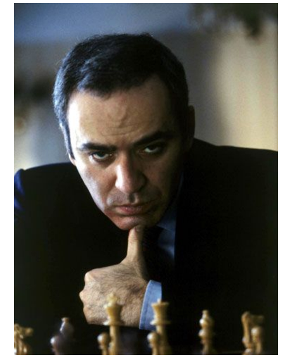
"Autonomous vehicles that safely execute missions in a complex urban environment with moving traffic."

- "The objective of this program is safe and correct autonomous driving capability in traffic at 20 mph."

[DARPA Urban Challenge Summary Video]

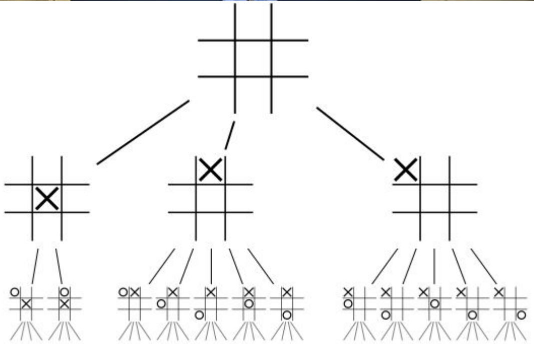
[Google Driverless Car Video]

Deep Blue



In 1997 Deep Blue became the first machine to win a match against a reigning world chess champion (by 3.5-2.5)

Deep Blue Combined



Parallel and special purpose hardware

- A 30-node IBM RS/6000, enhanced with
- 480 special purpose VLSI chess chips

A heuristic game-tree search algorithm

- Capable of searching 200M positions/sec
- Searched 6-12 moves deep on average, sometimes to 40

Chess knowledge

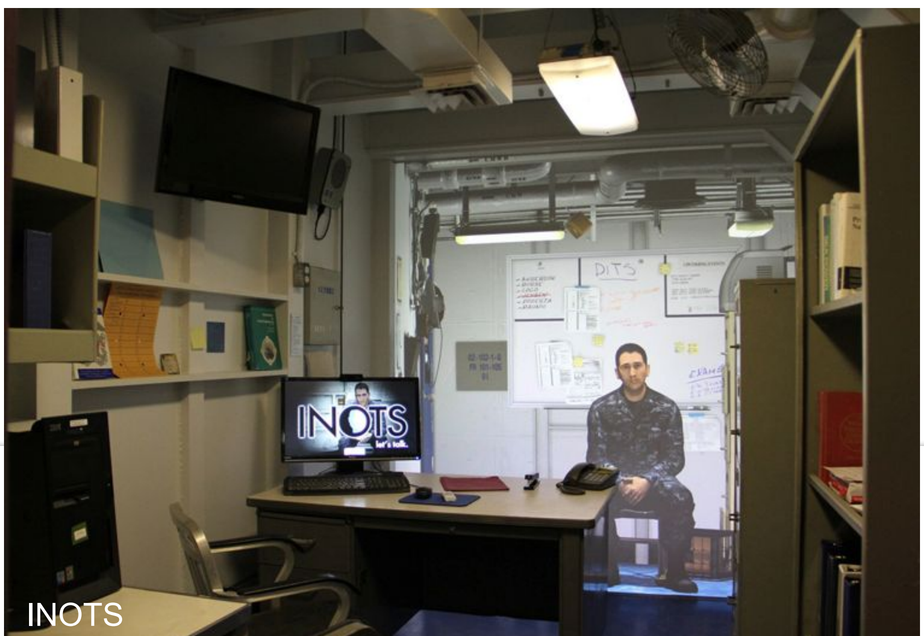
- An opening book of 4K positions and 700K GM games
- An endgame database for when only 5-6 pieces left
- A positional evaluation function with 8K parts and many parameters that were tuned by learning over thousands of Master games

[Watson & Jeopardy]

Example Virtual Humans (USC/ICT)


Ada & Grace


Gunslinger



INOTS


SASO

CSCI 561: Artificial Intelligence

Course overview: foundations of symbolic intelligent systems. Agents, search, problem solving, logic, representation, reasoning, symbolic programming, and robotics.

Prerequisites: Good programming and algorithm analysis skills. Basic probability theory desirable.

Exams: 4:00pm – 5:20pm on Thursdays June 8 and June 29
Tuesday July 25 (all in class)

Textbook:
Russell & Norvig, *Artificial Intelligence: A Modern Approach*

CSCI 561: Artificial Intelligence

Professor

Sheila Tejada, PhD (stejada@usc.edu)

Teaching Assistant

Yue Shi (yueshi@usc.edu)

Eric Heiden (heiden@usc.edu)

Lectures

4:00pm – 6:30pm on Tuesdays and Thursdays in OHE122

Office Hours

After lecture on Tuesdays and Thursdays in OHE122 and by appointment

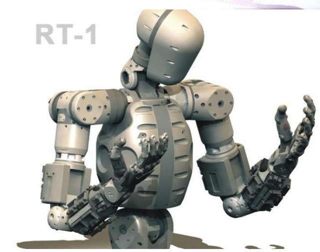
Grand Challenges for AI

Human-level AI

- *Intelligent virtual humans*
- Humanoid robots

Superhuman performance in limited domains

- *Beat the world champion at chess* (or Go)
- Effective control of very complex systems



Specific advanced capabilities of interest

- Mathematical or scientific discovery
- *Automatically drive a vehicle in real world*
- *Answer a wide range of questions*
- Autonomously behave & learn continuously over years
- *Automated real-time language* and speech translation
- Learn to perform a new task from scratch

CSCI 561: Artificial Intelligence

Grading

Grades for this course will be based on performance on participation exercises, homework and exams.

