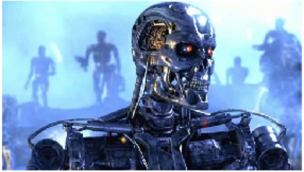
Lecture 1: Introduction to Artificial Intelligence

CS 580 (001) - Spring 2018

Amarda Shehu

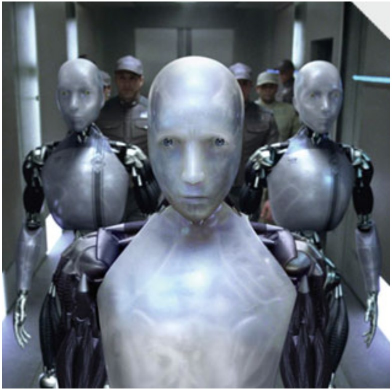
Department of Computer Science
George Mason University, Fairfax, VA, USA

January 24, 2018

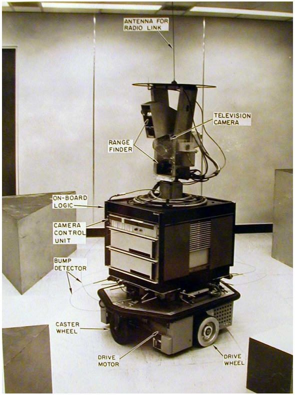



BATTLESTAR
GALACTICA
THE PLAN

Is this AI?



Is this an Intelligent Artifact?



http://en.wikipedia.org/wiki/Shakey_the_robot

What about these?

[auto] [bdog] [rhex] [heli1] [snake] [asimo]

What would you Consider an Intelligent Artifact?

What would you Consider an Intelligent Artifact?

Is Intelligence to be Human or to be Rational?

What would you Consider an Intelligent Artifact?

Is Intelligence to be Human or to be Rational?

Systems that think like humans	**Systems that think rationally**
Systems that act like humans	**Systems that act rationally**

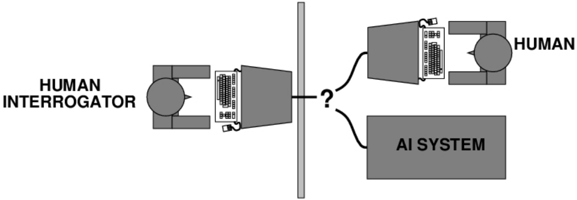
To be intelligent is to act humanly

The Turing Test for an Intelligent Artifact

♢ Turing focused on “acting like a human” as an operational definition of artificial intelligence: Turing (1950) “Computing machinery and intelligence”

♢ Proposed the “imitation” game as a test for a **hidden** intelligent artifact who could be fed visual and material information

♢ Objective: fool a human 30% of the time in a 5-minute test.

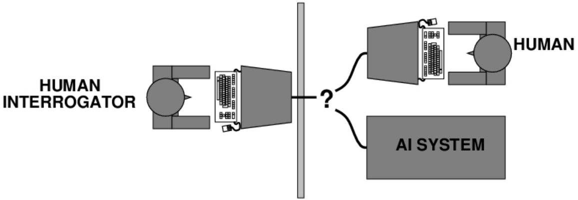


The Turing Test for an Intelligent Artifact

♢ Turing focused on "acting like a human" as an operational definition of artificial intelligence: Turing (1950) "Computing machinery and intelligence"

♢ Proposed the "imitation" game as a test for a **hidden** intelligent artifact who could be fed visual and material information

♢ Objective: fool a human 30% of the time in a 5-minute test.



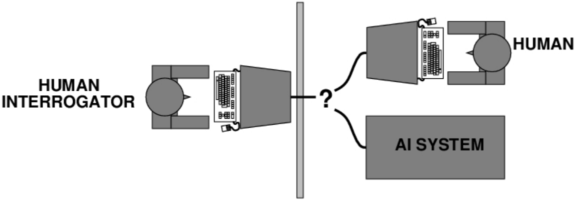
♢ Test introduced major components of AI:
knowledge, reasoning, language, understanding, learning
(computer vision, robotics)

The Turing Test for an Intelligent Artifact

◇ Turing focused on "acting like a human" as an operational definition of artificial intelligence: Turing (1950) "Computing machinery and intelligence"

◇ Proposed the "imitation" game as a test for a **hidden** intelligent artifact who could be fed visual and material information

◇ Objective: fool a human 30% of the time in a 5-minute test.



◇ Test introduced major components of AI:
knowledge, reasoning, language, understanding, learning
(computer vision, robotics)

◇ Predicted objective would be reached by year 2000

◇ Anticipated all major arguments against AI

The Truth about the Turing Test

♢ Problem: Turing test is not **reproducible**, **informative/constructive**, or amenable to **mathematical analysis**

♢ Weak vs. Strong AI argument: one can simulate intelligence but not possess it

♢ AI researchers largely interested in underlying principles

♢ Turing test evolved in pop culture

What would you Consider an Intelligent Artifact?

Fooling a human is not enough – need proof of consciousness.

Fooling a human is not enough – need proof of consciousness.

Figure: Popular culture beyond the Turing Test – Ex Machina (2015)

Fooling a human is not enough – need proof of consciousness.

Figure: Popular culture beyond the Turing Test – Ex Machina (2015)

From fooling to manipulating a human as ultimate test.

To be intelligent is to think humanly

- 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. 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**
- All three fields share one principal direction!

Shouldn't understanding how the brain works precede any effort to build artificial intelligence?

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

Shouldn't understanding how the brain works precede any effort to build artificial intelligence?

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

Shouldn't understanding how the brain works precede any effort to build artificial intelligence?

- Brains (human minds) are very good at making rational decisions, but not perfect
  - Is rational perfect?
- Brains arent as modular as software, so hard to reverse engineer!

What about the Brain?

Shouldn't understanding how the brain works precede any effort to build artificial intelligence?

- Brains (human minds) are very good at making rational decisions, but not perfect
  - Is rational perfect?
- Brains arent as modular as software, so hard to reverse engineer!
- "Brains are to intelligence as wings are to flight"

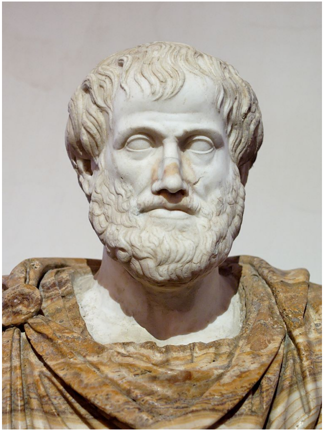
Shouldn't understanding how the brain works precede any effort to build artificial intelligence?

- Brains (human minds) are very good at making rational decisions, but not perfect
  - Is rational perfect?
- Brains arent as modular as software, so hard to reverse engineer!
- "Brains are to intelligence as wings are to flight"
- Lessons learned from the brain: memory and simulation are key to decision making

To be intelligent is to think rationally

Or 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 out of all the thoughts (logical or otherwise) that I **could** have?



Aristotle, 384 322 BC

To be intelligent is to act rationally

and Acting Rationally

Rational behavior: doing the right thing

Rational behavior: doing the right thing

We use the term rational in a **specific, technical** way

Rational behavior: doing the right thing
We use the term rational in a **specific, technical** way

- The right thing: that which is expected
  to maximize goal achievement,
  given the available information

Rational behavior: doing the right thing
We use the term rational in a **specific, technical** way

- 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

Rational behavior: doing the right thing
We use the term rational in a **specific, technical** way

- 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):

and Acting Rationally

Rational behavior: doing the right thing
We use the term rational in a **specific, technical** way

- 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 behavior: doing the right thing
We use the term rational in a **specific, technical** way

- 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: maximally achieving pre-defined goals

Rational behavior: doing the right thing
We use the term rational in a **specific, technical** way

- 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: maximally achieving pre-defined goals
- Rationality only concerns what decisions are made (not the thought process behind them)

Rational behavior: doing the right thing
We use the term rational in a **specific, technical** way

- 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: maximally achieving pre-defined goals
- Rationality only concerns what decisions are made (not the thought process behind them)
- Goals are expressed in terms of the utility of outcomes

Rational behavior: doing the right thing
We use the term rational in a **specific, technical** way

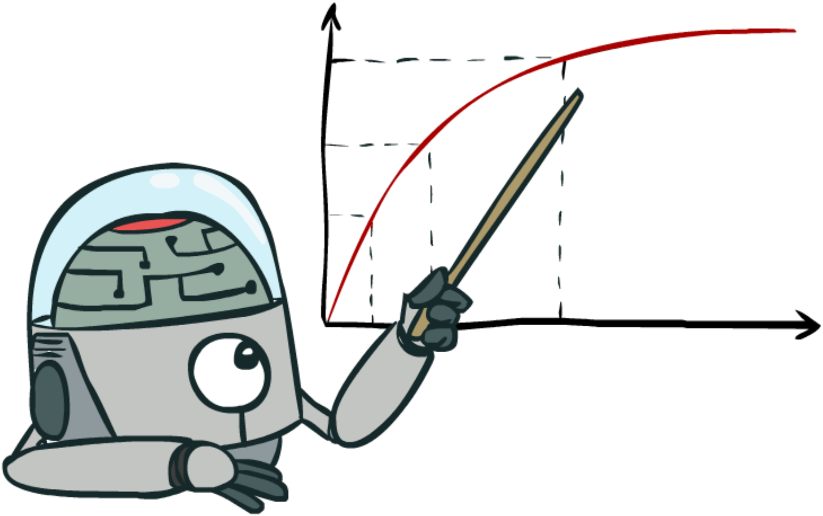
- The right thing: that which is expected to maximize goal achievement, given the available information
- Doesn't necessarily involve thinking—e.g., blinking reflex—but thinking should be in the service of rational action
- Aristotle (Nicomachean Ethics):
  - **Every art and every inquiry, and similarly every action and pursuit, is thought to aim at some good**

- Rational: maximally achieving pre-defined goals
- Rationality 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**

Rational behavior: doing the right thing
We use the term rational in a **specific, technical** way

- 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: maximally achieving pre-defined goals
- Rationality 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**



Rational Agents

An agent is an entity that perceives and acts

Rational Agents

An agent is an entity that perceives and acts (Why need to perceive?)

