CS 4700:
Foundations of Artificial Intelligence

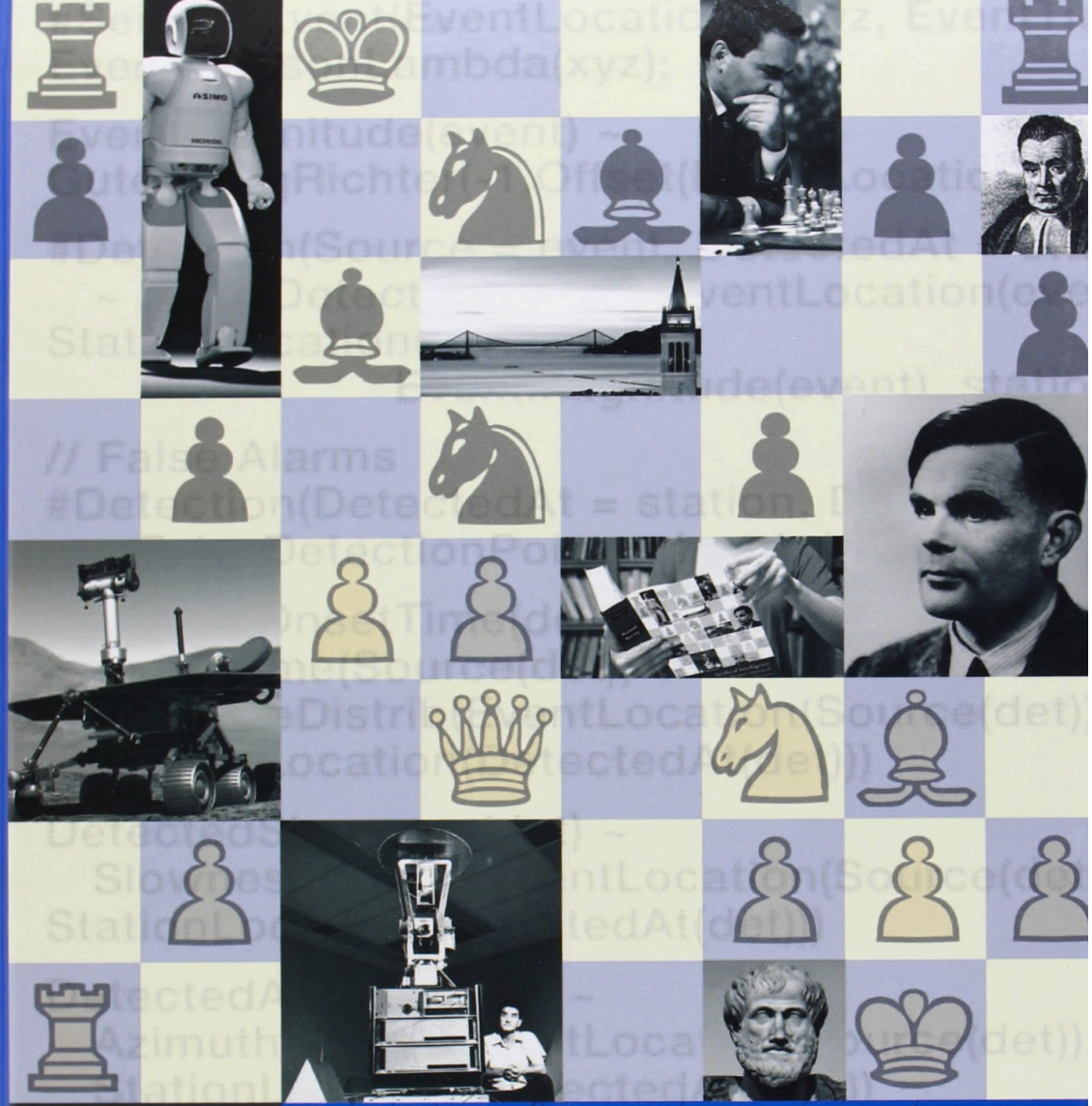
Instructor:

Prof. Selman

selman@cs.cornell.edu

Introduction

(Reading R&N: Chapter 1)



Stuart
Russell
Peter
Norvig

Artificial Intelligence
A Modern Approach
Third Edition

Course Administration (separate slides) ✓

What is Artificial Intelligence?