Participation: 10%

Homework: 25%, 3 assignments (5%, 10% 10%)

Midterms: 40%, 2 exams (20%, 20%)

Final: 25%, comprehensive exam

CSCI 561: Artificial Intelligence

Grading

Final letter grades for this course are entirely at the discretion of the course professor. Based on previous semesters, we expect that letter grades for this course will follow this scale:

- 85% or higher : A
- 80-85% : A-
- 75-80% : B+
- 70-75% : B
- 65-70 : B-
- 60-65% : C+
- 55-60% : C
- 50-55% : C-

CSCI 561: Artificial Intelligence

More on Homework Assignments

- There will be three homework assignments, which may consist of programming problems, open-ended essay questions, and questions representative of those that will appear on course exams.
- Programming problems may require the use of a specific programming language, e.g. Python.
- Programming assignments will be corrected and checked for cheating by automated agents.

CSCI 561: Artificial Intelligence

Learning Management System

This course will make extensive use of the online learning management system, DEN@Viterbi. Students will submit all homework assignments via this system. This system will also be used to provide online discussion forums where students can discuss topics with their peers, the teaching assistant, and course instructors. DEN@Viterbi can be found online at: **https://courses.uscden.net/d2l/login**

Academic Integrity

Familiarize yourself with the USC Academic Integrity guidelines.

Violations of the Student Conduct Code will be filed with the Office of Student Judicial Affairs, and appropriate sanctions will be given.

Homework assignments are to be solved individually.

You are welcome to discuss class material in review groups, but do not discuss how to solve the homeworks.

Exams are closed-book with no questions allowed.

Academic Integrity

All students are responsible for reading and following the Student Conduct Code. **Note that the USC Student Conduct Code prohibits plagiarism.**

Some examples of what is **not** allowed by the conduct code: **copying all or part of someone else's work** (by hand or by looking at others' files, either secretly or if shown), and submitting it as your own; **giving another student in the class a copy of your assignment solution;** or **consulting with another student during an exam.**

=> If you have questions about what is allowed, please discuss it with the instructor.

Students who violate university standards of academic integrity are **subject to disciplinary sanctions**, including **failure** in the course and **suspension from the university**. Since dishonesty in any form harms the individual, other students, and the university, policies on academic integrity will be **strictly enforced**. Violations of the Student Conduct Code will be filed with the Office of Student Judicial Affairs.

Review

Intelligence

- Does the right thing given what it knows (*rational*)
- The common underlying capabilities that enable a system to be general, literate, rational, autonomous and collaborative

Artificial Intelligence

- The scientific understanding of the mechanisms underlying thought and intelligent behavior and their embodiment in machines

Intelligent Agents

- Goals, knowledge, perception and action

Lecture 1 (part 2)