Rational Agents

An agent is an entity that perceives and acts (Why need to perceive?)

This course is about designing rational agents

Rational Agents

An agent is an entity that perceives and acts (Why need to perceive?)

This course is about designing rational agents

Abstractly, an agent is a function from percept histories to actions:

$$f : \mathcal{P}^* \rightarrow \mathcal{A}$$

An agent is an entity that perceives and acts (Why need to perceive?)

This course is about designing rational agents

Abstractly, an agent is a function from percept histories to actions:

$$f : \mathcal{P}^* \rightarrow \mathcal{A}$$

For any given class of **environments** and **tasks**, we seek the agent (or class of agents) with the best performance

An agent is an entity that perceives and acts (Why need to perceive?)

This course is about designing rational agents

Abstractly, an agent is a function from percept histories to actions:

$$f : \mathcal{P}^* \to \mathcal{A}$$

For any given class of **environments** and **tasks**, we seek the agent (or class of agents) with the best performance

Caveat: **computational limitations may make perfect rationality unachievable**

Rational Agents

An agent is an entity that perceives and acts (Why need to perceive?)

This course is about designing rational agents

Abstractly, an agent is a function from percept histories to actions:

$$f : \mathcal{P}^* \rightarrow \mathcal{A}$$

For any given class of **environments** and **tasks**, we seek the agent (or class of agents) with the best performance

Caveat: **computational limitations may make perfect rationality unachievable**

$\rightarrow$ design best program for given machine resources

An agent is an entity that perceives and acts (Why need to perceive?)

This course is about designing rational agents

Abstractly, an agent is a function from percept histories to actions:

$$f : \mathcal{P}^* \rightarrow \mathcal{A}$$

For any given class of **environments** and **tasks**, we seek the agent (or class of agents) with the best performance

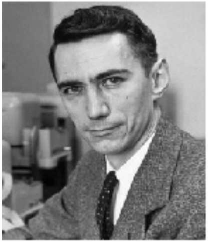
Caveat: **computational limitations may make perfect rationality unachievable**

$\rightarrow$ design best program for given machine resources

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

- Dartmouth Workshop Proposal:
  - "An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves.
  - We think that a significant advance can be made if we work on it together for a summer."



John McCarthy and Claude Shannon

Potted History of AI

1940-1950: The Early Days	
1943	McCulloch & Pitts: Boolean circuit model of brain
1950	Turing's "Computing Machinery and Intelligence"
1950-1970: Excitement Look Ma, No Hands!	
1950s	Early programs, Samuel's checkers program, Newell-Simon's Logic Theorist, Gelernter's Geometry Engine
1956	Dartmouth meeting: coined "AI"
1965	Robinson's complete logical reasoning algorithm
1970-1990: AI Winter	
1966–74	AI discovers computational complexity, neural network research almost disappears
1969–79	Early development of knowledge-based systems
1980–93	Expert systems industry booms and then busts
1985–95	Neural networks popular again
1990s: Statistical Approaches	
1988–	Resurgence of probability; increase in technical depth "Nouvelle AI": ALife, GAs, soft computing
1995–	Agents, agents, everywhere ...
2003–	Human-level AI back on the agenda
Where are we now?	

Quiz: Which of the following can be done at present?

Quiz: Which of the following can be done at present?

- ♢ **Play a decent game of table tennis?**

Quiz: Which of the following can be done at present?

♢ **Play a decent game of table tennis?** ✓

State of the Art - What can AI Currently Do?

Quiz: Which of the following can be done at present?

- ♢ **Play a decent game of table tennis?** ✓
- ♢ **Play a decent game of jeopardy?**

Quiz: Which of the following can be done at present?

- ♢ **Play a decent game of table tennis?** ✓
- ♢ **Play a decent game of jeopardy?** ✓

State of the Art - What can AI Currently Do?

Quiz: Which of the following can be done at present?

- ♢ **Play a decent game of table tennis?** ✓
- ♢ **Play a decent game of jeopardy?** ✓
- ♢ **Drive safely along a curving mountain road?**

Quiz: Which of the following can be done at present?

- ♢ **Play a decent game of table tennis?** ✓
- ♢ **Play a decent game of jeopardy?** ✓
- ♢ **Drive safely along a curving mountain road?** ✓

Quiz: Which of the following can be done at present?

- ♢ **Play a decent game of table tennis?** ✓
- ♢ **Play a decent game of jeopardy?** ✓
- ♢ **Drive safely along a curving mountain road?** ✓
- ♢ **Buy a week's worth of groceries on the web?**

State of the Art - What can AI Currently Do?

Quiz: Which of the following can be done at present?

- ♢ **Play a decent game of table tennis?** ✓
- ♢ **Play a decent game of jeopardy?** ✓
- ♢ **Drive safely along a curving mountain road?** ✓
- ♢ **Buy a week's worth of groceries on the web?** ≈

Quiz: Which of the following can be done at present?

- ♢ **Play a decent game of table tennis?** ✓
- ♢ **Play a decent game of jeopardy?** ✓
- ♢ **Drive safely along a curving mountain road?** ✓
- ♢ **Buy a week's worth of groceries on the web?** ≈
- ♢ **Discover and prove a new mathematical theorem?**

Quiz: Which of the following can be done at present?

- ♢ **Play a decent game of table tennis?** ✓
- ♢ **Play a decent game of jeopardy?** ✓
- ♢ **Drive safely along a curving mountain road?** ✓
- ♢ **Buy a week's worth of groceries on the web?** $\approx$
- ♢ **Discover and prove a new mathematical theorem?** ✓

Quiz: Which of the following can be done at present?

- ♢ **Play a decent game of table tennis?** ✓
- ♢ **Play a decent game of jeopardy?** ✓
- ♢ **Drive safely along a curving mountain road?** ✓
- ♢ **Buy a week's worth of groceries on the web?** $\approx$
- ♢ **Discover and prove a new mathematical theorem?** ✓
- ♢ **Converse successfully with a human for an hour?**

Quiz: Which of the following can be done at present?

- ♢ **Play a decent game of table tennis?** ✓
- ♢ **Play a decent game of jeopardy?** ✓
- ♢ **Drive safely along a curving mountain road?** ✓
- ♢ **Buy a week's worth of groceries on the web?** ≈
- ♢ **Discover and prove a new mathematical theorem?** ✓
- ♢ **Converse successfully with a human for an hour?** X

Quiz: Which of the following can be done at present?

- ♢ **Play a decent game of table tennis?** ✓
- ♢ **Play a decent game of jeopardy?** ✓
- ♢ **Drive safely along a curving mountain road?** ✓
- ♢ **Buy a week's worth of groceries on the web?** ≈
- ♢ **Discover and prove a new mathematical theorem?** ✓
- ♢ **Converse successfully with a human for an hour?** X
- ♢ **Perform surgery?**

State of the Art - What can AI Currently Do?

Quiz: Which of the following can be done at present?

- ♢ **Play a decent game of table tennis?** ✓
- ♢ **Play a decent game of jeopardy?** ✓
- ♢ **Drive safely along a curving mountain road?** ✓
- ♢ **Buy a week's worth of groceries on the web?** ≈
- ♢ **Discover and prove a new mathematical theorem?** ✓
- ♢ **Converse successfully with a human for an hour?** X
- ♢ **Perform surgery?** X

Quiz: Which of the following can be done at present?

- ♢ **Play a decent game of table tennis?** ✓
- ♢ **Play a decent game of jeopardy?** ✓
- ♢ **Drive safely along a curving mountain road?** ✓
- ♢ **Buy a week's worth of groceries on the web?** ≈
- ♢ **Discover and prove a new mathematical theorem?** ✓
- ♢ **Converse successfully with a human for an hour?** X
- ♢ **Perform surgery?** X
- ♢ **Put away the dishes and fold laundry?**

State of the Art - What can AI Currently Do?

Quiz: Which of the following can be done at present?

♢ **Play a decent game of table tennis?**	✓
♢ **Play a decent game of jeopardy?**	✓
♢ **Drive safely along a curving mountain road?**	✓
♢ **Buy a week's worth of groceries on the web?**	$\approx$
♢ **Discover and prove a new mathematical theorem?**	✓
♢ **Converse successfully with a human for an hour?**	X
♢ **Perform surgery?**	X
♢ **Put away the dishes and fold laundry?**	Getting there...

State of the Art - What can AI Currently Do?

Quiz: Which of the following can be done at present?

♢ **Play a decent game of table tennis?**	✓
♢ **Play a decent game of jeopardy?**	✓
♢ **Drive safely along a curving mountain road?**	✓
♢ **Buy a week's worth of groceries on the web?**	$\approx$
♢ **Discover and prove a new mathematical theorem?**	✓
♢ **Converse successfully with a human for an hour?**	X
♢ **Perform surgery?**	X
♢ **Put away the dishes and fold laundry?**	Getting there...
♢ **Translate spoken Chinese into English?**	

State of the Art - What can AI Currently Do?

Quiz: Which of the following can be done at present?

♢ **Play a decent game of table tennis?**	✓
♢ **Play a decent game of jeopardy?**	✓
♢ **Drive safely along a curving mountain road?**	✓
♢ **Buy a week's worth of groceries on the web?**	$\approx$
♢ **Discover and prove a new mathematical theorem?**	✓
♢ **Converse successfully with a human for an hour?**	X
♢ **Perform surgery?**	X
♢ **Put away the dishes and fold laundry?**	Getting there...
♢ **Translate spoken Chinese into English?**	$\approx$

State of the Art - What can AI Currently Do?