Agents and environments

- Introducing the vacuum-cleaner world

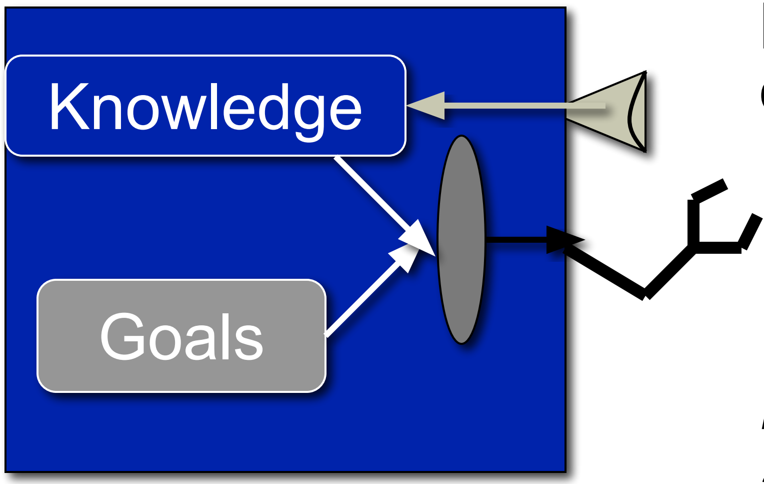
The concept of rational behavior

Environments

Agent types and variations

- Book cares most about *rational agents*

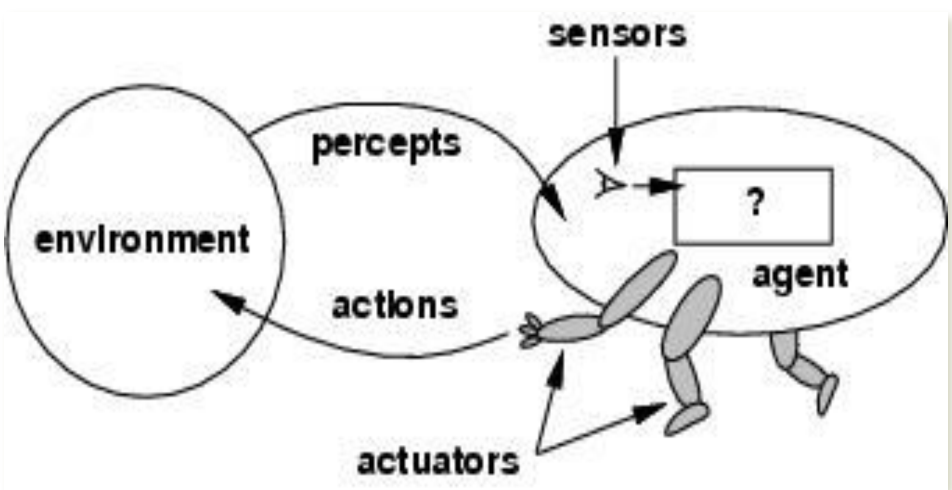
Agents and Environments

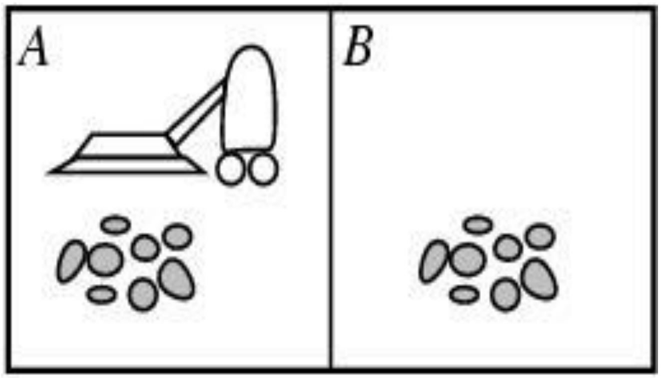


Agents **are systems that perceive and act in some environment**

- Including humans, robots, softbots, thermostats, etc.
- Ignore Knowledge & Goals for now

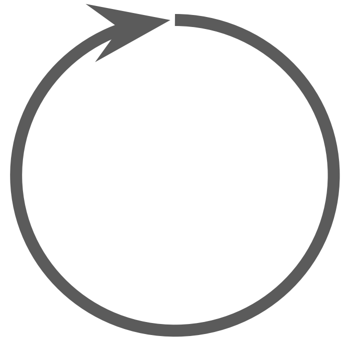
Environment **is world in which agent operates**





Vacuum-Cleaner World

Cognitive Cycle



Agent repeatedly decides what to do next

- The *cognitive cycle* that repeats for agent lifetime

- In humans, the cycle runs at ~50-100ms
  - This is minimum time to choose an action, but many such cycles can be combined to make harder choices

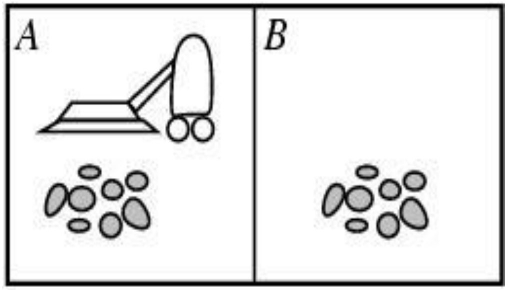
On each cycle, agent can be considered to be computing a function for decision making

Agent Function

The *agent function* is a mathematical relationship that maps percept sequences to actions in the environment

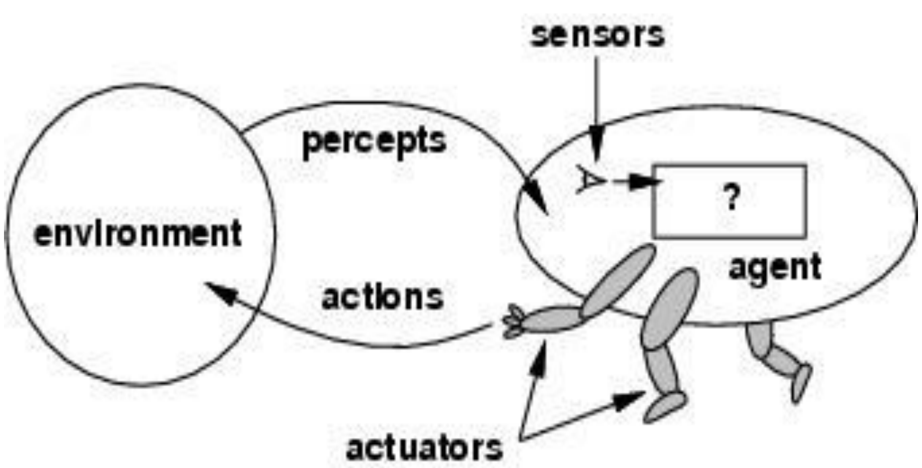
- **f: P* >A**

For example:



- *[A, Dirty] > SUCK*
- *[Car 20' away][Car 10' away] > RUN*

From Agent Functions to Agent Programs

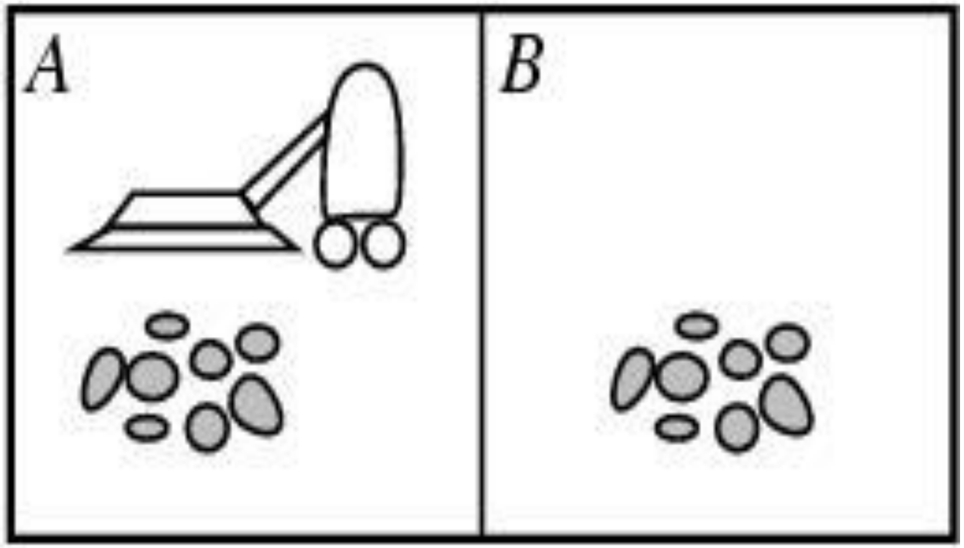


The *agent function* is computed by an *agent program*

The agent program runs on the physical *architecture* to implement the function

- I.e., generates actions from percept sequences

The Vacuum-Cleaner World

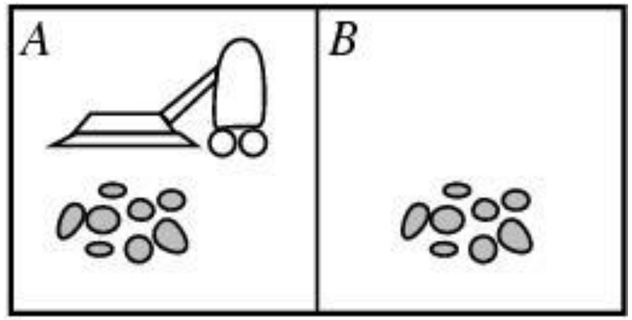


Environment: squares A and B

Percepts: [Location, Content] e.g. *[A, Dirty]*

Actions: left, right, suck, and no-op

A Table-Based Agent Program



Percept sequence	Action
[A,Clean]	Right
[A, Dirty]	Suck
[B, Dirty]	Suck
[A, Clean],[A, Clean]	Right
[A, Clean],[A, Dirty]	Suck
…	…

This Blows Up as Time Goes On

Rationality & Rational Agents

What is rational at a given time depends on:

- Performance measure
  - Ideally objective, external, based on what is to be achieved
- Prior environment knowledge
- Actions
- Percept sequence to date (sensors)

A *rational agent* chooses whichever action maximizes the expected value of the performance measure given the percept sequence to date and the prior environment knowledge

What Rationality Isn't

Generality, literacy, autonomy or collaboration

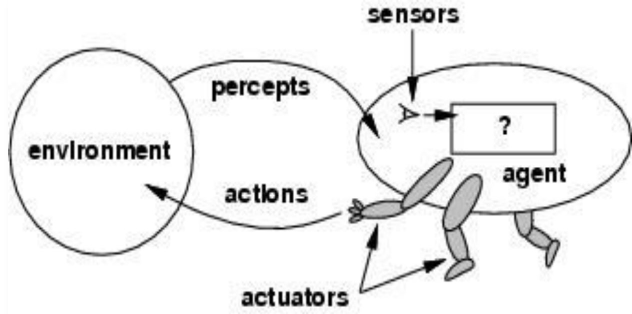
Omniscience

- An omniscient agent knows what will occur

Perfection

- Rationality maximizes *expected (prospective)* performance, while perfection maximizes *actual (retrospective)* performance

Environments



Environment: World in which the agent operates

- To understand agent behavior - or to design a special purpose agent - need to understand its environment
  - Simon's Ant: Complex behavior may arise from a simple program in a complex environment
  - Rationality defined, at least in part, in terms of agent's environment

PEAS description of the *task environment:*

- **P**erformance: Measure for success/progress/quality
- **E**nvironment: The world in which the agent operates
- **A**ctuators: How the agent affects the environment
- **S**ensors: How the agent perceives the environment

Example: Automated Taxi

Performance

- Safety, destination, profits, legality, comfort, …

Environment

- Streets/freeways, other traffic, pedestrians, weather, …

Actuators

- Steering, accelerating, braking, horn, speaker/display,…

Sensors

- Video, sonar, speedometer, engine sensors, keyboard, GPS, laser rangefinder, audio…