Quiz: Which of the following can be done at present?

♢ **Play a decent game of table tennis?**	✓
♢ **Play a decent game of jeopardy?**	✓
♢ **Drive safely along a curving mountain road?**	✓
♢ **Buy a week's worth of groceries on the web?**	≈
♢ **Discover and prove a new mathematical theorem?**	✓
♢ **Converse successfully with a human for an hour?**	X
♢ **Perform surgery?**	X
♢ **Put away the dishes and fold laundry?**	Getting there...
♢ **Translate spoken Chinese into English?**	≈
♢ **Write an intentionally funny story?**	

State of the Art - What can AI Currently Do?

Quiz: Which of the following can be done at present?

♢ **Play a decent game of table tennis?**	✓
♢ **Play a decent game of jeopardy?**	✓
♢ **Drive safely along a curving mountain road?**	✓
♢ **Buy a week's worth of groceries on the web?**	$\approx$
♢ **Discover and prove a new mathematical theorem?**	✓
♢ **Converse successfully with a human for an hour?**	X
♢ **Perform surgery?**	X
♢ **Put away the dishes and fold laundry?**	Getting there...
♢ **Translate spoken Chinese into English?**	$\approx$
♢ **Write an intentionally funny story?**	X

One day Joe Bear was hungry. He asked his friend Irving Bird where some honey was. Irving told him there was a beehive in the oak tree. Joe threatened to hit Irving if he didn't tell him where some honey was. The End.

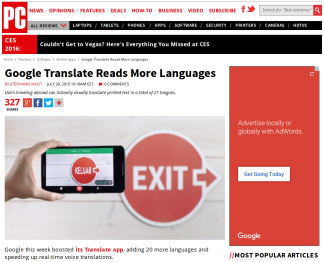
Henry Squirrel was thirsty. He walked over to the river bank where his good friend Bill Bird was sitting. Henry slipped and fell in the river. Gravity drowned. The End.

Once upon a time there was a dishonest fox and a vain crow. One day the crow was sitting in his tree, holding a piece of cheese in his mouth. He noticed that he was holding the piece of cheese. He became hungry, and swallowed the cheese. The fox walked over to the crow. The End.

Joe Bear was hungry. He asked Irving Bird where some honey was. Irving refused to tell him, so Joe offered to bring him a worm if he'd tell him where some honey was. Irving agreed. But Joe didn't know where any worms were, so he asked Irving, who refused to say. So Joe offered to bring him a worm if he'd tell him where a worm was. Irving agreed. But Joe didn't know where any worms were, so he asked Irving, who refused to say. So Joe offered to bring him a worm if he'd tell him where a worm was ...

Natural Language

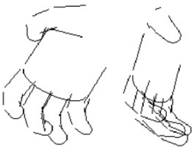
- Speech technologies (e.g. Siri)
  - Automatic speech recognition (ASR)
  - Text-to-speech synthesis (TTS)
  - Dialog systems
- Language processing technologies
  - Question answering
  - Machine translation
- Web search
- Text classification, spam filtering, etc

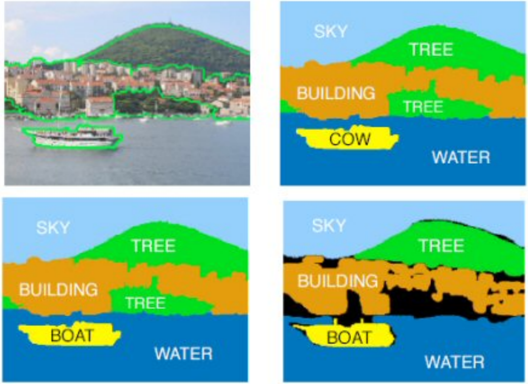


- Robotics
  - At interface between mechanical engineering and AI
  - Algorithms in the Field (AiF) much harder than simulations
- Current technologies
  - Autonomous vehicles
  - Search and Rescue
  - Exploration/Navigation
  - Data Collection
  - Entertainment (Soccer)
  - Service for elderly
  - Autonomous vs. semi-autonomous
- Mechanical aspects ignored here (treated in Robotics courses, such as CS485, CS689)
- Here we will focus on algorithms for planning and control

Vision

- Now central to Robotics
- Object and face recognition
- Image classification
- Object tracking and behavior recognition





- Logical Systems
  - Theorem provers
  - NASA fault diagnostics
  - System Verification (software)
  - Question and answering
- Methods
  - Deduction systems
  - Constraint satisfaction
  - Satisfiability solvers (significant advances)

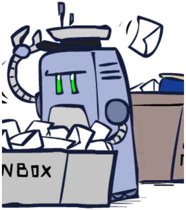

THE PROOF

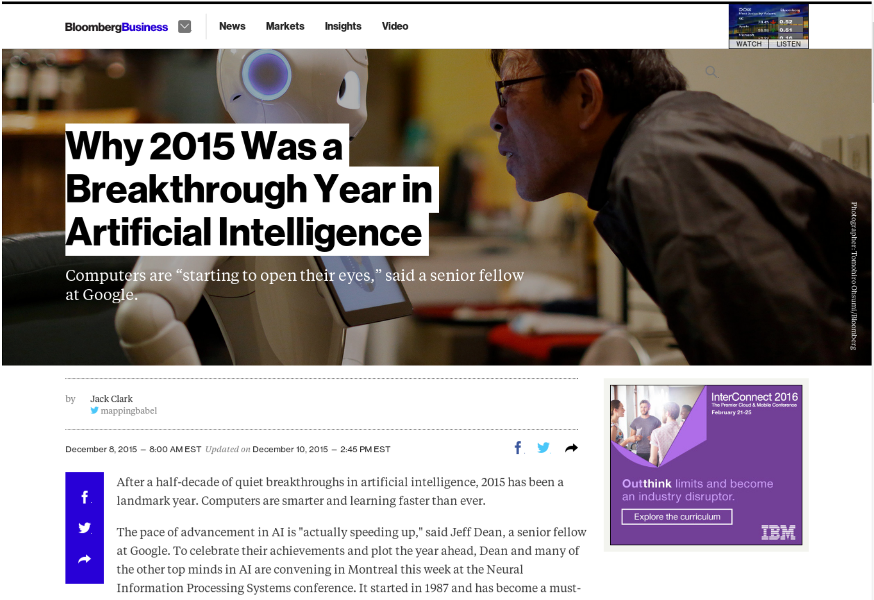
As Eas

Game Playing

- Classic Moment: May, '97: Deep Blue vs. Kasparov
  - First match won against world champion
  - "Intelligent creative" play
  - 200 million board positions per second
  - Humans understood 99.9 of Deep Blue's moves
  - Can do about the same now with a PC cluster
- Open question:
  - How does human cognition deal with the search space explosion of chess?
  - Or: how can humans compete with computers at all??
- 1996: Kasparov Beats Deep Blue
  - "I could feel — I could smell — a new kind of intelligence across the table."
- 1997: Deep Blue Beats Kasparov
  - "Deep Blue hasn't proven anything."
- Computers now match or beat humans in:
  - Chess
  - Othello
  - Scrabble
  - Backgammon
  - Poker
  - Jeorpardy
  - Even Go (Google's AlphaGo - 'Like a God' - beats Ke Jie, world's best player in 2017)

- AI in many other useful systems
  - Scheduling, e.g. airline routing, military
  - Route planning, e.g. Google maps
  - Medical diagnosis, e.g., EKGs
  - Automated surveillance
  - Web search engines
  - Spam classifiers
  - Automated help desks
  - Fraud detection
  - Product recommendations





BloombergBusiness News Markets Insights Video

WATCH LISTEN

Why 2015 Was a Breakthrough Year in Artificial Intelligence

Computers are "starting to open their eyes," said a senior fellow at Google.

Photographer: Tomohiro Ohsumi/Bloomberg

by Jack Clark
mappingbabel

December 8, 2015 – 8:00 AM EST *Updated on* December 10, 2015 – 2:45 PM EST

After a half-decade of quiet breakthroughs in artificial intelligence, 2015 has been a landmark year. Computers are smarter and learning faster than ever.

The pace of advancement in AI is "actually speeding up," said Jeff Dean, a senior fellow at Google. To celebrate their achievements and plot the year ahead, Dean and many of the other top minds in AI are convening in Montreal this week at the Neural Information Processing Systems conference. It started in 1987 and has become a must-

2015: Computers Open their Eyes

♢ Google researchers nabbed the cover of scientific journal Nature with a system that can learn to play and master old Atari games without directions.

♢ Facebook built a way to let computers describe images to blind people.

♢ Microsoft showed off a new Skype system that can automatically translate from one language to another.

♢ IBM singled out AI as one of its greatest potential growth areas.

♢ Preferred Networks is making AI systems that will go into industrial robots made by Japan's Fanuc.

♢ Indico Data Labs worked with a Facebook researcher to teach a computer how to paint faces using its own sort of imagination.

♢ The rise of deep learning (neural networks re-awakened)

♢ New paradigm: teaching computers to think for themselves and improvise solutions to common problems.

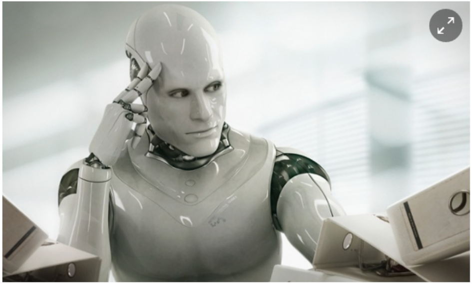
Artificial intelligence (AI)

Elon Musk: artificial intelligence is our biggest existential threat

The AI investor says that humanity risks 'summoning a demon' and calls for more regulatory oversight

Samuel Gibbs
@SamuelGibbs

Monday 27 October 2014 06.26 EDT



Artificial intelligence should be regulated, says Elon Musk. Photograph: Blutgruppe/Blutgruppe/Corbis

Elon Musk has spoken out against artificial intelligence (AI), declaring it the most serious threat to the survival of the human race.