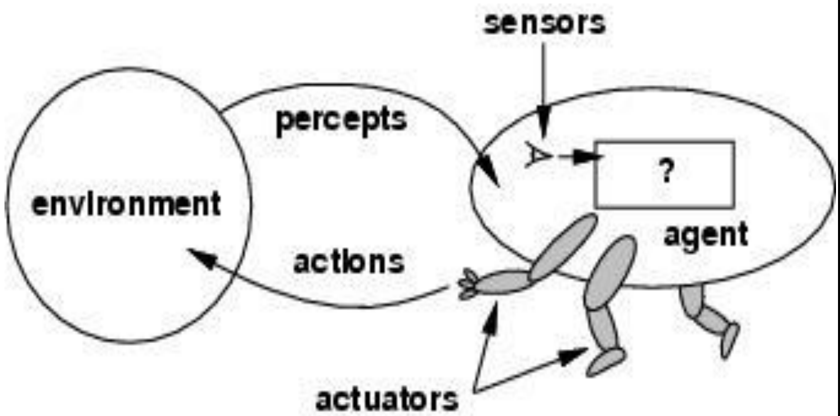
The Agent-Environment Boundary

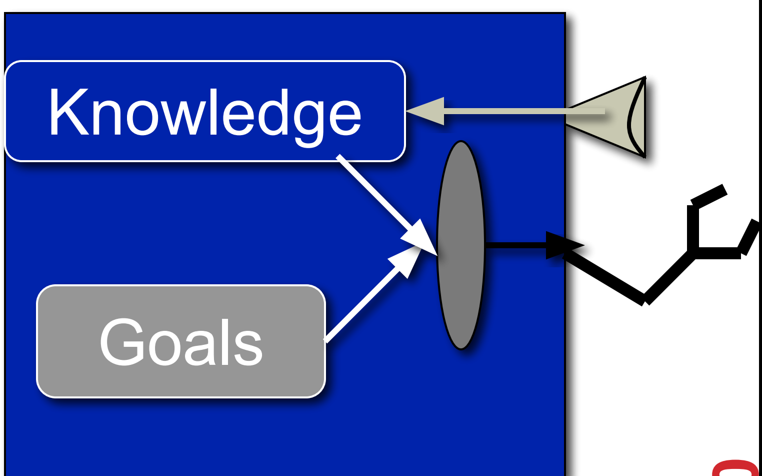
Where draw boundary?

- Does agent include just agent program
- Or does it also include:
  - Goals (performance measure)
  - Body (sensors/actuators)
    - Steering wheel is part of environment, but are arms that grasp them?

A matter of how define analyze and discuss, but not a scientific question

- Will generally use broader sense, with agent itself assumed to have goals and body, but textbook varies on this





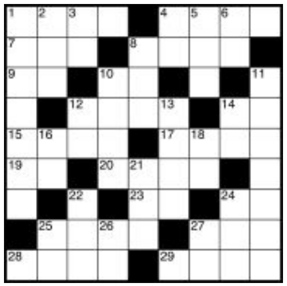
Environment Types

However you define *environment*, its nature can dramatically impact the complexity of the required agent program as well as the difficulty of achieving goals in it

Next few slides look at some key attributes of environments

Part-Picking Robot

Environment Types



	Crossword	Backgammon	Part-Picking Robot	Robot Taxi
Observable??				
Deterministic??				
Episodic??				
Static??				
Discrete??				
Single-agent??				

Environment Types

Fully vs. partially observable: an environment is fully observable when the sensors can detect all aspects that are *relevant* to the choice of action.

	Crossword	Backgammon	Part-Picking Robot	Robot Taxi
Observable??				
Deterministic??				
Episodic??				
Static??				
Discrete??				
Single-agent??				

Environment Types

Fully vs. partially observable: an environment is fully observable when the sensors can detect all aspects that are *relevant* to the choice of action.

	Crossword	Backgammon	Part-Picking Robot	Robot Taxi
Observable??	FULL	FULL	PARTIAL	PARTIAL
Deterministic??				
Episodic??				
Static??				
Discrete??				
Single-agent??				

Environment Types

Deterministic vs. stochastic: if the next environment state is completely determined by the current state and the executed action then the environment is deterministic

	Crossword	Backgammon	Part-Picking Robot	Robot Taxi
Observable??	FULL	FULL	PARTIAL	PARTIAL
Deterministic??				
Episodic??				
Static??				
Discrete??				
Single-agent??				

Environment Types

Deterministic vs. stochastic: if the next environment state is completely determined by the current state and the executed action then the environment is deterministic

	Crossword	Backgammon	Part-Picking Robot	Robot Taxi
Observable??	FULL	FULL	PARTIAL	PARTIAL
Deterministic??	YES	NO	NO	NO
Episodic??				
Static??				
Discrete??				
Single-agent??				

Environment Types

Episodic vs. sequential: In an episodic environment the agent's experience can be divided into atomic steps where the agent perceives and then performs a single action. The choice of action depends only on the episode itself, not on previous actions/episodes & does not affect future actions or choices

	Crossword	Backgammon	Part-Picking Robot	Robot Taxi
Observable??	FULL	FULL	PARTIAL	PARTIAL
Deterministic??	YES	NO	NO	NO
Episodic??				
Static??				
Discrete??				
Single-agent??				

Environment Types

Episodic vs. sequential: In an episodic environment the agent's experience can be divided into atomic steps where the agent perceives and then performs a single action. The choice of action depends only on the episode itself, not on previous actions/episodes & does not affect future actions or choices

	Crossword	Backgammon	Part-Picking Robot	Taxi
Observable??	FULL	FULL	PARTIAL	PARTIAL
Deterministic??	YES	NO	NO	NO
Episodic??	NO	NO	YES	NO
Static??				
Discrete??				
Single-agent??				

Environment Types

Static vs. dynamic: If the environment can change while the agent is choosing an action, the environment is dynamic

	Crossword	Backgammon	Part-Picking Robot	Robot Taxi
Observable??	FULL	FULL	PARTIAL	PARTIAL
Deterministic??	YES	NO	NO	NO
Episodic??	NO	NO	YES	NO
Static??				
Discrete??				
Single-agent??				

Environment Types

Static vs. dynamic: If the environment can change while the agent is choosing an action, the environment is dynamic

	Crossword	Backgammon	Part-Picking Robot	Robot Taxi
Observable??	FULL	FULL	PARTIAL	PARTIAL
Deterministic??	YES	NO	NO	NO
Episodic??	NO	NO	YES	NO
Static??	YES	YES	NO	NO
Discrete??				
Single-agent??				

Environment Types

Discrete vs. continuous: This distinction can be applied to the state of the environment, the way time is handled and to the percepts/actions of the agent

	Crossword	Backgammon	Part-Picking Robot	Robot Taxi
Observable??	FULL	FULL	PARTIAL	PARTIAL
Deterministic??	YES	NO	NO	NO
Episodic??	NO	NO	YES	NO
Static??	YES	YES	NO	NO
Discrete??				
Single-agent??				

Environment Types

Discrete vs. continuous: This distinction can be applied to the state of the environment, the way time is handled and to the percepts/actions of the agent

	Crossword	Backgammon	Part-Picking Robot	Robot Taxi
Observable??	FULL	FULL	PARTIAL	PARTIAL
Deterministic??	YES	NO	NO	NO
Episodic??	NO	NO	YES	NO
Static??	YES	YES	NO	NO
Discrete??	YES	YES	NO	NO
Single-agent??				

Environment Types

Single vs. multi-agent: Does the environment contain more than one agent whose behavior interacts in some relevant way?

	Crossword	Backgammon	Part-Picking Robot	Robot Taxi
Observable??	FULL	FULL	PARTIAL	PARTIAL
Deterministic??	YES	NO	NO	NO
Episodic??	NO	NO	YES	NO
Static??	YES	YES	NO	NO
Discrete??	YES	YES	NO	NO
Single-agent??				

Environment Types

Single vs. multi-agent: Does the environment contain more than one agent whose behavior interacts in some relevant way?

	Crossword	Backgammon	Part-Picking Robot	Robot Taxi
Observable??	FULL	FULL	PARTIAL	PARTIAL
Deterministic??	YES	NO	NO	NO
Episodic??	NO	NO	YES	NO
Static??	YES	YES	NO	NO
Discrete??	YES	YES	NO	NO
Single-agent??	YES	NO	YES	NO

Environment Difficulty

The simplest environment is

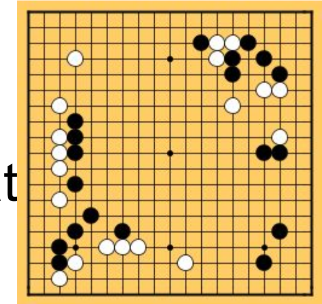
- Fully observable, deterministic, episodic, static, discrete and single-agent

Real world situations are frequently

- Partially observable, stochastic, sequential, dynamic, continuous and multi-agent

Other factors that determine difficulty include

- Difficulty of individual actions
  - E.g., Crosswords, part picking
- Size/combinatorics of environment
  - E.g., The game of Go has ~3^{19*19} (= ~10^{172}) stat



Agent Behavior in Environments

How does the inside of the agent work?

- Agent = architecture + program

All agents types discussed here have the same skeleton:

- Input = current percepts
- Output = action(s)
- Cognitive Cycle = generate output from input (and internal structure)

Agent Types

Four basic kinds of agent programs will be discussed:

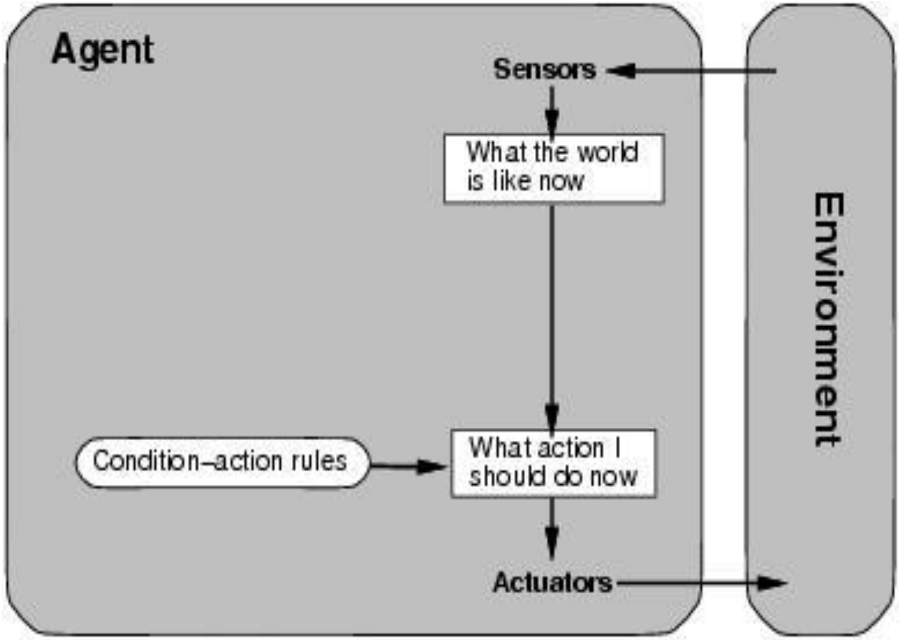
- Simple reflex agents
- Model-based reflex agents
- Goal-based agents
- Utility-based agents