AI in the News

Elon Musk Donates $10M To Make Sure AI Doesn't Go The Way Of Skynet

Posted Jan 15, 2015 by Darrell Etherington (@drizzled)

8,386 SHARES

Next Story

Tesla and SpaceX chief executive Elon Musk has gone on record before proclaiming the potential risks of artificial intelligence, and now he's putting his money where his mouth is. The intrepid inventor and entrepreneur announced a donation of $10 million to help fund research to "keep AI beneficial" to humanity today. The funds go to the Future of Life Institute (FLI), an organization run by volunteers dedicated to research aimed at "mitigate[ing] existential risks facing

SCIENCE

Study to Examine Effects of Artificial Intelligence

By **JOHN MARKOFF** DEC. 15, 2014

Scientists have begun what they say will be a century-long study of the effects of artificial intelligence on society, including on the economy, war and crime, officials at Stanford University announced Monday.

The project, hosted by the university, is unusual not just because of its duration but because it seeks to track the effects of these technologies as they reshape the roles played by human beings in a broad range of endeavors.

"My take is that A.I. is taking over," said Sebastian Thrun, a well-known roboticist who led the development of Google's self-driving car. "A few humans might still be 'in charge,' but less and less so."

Artificial intelligence describes computer systems that perform tasks traditionally requiring human intelligence and perception. In 2009, the president of the Association for the Advancement of Artificial Intelligence, Eric Horvitz, organized a meeting of computer scientists in California to discuss the possible ramifications of A.I. advances. The group concluded that the advances were largely positive and lauded the "relatively graceful" progress.

The Future of AI

- We are doing AI to:
  - create intelligent systems
    - The more intelligent, the better
  - gain a better understanding of human intelligence
  - magnify those benefits that flow from it
- Progress is accelerating, partly due to an industry arms race
- Once performance reaches a minimum level, every 1% improvement is worth billions
  - Speech
  - Text understanding
  - Object recognition
  - Automated vehicles
  - Domestic robots

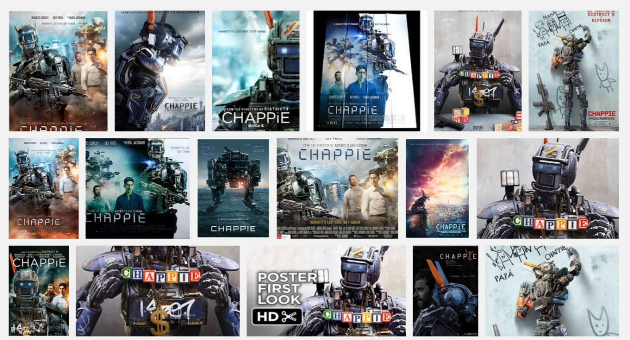
"The first ultra intelligent machine is the last invention that man need ever make. I. J. Good, 1965

- Might help us avoid war and ecological catastrophes, **achieve immortality**, and expand throughout the universe

Pop Culture Fascination with Immortality through AI

Transcendence 2015 (transfer of consciousness on machine)

Pop Culture Fascination with Immortality through AI



Chappie 2015 (a learning robot)

The Future of AI

"The first ultra intelligent machine is the last invention that man need ever make. I. J. Good, 1965

- Might help us avoid war and ecological catastrophes, achieve immortality, and expand throughout the universe
- Success would be the biggest event in human history
- and perhaps the last

Should we be Worried?

- "AI will never reach human levels of intelligence"

Should we be Worried?

- "AI will never reach human levels of intelligence"
- "OK, maybe it will, but I'll be dead before it does"

Should we be Worried?

- “AI will never reach human levels of intelligence”
- “OK, maybe it will, but I’ll be dead before it does”
- ”Machines will never be conscious”

Should we be Worried?

- “AI will never reach human levels of intelligence”
- “OK, maybe it will, but I’ll be dead before it does”
- "Machines will never be conscious”
- Consciousness isnt the problem, it’s competence!

Should we be Worried?

- “AI will never reach human levels of intelligence”
- “OK, maybe it will, but I’ll be dead before it does”
- "Machines will never be conscious”
- Consciousness isnt the problem, it’s competence!
- "We design these things, right?”

Should we be Worried?

- “AI will never reach human levels of intelligence”
- “OK, maybe it will, but I’ll be dead before it does”
- ”Machines will never be conscious”
- Consciousness isnt the problem, it’s competence!
- ”We design these things, right?”
- Yes, and the genie grants three wishes

Should we be Worried?

- “AI will never reach human levels of intelligence”
- “OK, maybe it will, but I’ll be dead before it does”
- "Machines will never be conscious”
- Consciousness isnt the problem, it’s competence!
- "We design these things, right?”
- Yes, and the genie grants three wishes
- For almost any goal, a super intelligent system will acquire as many resources as possible and improve its own algorithms

Should we be Worried?

- “AI will never reach human levels of intelligence”
- “OK, maybe it will, but I’ll be dead before it does”
- "Machines will never be conscious”
- Consciousness isnt the problem, it’s competence!
- "We design these things, right?”
- Yes, and the genie grants three wishes
- For almost any goal, a super intelligent system will acquire as many resources as possible and improve its own algorithms

Should we be Worried?

- “AI will never reach human levels of intelligence”
- “OK, maybe it will, but I’ll be dead before it does”
- ”Machines will never be conscious”
- Consciousness isnt the problem, it’s competence!
- ”We design these things, right?”
- Yes, and the genie grants three wishes
- For almost any goal, a super intelligent system will acquire as many resources as possible and improve its own algorithms
- Protect itself against any attempt to switch it off or change the goal

Precedent: Nuclear Physics

- Rutherford (1933): anyone who looks for a source of power in the transformation of the atom is talking moonshine.
- Sept 12, 1933: The stoplight changed to green.
- Szilard stepped off the curb. As he crossed the street time cracked open before him and he saw a way to the future, death into the world and all our woes, the shape of things to come.
- Szilard (1934): patent on nuclear chain reaction; kept secret

What is the Future of AI?

- Along what paths will AI evolve?
- What is the (plausibly reachable) best case? Worst case?
- Can we affect the future of AI?
    - Can we reap the benefits of super intelligent machines and avoid the risks?
    - "The essential task of our age." Nick Bostrom, Professor of Philosophy, Oxford University.

on that cheery note...

on that cheery note...

What does this Course Cover?

on that cheery note...

What does this Course Cover?

This Course

- Introduction to AI: History of AI. What is artificial intelligence? What are intelligent agents?

This Course

- Introduction to AI: History of AI. What is artificial intelligence? What are intelligent agents?
- Problem-Solving and Search: Uninformed vs. Informed Search, Adversarial Search (Game Playing), Constraint Satisfaction.

This Course

- Introduction to AI: History of AI. What is artificial intelligence? What are intelligent agents?
- Problem-Solving and Search: Uninformed vs. Informed Search, Adversarial Search (Game Playing), Constraint Satisfaction.
- Knowledge and Reasoning: Logical Agents, Propositional, First-order Logic and Inference.

This Course

- Introduction to AI: History of AI. What is artificial intelligence? What are intelligent agents?
- Problem-Solving and Search: Uninformed vs. Informed Search, Adversarial Search (Game Playing), Constraint Satisfaction.
- Knowledge and Reasoning: Logical Agents, Propositional, First-order Logic and Inference.
- Planning: Classical Planning, Knowledge Representation, Robot Motion Planning.

This Course

- Introduction to AI: History of AI. What is artificial intelligence? What are intelligent agents?
- Problem-Solving and Search: Uninformed vs. Informed Search, Adversarial Search (Game Playing), Constraint Satisfaction.
- Knowledge and Reasoning: Logical Agents, Propositional, First-order Logic and Inference.
- Planning: Classical Planning, Knowledge Representation, Robot Motion Planning.
- Uncertainty and Probabilistic Reasoning: Bayesian Networks, Inference in Bayesian Networks, Temporal Probability Models.

This Course

- Introduction to AI: History of AI. What is artificial intelligence? What are intelligent agents?
- Problem-Solving and Search: Uninformed vs. Informed Search, Adversarial Search (Game Playing), Constraint Satisfaction.
- Knowledge and Reasoning: Logical Agents, Propositional, First-order Logic and Inference.
- Planning: Classical Planning, Knowledge Representation, Robot Motion Planning.
- Uncertainty and Probabilistic Reasoning: Bayesian Networks, Inference in Bayesian Networks, Temporal Probability Models.
- Learning: Bayesian Methods, Neural Networks, Perception/Vision, Probabilistic Robotics, NLP.

This Course

- Introduction to AI: History of AI. What is artificial intelligence? What are intelligent agents?
- Problem-Solving and Search: Uninformed vs. Informed Search, Adversarial Search (Game Playing), Constraint Satisfaction.
- Knowledge and Reasoning: Logical Agents, Propositional, First-order Logic and Inference.
- Planning: Classical Planning, Knowledge Representation, Robot Motion Planning.
- Uncertainty and Probabilistic Reasoning: Bayesian Networks, Inference in Bayesian Networks, Temporal Probability Models.
- Learning: Bayesian Methods, Neural Networks, Perception/Vision, Probabilistic Robotics, NLP.

Emphasis on concepts and algorithms. Illustrated with practical applications arising in diverse areas, including mobile systems.