All can be turned into learning agents

A more complex variation

- Hybrid agents

Simple Reflex Agent



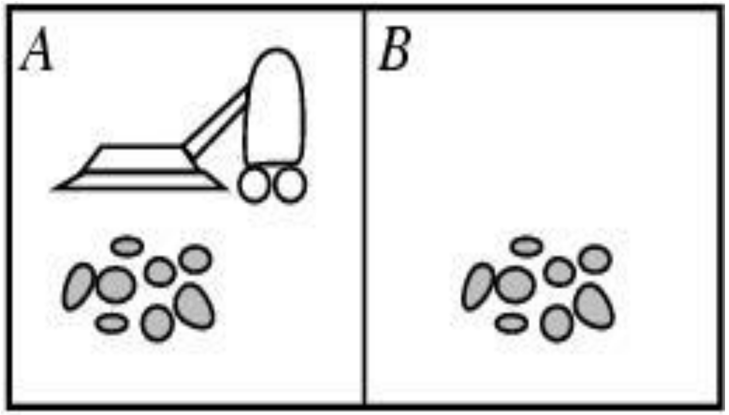
Select action on the basis of *only the current* percept

Large reduction in possible percept/action situations *(next slide)*

May be implemented as *condition-action rules*

- E.g., “If dirty then suck”

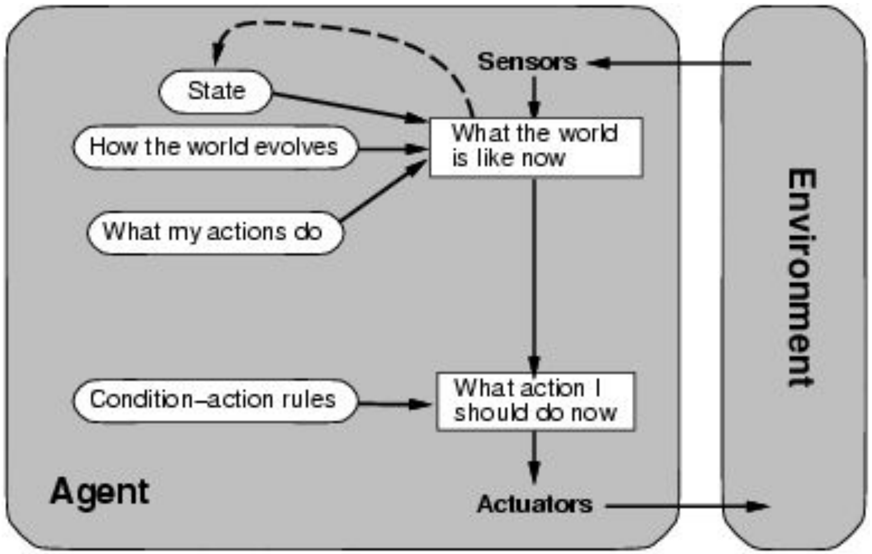
Simple Reflex Program in the Vacuum-Cleaner World



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*

Reduction from 4^T to 3 entries

Model-Based Reflex Agent



To tackle *partially observable* environments

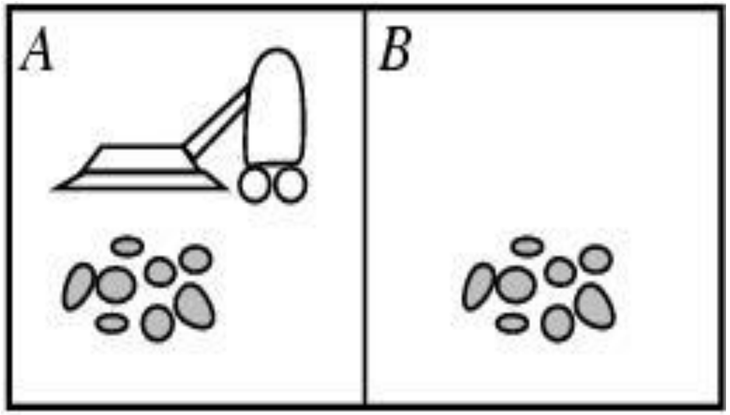
- Maintain internal state representing best estimate of current world situation

Over time update state using world knowledge

- How world changes
- How actions affect world

⇒ ***Model of World***

Model-Based Example

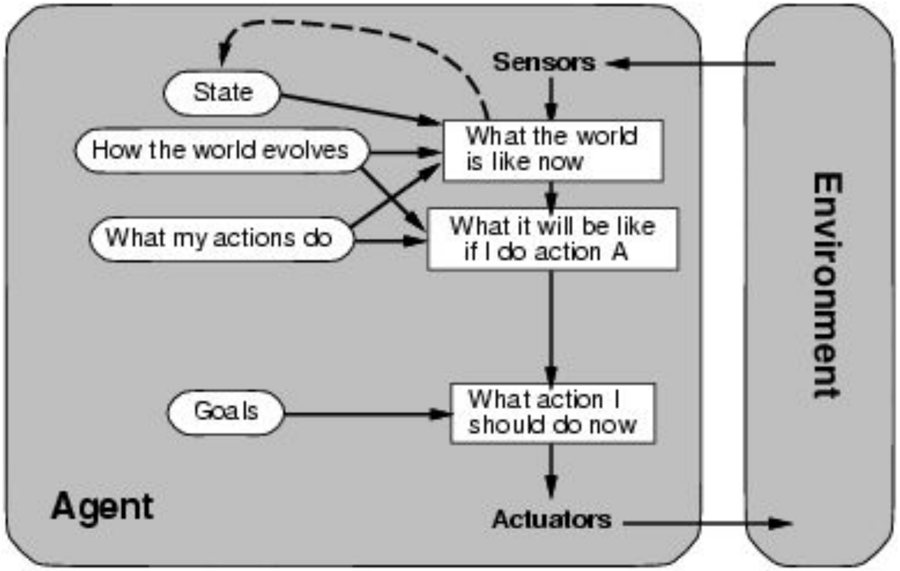


Suppose you are blind, but know that

- *Suck* always cleans the current room
- *Left* always leaves you in room A
- *Right* always leaves you in room B

Totally unobservable world

Goal-Based Agent



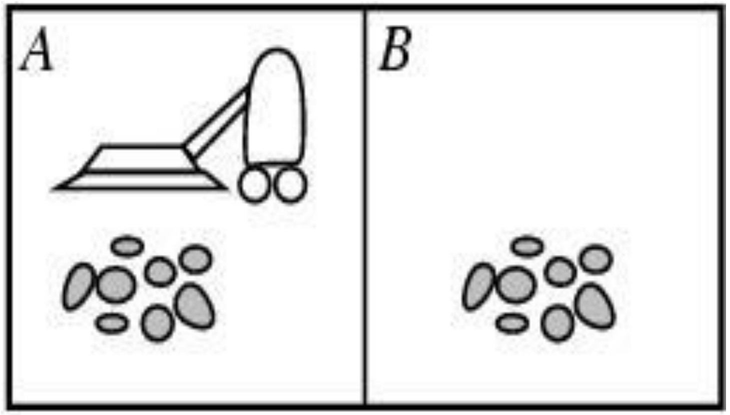
Goals describe what agent wants

- By changing goals, can change what agent does in same situation

Combining models and goals enables determining which possible future paths could lead to goals

- Typically investigated in **search**, **problem solving** and **planning** research

Goal Based Example



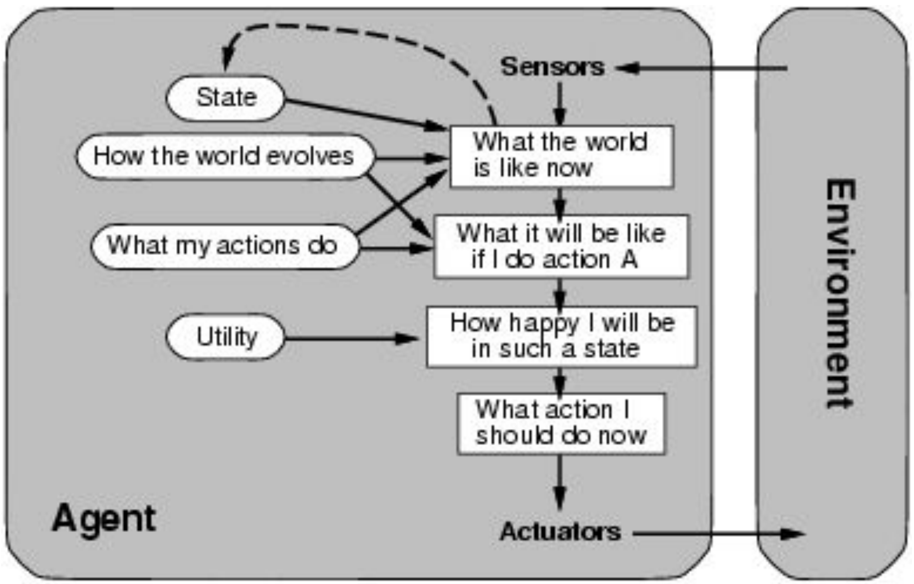
If goal is to clean both rooms

- *Suck*

If goal is just to clean room B

- *Right*

Utility-Based Agent



Some goals can be solved in different ways

- Some solutions may be "better" – have higher utility

Utility function maps a (sequence of) state(s) onto a real number

- Can think of goal achievement as 1 versus 0

Can help in optimization or in arbitration among goals or solutions

Summing Up

Key things to know:

- What is an **agent**? What is an **agent function?**
- What is a **performance measure?**
- What is (and isn't) **rational** behavior?
- What is a **task environment?** What is **PEAS?**
- What are the dimensions and descriptors to characterize a task environment? (fully vs. partially observable, single- vs. multi-agent, deterministic vs. stochastic, etc)
- What kinds of agent programs are there? How are they different?

For next time...

Read Chapter 3 in AIMA “Solving Problems by Searching”

Practice problems:

Written Exercises:

from the textbook AIMA 3rd edition:

Chapter 1 Problem# 1.14

Chapter 2 Problem# 2.2, 2.5