Syllabus

Intro to AI (1 WK)

Problem-Solving and Search (3-4 WKS)

- Uninformed vs. Informed Search
- Constraint Satisfaction, Adversarial Search

Knowledge and Reasoning (4-5 WKS)

- Logical Agents, Propositional Logic
- First-order Logic and Inference in First-order Logic
- Planning, Knowledge Representation, Robot Motion Planning

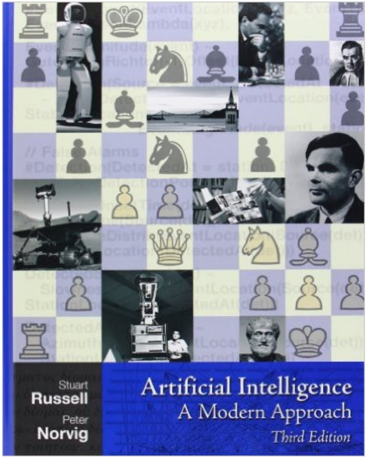
Uncertainty and Probabilistic Reasoning (3 WKS)

- Bayesian Networks
- Inference in Bayesian Networks
- Temporal Probability Models

Learning (2 WKS)

- Bayesian Methods, Neural Networks
- Perception, Robotics
- NLP

Required:



Many online supplemental materials posted on the class syllabus.

Grading and Contact Information

Grading

3 Homeworks (30%)
Midterm Exam (15%)
Final Exam (30%)
Project (25%)

Instructor: Amarda Shehu

Office: ENG #4452
Email: amarda AT gmu.edu
Web: cs.gmu.edu/∼ashehu

CS580 Hours

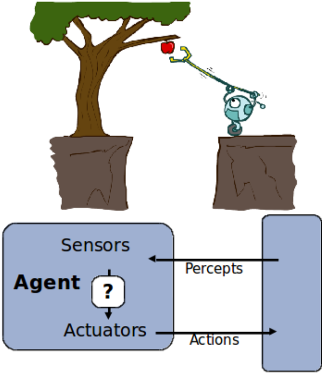
Class: W 4:30 - 7:10 pm
Place: Innovation Hall # 136
Instructor's Office Hours: M 2:30 - 4:30 pm

TA: Rajesh Patel

Email: rpatel17@gmu.edu ENG#5321, W 2:30 - 4:30 pm

Designing Rational Agents

- An agent is an entity that **perceives** and **acts**
- A **rational** agent **selects actions** that **maximize its (expected) utility**
- Characteristics of the percepts, environment, and action space dictate techniques for selecting rational actions



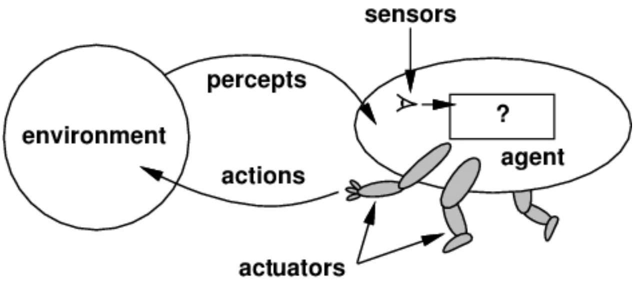
This course is about:

- 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

Intelligent Agents

Intelligent Agents - Outline

- ♢ Agents and Environments
- ♢ Rationality
- ♢ PEAS (Performance measure, Environment, Actuators, Sensors)
- ♢ Environment Types
- ♢ Agent Types

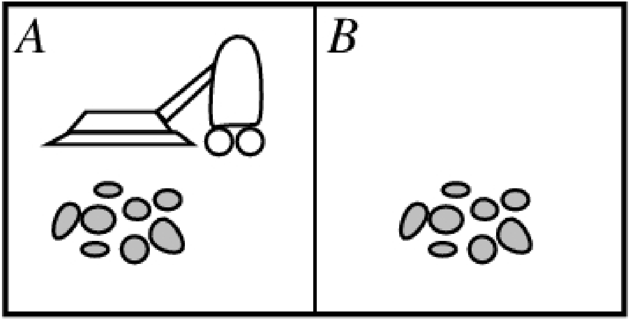


Agents include humans, robots, softbots, thermostats, etc.

The agent function maps from percept histories to actions:

$$f : \mathcal{P}^* \rightarrow \mathcal{A}$$

The agent program runs on the physical architecture to produce f



Percepts: location and contents, e.g., [*A*, *Dirty*]

Actions: *Left*, *Right*, *Suck*, *NoOp*

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 **right** agent function? Can it be implemented in a **small** agent program?
[note difference between agent function and agent program]

Fixed performance measure evaluates the sequence of environment states

Rationality

Fixed performance measure evaluates the sequence of environment states
Why is it important that it not be internal to the agent?

Fixed performance measure evaluates the sequence of environment states
Why is it important that it not be internal to the agent?

Possible performance measures:
– 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?

Rationality

Fixed performance measure evaluates the sequence of environment states
Why is it important that it not be internal to the agent?

Possible performance measures:
– 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**

Fixed performance measure evaluates the sequence of environment states
Why is it important that it not be internal to the agent?

Possible performance measures:
– 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 $\neq$ omniscient
– percepts may not supply all relevant information
Rational $\neq$ clairvoyant
– action outcomes may not be as expected
Hence, rational $\neq$ successful

Rationality

Fixed performance measure evaluates the sequence of environment states
Why is it important that it not be internal to the agent?

Possible performance measures:
– 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 $\neq$ omniscient
– percepts may not supply all relevant information
Rational $\neq$ clairvoyant
– action outcomes may not be as expected
Hence, rational $\neq$ successful

Rational $\Longrightarrow$ exploration, learning, autonomy

To design a rational agent, we must first specify the **task environment – PEAS**

Performance measure:??

Environment:??

Actuators:??

Sensors:??

Automated Taxi

Performance measure:??

Environment:??

Actuators:??

Sensors:??

Automated Taxi

Performance measure:?? safety, destination, profits, legality, comfort, . . .

Environment:??

Actuators:??

Sensors:??

Automated Taxi

Performance measure:?? safety, destination, profits, legality, comfort, ...

Environment:?? US streets/freeways, traffic, pedestrians, weather, ...

Actuators:??

Sensors:??

Automated Taxi

Performance measure:?? safety, destination, profits, legality, comfort, ...

Environment:?? US streets/freeways, traffic, pedestrians, weather, ...

Actuators:?? steering, accelerator, brake, horn, speaker/display, ...

Sensors:??

Automated Taxi

Performance measure:?? safety, destination, profits, legality, comfort, . . .

Environment:?? US streets/freeways, traffic, pedestrians, weather, . . .

Actuators:?? steering, accelerator, brake, horn, speaker/display, . . .

Sensors:?? video, accelerometers, gauges, engine sensors, keyboard, GPS, . . .

Internet Shopping Agent

Performance measure:??

Environment:??

Actuators:??

Sensors:??

Internet Shopping Agent

Performance measure:?? price, quality, appropriateness, efficiency

Environment:??

Actuators:??

Sensors:??

Internet Shopping Agent

Performance measure:?? price, quality, appropriateness, efficiency

Environment:?? current and future WWW sites, vendors, shippers

Actuators:??

Sensors:??

Internet Shopping Agent

Performance measure:?? price, quality, appropriateness, efficiency

Environment:?? current and future WWW sites, vendors, shippers

Actuators:?? display to user, follow URL, fill in form

Sensors:??

Internet Shopping Agent

Performance measure:?? price, quality, appropriateness, efficiency

Environment:?? current and future WWW sites, vendors, shippers

Actuators:?? display to user, follow URL, fill in form

Sensors:?? HTML pages (text, graphics, scripts)

Environment Types

- Do the agent's sensors give complete information (relevant to the choice of action) about the state of the environment at each point in time?
  - Fully- vs. partially-observable

Environment Types

- Do the agent's sensors give complete information (relevant to the choice of action) about the state of the environment at each point in time?
  - Fully- vs. partially-observable
- Does the agent operate in an environment with other agents?

Environment Types

- Do the agent's sensors give complete information (relevant to the choice of action) about the state of the environment at each point in time?
  - Fully- vs. partially-observable
- Does the agent operate in an environment with other agents?
  - Single vs. multi-agent (competitive, cooperative)

Environment Types

- Do the agent's sensors give complete information (relevant to the choice of action) about the state of the environment at each point in time?
  - Fully- vs. partially-observable
- Does the agent operate in an environment with other agents?
  - Single vs. multi-agent (competitive, cooperative)
- Is the next state of the environment completely determined by the current state and agent action?

- Do the agent's sensors give complete information (relevant to the choice of action) about the state of the environment at each point in time?
  - Fully- vs. partially-observable
- Does the agent operate in an environment with other agents?
  - Single vs. multi-agent (competitive, cooperative)
- Is the next state of the environment completely determined by the current state and agent action?
  - Deterministic vs. stochastic

Environment Types

- Do the agent's sensors give complete information (relevant to the choice of action) about the state of the environment at each point in time?
  - Fully- vs. partially-observable
- Does the agent operate in an environment with other agents?
  - Single vs. multi-agent (competitive, cooperative)
- Is the next state of the environment completely determined by the current state and agent action?
  - Deterministic vs. stochastic
- Does current action depend on previous actions?

- Do the agent's sensors give complete information (relevant to the choice of action) about the state of the environment at each point in time?
  - Fully- vs. partially-observable
- Does the agent operate in an environment with other agents?
  - Single vs. multi-agent (competitive, cooperative)
- Is the next state of the environment completely determined by the current state and agent action?
  - Deterministic vs. stochastic
- Does current action depend on previous actions?
  - Episodic vs. sequential

Environment Types

- Do the agent's sensors give complete information (relevant to the choice of action) about the state of the environment at each point in time?
  - Fully- vs. partially-observable
- Does the agent operate in an environment with other agents?
  - Single vs. multi-agent (competitive, cooperative)
- Is the next state of the environment completely determined by the current state and agent action?
  - Deterministic vs. stochastic
- Does current action depend on previous actions?
  - Episodic vs. sequential
- Can the environment change while the agent is deliberating?

Environment Types

- Do the agent's sensors give complete information (relevant to the choice of action) about the state of the environment at each point in time?
  - Fully- vs. partially-observable
- Does the agent operate in an environment with other agents?
  - Single vs. multi-agent (competitive, cooperative)
- Is the next state of the environment completely determined by the current state and agent action?
  - Deterministic vs. stochastic
- Does current action depend on previous actions?
  - Episodic vs. sequential
- Can the environment change while the agent is deliberating?
  - Static vs. dynamic

Environment Types

- Do the agent's sensors give complete information (relevant to the choice of action) about the state of the environment at each point in time?
  - Fully- vs. partially-observable
- Does the agent operate in an environment with other agents?
  - Single vs. multi-agent (competitive, cooperative)
- Is the next state of the environment completely determined by the current state and agent action?
  - Deterministic vs. stochastic
- Does current action depend on previous actions?
  - Episodic vs. sequential
- Can the environment change while the agent is deliberating?
  - Static vs. dynamic
- What is the domain of values for variables tracking environment state, agent state, and time?

Environment Types

- Do the agent's sensors give complete information (relevant to the choice of action) about the state of the environment at each point in time?
  - Fully- vs. partially-observable
- Does the agent operate in an environment with other agents?
  - Single vs. multi-agent (competitive, cooperative)
- Is the next state of the environment completely determined by the current state and agent action?
  - Deterministic vs. stochastic
- Does current action depend on previous actions?
  - Episodic vs. sequential
- Can the environment change while the agent is deliberating?
  - Static vs. dynamic
- What is the domain of values for variables tracking environment state, agent state, and time?
  - Discrete vs. continuous

Environment Types

- Do the agent's sensors give complete information (relevant to the choice of action) about the state of the environment at each point in time?
  - Fully- vs. partially-observable
- Does the agent operate in an environment with other agents?
  - Single vs. multi-agent (competitive, cooperative)
- Is the next state of the environment completely determined by the current state and agent action?
  - Deterministic vs. stochastic
- Does current action depend on previous actions?
  - Episodic vs. sequential
- Can the environment change while the agent is deliberating?
  - Static vs. dynamic
- What is the domain of values for variables tracking environment state, agent state, and time?
  - Discrete vs. continuous
- Does agent know outcomes of all its actions?

Environment Types

- Do the agent's sensors give complete information (relevant to the choice of action) about the state of the environment at each point in time?
  - Fully- vs. partially-observable
- Does the agent operate in an environment with other agents?
  - Single vs. multi-agent (competitive, cooperative)
- Is the next state of the environment completely determined by the current state and agent action?
  - Deterministic vs. stochastic
- Does current action depend on previous actions?
  - Episodic vs. sequential
- Can the environment change while the agent is deliberating?
  - Static vs. dynamic
- What is the domain of values for variables tracking environment state, agent state, and time?
  - Discrete vs. continuous
- Does agent know outcomes of all its actions?
  - Known vs. unknown

Most challenging environment:

Environment Types

- Do the agent's sensors give complete information (relevant to the choice of action) about the state of the environment at each point in time?
  - Fully- vs. partially-observable
- Does the agent operate in an environment with other agents?
  - Single vs. multi-agent (competitive, cooperative)
- Is the next state of the environment completely determined by the current state and agent action?
  - Deterministic vs. stochastic
- Does current action depend on previous actions?
  - Episodic vs. sequential
- Can the environment change while the agent is deliberating?
  - Static vs. dynamic
- What is the domain of values for variables tracking environment state, agent state, and time?
  - Discrete vs. continuous
- Does agent know outcomes of all its actions?
  - Known vs. unknown

Most challenging environment: Partially-observable,

Environment Types

- Do the agent's sensors give complete information (relevant to the choice of action) about the state of the environment at each point in time?
  - Fully- vs. partially-observable
- Does the agent operate in an environment with other agents?
  - Single vs. multi-agent (competitive, cooperative)
- Is the next state of the environment completely determined by the current state and agent action?
  - Deterministic vs. stochastic
- Does current action depend on previous actions?
  - Episodic vs. sequential
- Can the environment change while the agent is deliberating?
  - Static vs. dynamic
- What is the domain of values for variables tracking environment state, agent state, and time?
  - Discrete vs. continuous
- Does agent know outcomes of all its actions?
  - Known vs. unknown

Most challenging environment: Partially-observable, multi-agent,

Environment Types

- Do the agent's sensors give complete information (relevant to the choice of action) about the state of the environment at each point in time?
  - Fully- vs. partially-observable
- Does the agent operate in an environment with other agents?
  - Single vs. multi-agent (competitive, cooperative)
- Is the next state of the environment completely determined by the current state and agent action?
  - Deterministic vs. stochastic
- Does current action depend on previous actions?
  - Episodic vs. sequential
- Can the environment change while the agent is deliberating?
  - Static vs. dynamic
- What is the domain of values for variables tracking environment state, agent state, and time?
  - Discrete vs. continuous
- Does agent know outcomes of all its actions?
  - Known vs. unknown

Most challenging environment: Partially-observable, multi-agent, stochastic,

Environment Types

- Do the agent's sensors give complete information (relevant to the choice of action) about the state of the environment at each point in time?
  - Fully- vs. partially-observable
- Does the agent operate in an environment with other agents?
  - Single vs. multi-agent (competitive, cooperative)
- Is the next state of the environment completely determined by the current state and agent action?
  - Deterministic vs. stochastic
- Does current action depend on previous actions?
  - Episodic vs. sequential
- Can the environment change while the agent is deliberating?
  - Static vs. dynamic
- What is the domain of values for variables tracking environment state, agent state, and time?
  - Discrete vs. continuous
- Does agent know outcomes of all its actions?
  - Known vs. unknown

Most challenging environment: Partially-observable, multi-agent, stochastic, sequential,

Environment Types

- Do the agent's sensors give complete information (relevant to the choice of action) about the state of the environment at each point in time?
  - Fully- vs. partially-observable
- Does the agent operate in an environment with other agents?
  - Single vs. multi-agent (competitive, cooperative)
- Is the next state of the environment completely determined by the current state and agent action?
  - Deterministic vs. stochastic
- Does current action depend on previous actions?
  - Episodic vs. sequential
- Can the environment change while the agent is deliberating?
  - Static vs. dynamic
- What is the domain of values for variables tracking environment state, agent state, and time?
  - Discrete vs. continuous
- Does agent know outcomes of all its actions?
  - Known vs. unknown

Most challenging environment: Partially-observable, multi-agent, stochastic, sequential, dynamic,

Environment Types

- Do the agent's sensors give complete information (relevant to the choice of action) about the state of the environment at each point in time?
  - Fully- vs. partially-observable
- Does the agent operate in an environment with other agents?
  - Single vs. multi-agent (competitive, cooperative)
- Is the next state of the environment completely determined by the current state and agent action?
  - Deterministic vs. stochastic
- Does current action depend on previous actions?
  - Episodic vs. sequential
- Can the environment change while the agent is deliberating?
  - Static vs. dynamic
- What is the domain of values for variables tracking environment state, agent state, and time?
  - Discrete vs. continuous
- Does agent know outcomes of all its actions?
  - Known vs. unknown

Most challenging environment: Partially-observable, multi-agent, stochastic, sequential, dynamic, continuous, and

Environment Types

- Do the agent's sensors give complete information (relevant to the choice of action) about the state of the environment at each point in time?
  - Fully- vs. partially-observable
- Does the agent operate in an environment with other agents?
  - Single vs. multi-agent (competitive, cooperative)
- Is the next state of the environment completely determined by the current state and agent action?
  - Deterministic vs. stochastic
- Does current action depend on previous actions?
  - Episodic vs. sequential
- Can the environment change while the agent is deliberating?
  - Static vs. dynamic
- What is the domain of values for variables tracking environment state, agent state, and time?
  - Discrete vs. continuous
- Does agent know outcomes of all its actions?
  - Known vs. unknown

Most challenging environment: Partially-observable, multi-agent, stochastic, sequential, dynamic, continuous, and unknown.

Environment Types

- Do the agent's sensors give complete information (relevant to the choice of action) about the state of the environment at each point in time?
  - Fully- vs. partially-observable
- Does the agent operate in an environment with other agents?
  - Single vs. multi-agent (competitive, cooperative)
- Is the next state of the environment completely determined by the current state and agent action?
  - Deterministic vs. stochastic
- Does current action depend on previous actions?
  - Episodic vs. sequential
- Can the environment change while the agent is deliberating?
  - Static vs. dynamic
- What is the domain of values for variables tracking environment state, agent state, and time?
  - Discrete vs. continuous
- Does agent know outcomes of all its actions?
  - Known vs. unknown

Most challenging environment: Partially-observable, multi-agent, stochastic, sequential, dynamic, continuous, and unknown. Example?

Environment Types

- Do the agent's sensors give complete information (relevant to the choice of action) about the state of the environment at each point in time?
  - Fully- vs. partially-observable
- Does the agent operate in an environment with other agents?
  - Single vs. multi-agent (competitive, cooperative)
- Is the next state of the environment completely determined by the current state and agent action?
  - Deterministic vs. stochastic
- Does current action depend on previous actions?
  - Episodic vs. sequential
- Can the environment change while the agent is deliberating?
  - Static vs. dynamic
- What is the domain of values for variables tracking environment state, agent state, and time?
  - Discrete vs. continuous
- Does agent know outcomes of all its actions?
  - Known vs. unknown

Most challenging environment: Partially-observable, multi-agent, stochastic, sequential, dynamic, continuous, and unknown. Example?

Environment Types

	Solitaire	Backgammon	Internet shopping	Taxi
Observable??	Yes			

Environment Types

	Solitaire	Backgammon	Internet shopping	Taxi
Observable??	Yes	Yes		

Environment Types

	Solitaire	Backgammon	Internet shopping	Taxi
Observable??	Yes	Yes	No	

Environment Types

	Solitaire	Backgammon	Internet shopping	Taxi
Observable??	Yes	Yes	No	No
Deterministic??				

Environment Types

	Solitaire	Backgammon	Internet shopping	Taxi
Observable??	Yes	Yes	No	No
Deterministic??	Yes			

Environment Types

	Solitaire	Backgammon	Internet shopping	Taxi
Observable??	Yes	Yes	No	No
Deterministic??	Yes	No		

Environment Types

	Solitaire	Backgammon	Internet shopping	Taxi
Observable??	Yes	Yes	No	No
Deterministic??	Yes	No	Partly	

	Solitaire	Backgammon	Internet shopping	Taxi
Observable??	Yes	Yes	No	No
Deterministic??	Yes	No	Partly	No
Episodic??				

Environment Types

	Solitaire	Backgammon	Internet shopping	Taxi
Observable??	Yes	Yes	No	No
Deterministic??	Yes	No	Partly	No
Episodic??	No			

	Solitaire	Backgammon	Internet shopping	Taxi
Observable??	Yes	Yes	No	No
Deterministic??	Yes	No	Partly	No
Episodic??	No	No		

Environment Types

	Solitaire	Backgammon	Internet shopping	Taxi
Observable??	Yes	Yes	No	No
Deterministic??	Yes	No	Partly	No
Episodic??	No	No	No	

Environment Types

	Solitaire	Backgammon	Internet shopping	Taxi
Observable??	Yes	Yes	No	No
Deterministic??	Yes	No	Partly	No
Episodic??	No	No	No	No
Static??				

Environment Types

	Solitaire	Backgammon	Internet shopping	Taxi
Observable??	Yes	Yes	No	No
Deterministic??	Yes	No	Partly	No
Episodic??	No	No	No	No
Static??	Yes			

Environment Types

	Solitaire	Backgammon	Internet shopping	Taxi
Observable??	Yes	Yes	No	No
Deterministic??	Yes	No	Partly	No
Episodic??	No	No	No	No
Static??	Yes	Semi		

	Solitaire	Backgammon	Internet shopping	Taxi
Observable??	Yes	Yes	No	No
Deterministic??	Yes	No	Partly	No
Episodic??	No	No	No	No
Static??	Yes	Semi	Semi	

Environment Types

	Solitaire	Backgammon	Internet shopping	Taxi
Observable??	Yes	Yes	No	No
Deterministic??	Yes	No	Partly	No
Episodic??	No	No	No	No
Static??	Yes	Semi	Semi	No
Discrete??				

Environment Types

	Solitaire	Backgammon	Internet shopping	Taxi
Observable??	Yes	Yes	No	No
Deterministic??	Yes	No	Partly	No
Episodic??	No	No	No	No
Static??	Yes	Semi	Semi	No
Discrete??	Yes			

Environment Types

	Solitaire	Backgammon	Internet shopping	Taxi
Observable??	Yes	Yes	No	No
Deterministic??	Yes	No	Partly	No
Episodic??	No	No	No	No
Static??	Yes	Semi	Semi	No
Discrete??	Yes	Yes		

Environment Types

	Solitaire	Backgammon	Internet shopping	Taxi
Observable??	Yes	Yes	No	No
Deterministic??	Yes	No	Partly	No
Episodic??	No	No	No	No
Static??	Yes	Semi	Semi	No
Discrete??	Yes	Yes	Yes	

Environment Types

	Solitaire	Backgammon	Internet shopping	Taxi
Observable??	Yes	Yes	No	No
Deterministic??	Yes	No	Partly	No
Episodic??	No	No	No	No
Static??	Yes	Semi	Semi	No
Discrete??	Yes	Yes	Yes	No
Single-agent??				

Environment Types

	Solitaire	Backgammon	Internet shopping	Taxi
Observable??	Yes	Yes	No	No
Deterministic??	Yes	No	Partly	No
Episodic??	No	No	No	No
Static??	Yes	Semi	Semi	No
Discrete??	Yes	Yes	Yes	No
Single-agent??	Yes			

Environment Types

	Solitaire	Backgammon	Internet shopping	Taxi
Observable??	Yes	Yes	No	No
Deterministic??	Yes	No	Partly	No
Episodic??	No	No	No	No
Static??	Yes	Semi	Semi	No
Discrete??	Yes	Yes	Yes	No
Single-agent??	Yes	No		

Environment Types

	Solitaire	Backgammon	Internet shopping	Taxi
Observable??	Yes	Yes	No	No
Deterministic??	Yes	No	Partly	No
Episodic??	No	No	No	No
Static??	Yes	Semi	Semi	No
Discrete??	Yes	Yes	Yes	No
Single-agent??	Yes	No	Yes (except auctions)	

Environment Types

	Solitaire	Backgammon	Internet shopping	Taxi
Observable??	Yes	Yes	No	No
Deterministic??	Yes	No	Partly	No
Episodic??	No	No	No	No
Static??	Yes	Semi	Semi	No
Discrete??	Yes	Yes	Yes	No
Single-agent??	Yes	No	Yes (except auctions)	No

	Solitaire	Backgammon	Internet shopping	Taxi
Observable??	Yes	Yes	No	No
Deterministic??	Yes	No	Partly	No
Episodic??	No	No	No	No
Static??	Yes	Semi	Semi	No
Discrete??	Yes	Yes	Yes	No
Single-agent??	Yes	No	Yes (except auctions)	No

The environment type largely determines the agent design

Environment Types

	Solitaire	Backgammon	Internet shopping	Taxi
Observable??	Yes	Yes	No	No
Deterministic??	Yes	No	Partly	No
Episodic??	No	No	No	No
Static??	Yes	Semi	Semi	No
Discrete??	Yes	Yes	Yes	No
Single-agent??	Yes	No	Yes (except auctions)	No

The environment type largely determines the agent design

The real world is (of course) partially observable, stochastic, sequential, dynamic, continuous, multi-agent

Environment Types

	Solitaire	Backgammon	Internet shopping	Taxi
Observable??	Yes	Yes	No	No
Deterministic??	Yes	No	Partly	No
Episodic??	No	No	No	No
Static??	Yes	Semi	Semi	No
Discrete??	Yes	Yes	Yes	No
Single-agent??	Yes	No	Yes (except auctions)	No

The environment type largely determines the agent design

The real world is (of course) partially observable, stochastic, sequential, dynamic, continuous, multi-agent

Agent Types

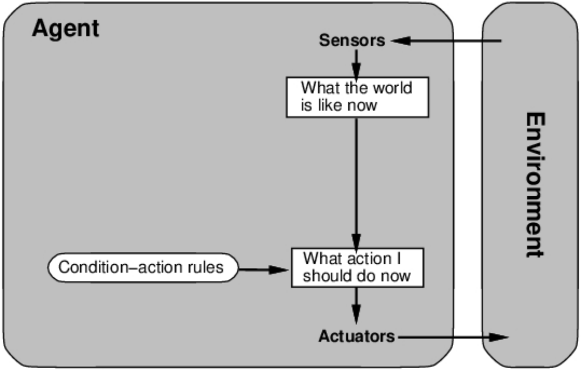
Categorization of agent programs
Four basic types in order of increasing generality:

- ♢ simple reflex agents
- ♢ reflex agents with state (model-based reflex agents)
- ♢ goal-based agents
- ♢ utility-based agents

All these can be turned into:

- ♢ learning agents

- Choose action based on current percept (from list of condition-action rules)
- Consider how the world IS



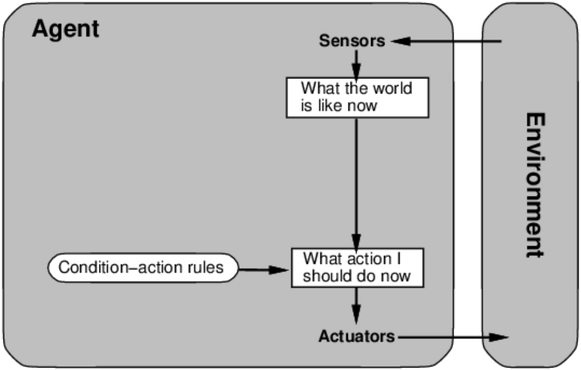
Example of a Reflex Agent

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*

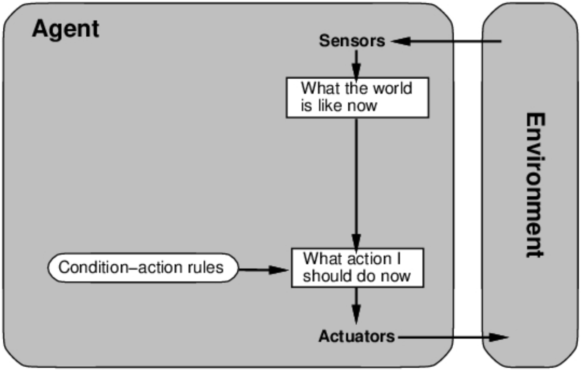
```
(setq joe (make-agent :name 'joe :body (make-agent-body)
                      :program (make-reflex-vacuum-agent-program)))

(defun make-reflex-vacuum-agent-program ()
  #'(lambda (percept)
      (let ((location (first percept)) (status (second percept)))
        (cond ((eq status 'dirty) 'Suck)
              ((eq location 'A) 'Right)
              ((eq location 'B) 'Left)))))
```

- Choose action based on current percept (from list of condition-action rules)
- Consider how the world IS



- Can a reflex agent be rational?
- What to do in partially-observable environments?

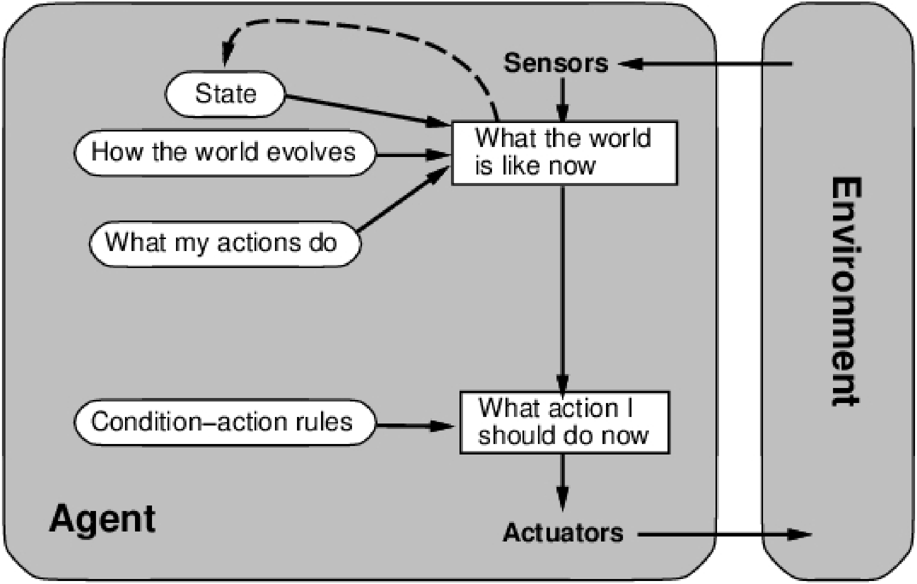
- Choose action based on current percept (from list of condition-action rules)
- Consider how the world IS



- Can a reflex agent be rational?
- What to do in partially-observable environments?

Planning Agents

- Ask "what if"
- Must formulate a goal (test)
  - Goal-based agent (atomic environment state)
  - Planning agent (factored or structured environment state)
- Decisions based on (hypothesized) consequences of actions
- Consider how the world WOULD BE
- Must have a model of how the world evolves in response to actions
- Complete versus Optimal planning
- Planning versus replanning (for dynamic environments)


Sensors
State
How the world evolves
What my actions do
What the world is like now
Condition–action rules
What action I should do now
Agent
Actuators
Environment

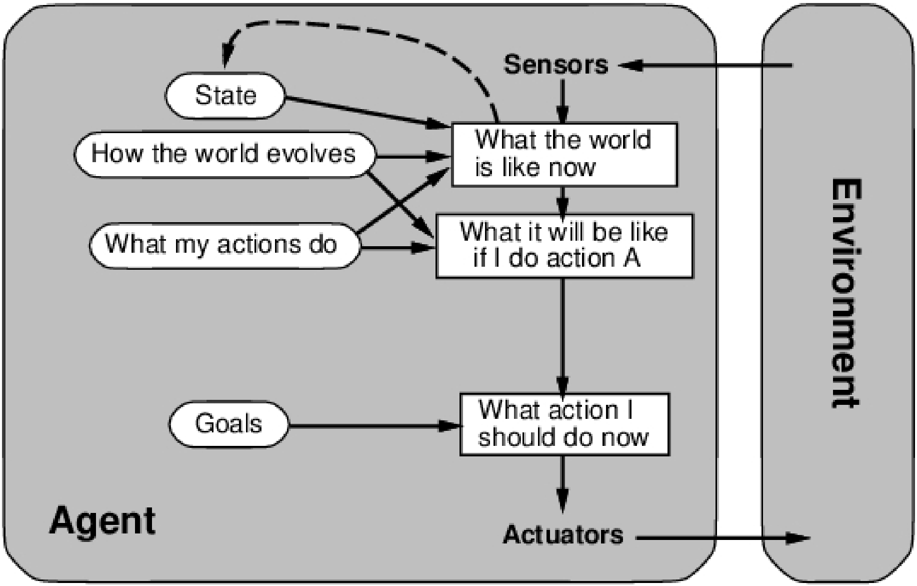
Example

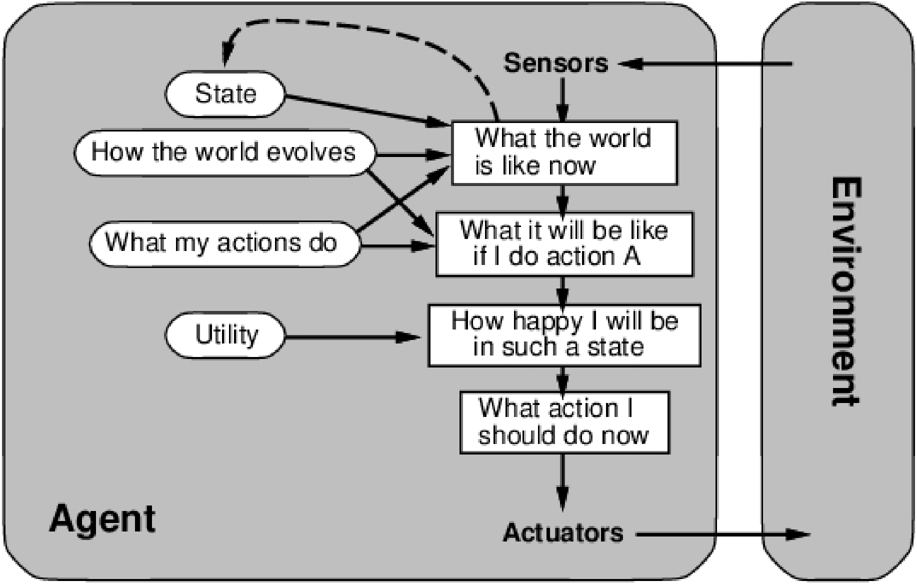
function REFLEX-VACUUM-AGENT([*location*,*status*]) **returns** an action
static: *last_A*, *last_B*, numbers, initially ∞

if *status* = *Dirty* **then** ...

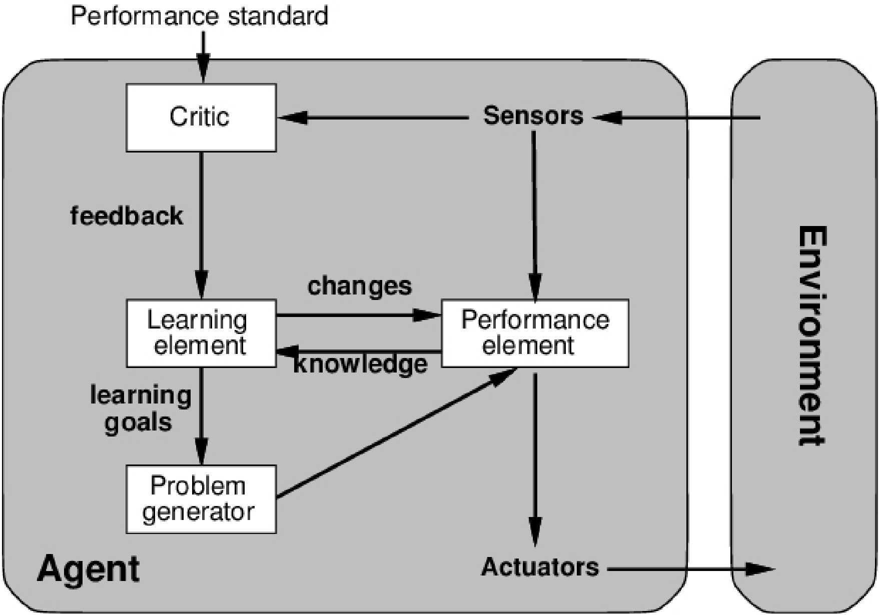
```
(defun make-reflex-vacuum-agent-with-state-program ()
  (let ((last-A infinity) (last-B infinity))
  #'(lambda (percept)
      (let ((location (first percept)) (status (second percept)))
        (incf last-A) (incf last-B)
        (cond
         ((eq status 'dirty)
          (if (eq location 'A) (setq last-A 0) (setq last-B 0))
          'Suck)
         ((eq location 'A) (if (> last-B 3) 'Right 'NoOp))
         ((eq location 'B) (if (> last-A 3) 'Left 'NoOp)))))))
```

Goal-based Agents





State
How the world evolves
What my actions do
Utility
Sensors
What the world is like now
What it will be like if I do action A
How happy I will be in such a state
What action I should do now
Actuators
Agent
Environment


Performance standard
Critic
Sensors
feedback
changes
Learning element
Performance element
knowledge
learning goals
Problem generator
Agent
Actuators
Environment

Atomic:

- Each state of the world is indivisible
- E.g., cities to a traveling salesman
- Search and game playing (chapters 3-5), Hidden Markov models (chapter 15), and Markov decision processes (chapter 17)

Factored:

- A state split into a fixed set of variables/attributes
- States can share attributes (and more easily transform into each-other)
- Can represent uncertainty (unevaluated attributes)
- Constraint satisfaction (chapter 6), propositional logic (chapter 7), planning (chapters 10-11), Bayesian networks (chapters 13-16), machine learning (chapters 18, 20, 21)

Structured:

- Represent objects and relations
- Relational databases and first-order logic (chapters 8,9, 12), first-order probability models (chapter 14), knowledge-based learning (chapter 19), natural language processing (chapters 22-23)

Summary on Intelligent Agents

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

- PEAS descriptions define task environments
- Environments are categorized along several dimensions:
  - fully- vs. partially-observable
  - deterministic vs. stochastic
  - episodic vs. sequential
  - static vs. dynamic
  - discrete vs. continuous
  - single- vs. multi-agent
- Agents are categorized by the agent programs:
  - reflex
  - reflex with state
  - goal-based
  - utility-based
  - learning-based