Course Themes, Goals, and Syllabus

AI: Goals

Ambitious goals:

- **understand "intelligent" behavior**
- **build "intelligent" agents / artifacts**

autonomous systems

understand human cognition (learning, reasoning, planning, and decision making) as a computational process.

What is Intelligence?

Intelligence:

- **capacity to learn and solve problems"**
  (Webster dictionary)
- **the ability to act rationally**

Hmm… Not so easy to define.

What is AI?

Views of AI fall into four different perspectives --- two dimensions:

1) Thinking versus Acting
2) Human versus Rational (which is "easier"?)

	Human-like Intelligence	"Ideal" Intelligent/ Pure Rationality
Thought/ Reasoning ("modeling thought / brain)	**2. Thinking humanly**	**3. Thinking Rationally**
Behavior/ Actions "behaviorism" "mimics behavior"	**1. Acting Humanly**	**4. Acting Rationally**

1. Acting Humanly

	Human-like Intelligence	"Ideal" Intelligent/ Rationally
Thought/ Reasoning	2. Thinking humanly	3. Thinking Rationally
Behavior/ Actions	1. Acting Humanly →Turing Test	4. Acting Rationally

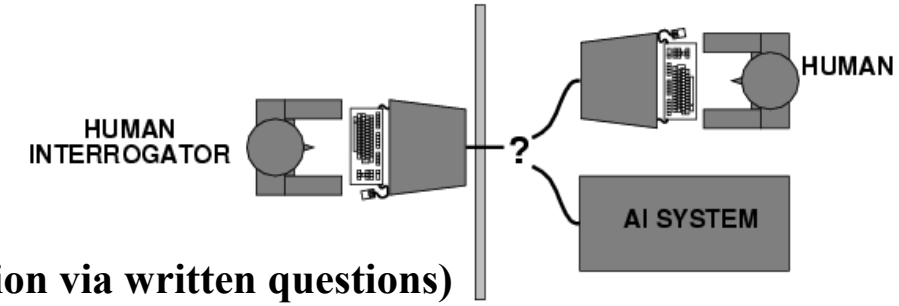
Alan Turing

Acting humanly: Turing Test

Turing (1950) "Computing machinery and intelligence"

"Can machines think?" "Can machines behave intelligently?"

- **Operational test for intelligent behavior: the Imitation Game**



(interaction via written questions)

AI system passes if interrogator cannot tell which one is the machine.

No computer vision or robotics or physical presence required!

Predicted that by 2000, a machine might have a 30% chance of fooling a lay person for 5 minutes.

Achieved. (Siri! ☺)

But, by scientific consensus, we are still several decades away from truly passing the Turing test (as the test was intended).

Trying to pass the Turing test: Some Famous Human Imitation "Games"

1960s ELIZA

- **Joseph Weizenbaum**
- **Rogerian psychotherapist**

1990s ALICE

Loebner prize

- **win $100,000 if you pass the test**

Still, passing Turing test is of somewhat questionable value.

Because, deception appears *required* and *allowed*!

Consider questions: Where were you born? How tall are you?

ELIZA: impersonating a Rogerian psychotherapist

1960s ELIZA Joseph Weizenbaum

You: Well, I feel sad
Eliza: Do you often feel sad?
You: not very often.
Eliza: Please go on.

Alternative

See: The New Yorker, August 16, 2013
Why Can't My Computer Understand Me?
Posted by *Gary Marcus*

http://www.newyorker.com/online/blogs/elements/2013/08/why-cant-my-computer-understand-me.html

Discusses alternative test by Hector Levesque:
http://www.cs.toronto.edu/~hector/Papers/ijcai-13-paper.pdf

2. Thinking Humanly

	Human-like Intelligence	"Ideal" Intelligent/ Rationally
Thought/ Reasoning	2. Thinking humanly → **Cognitive Modeling**	Thinking Rationally
Behavior/ Actions	Acting Humanly →Turing Test	Acting Rationally

here

Thinking humanly: modeling cognitive processes

Requires scientific theories of internal activities of the brain.

1) Cognitive Science (top-down) computer models + experimental techniques from psychology

→ Predicting and testing behavior of human subjects

2) Cognitive Neuroscience (bottom-up)

→ Direct identification from neurological data

Distinct disciplines but especially 2) has become very active. Connection to AI: Neural Nets. (Large Google / MSR / Facebook AI Lab efforts.)

Neuroscience: The Hardware

The brain

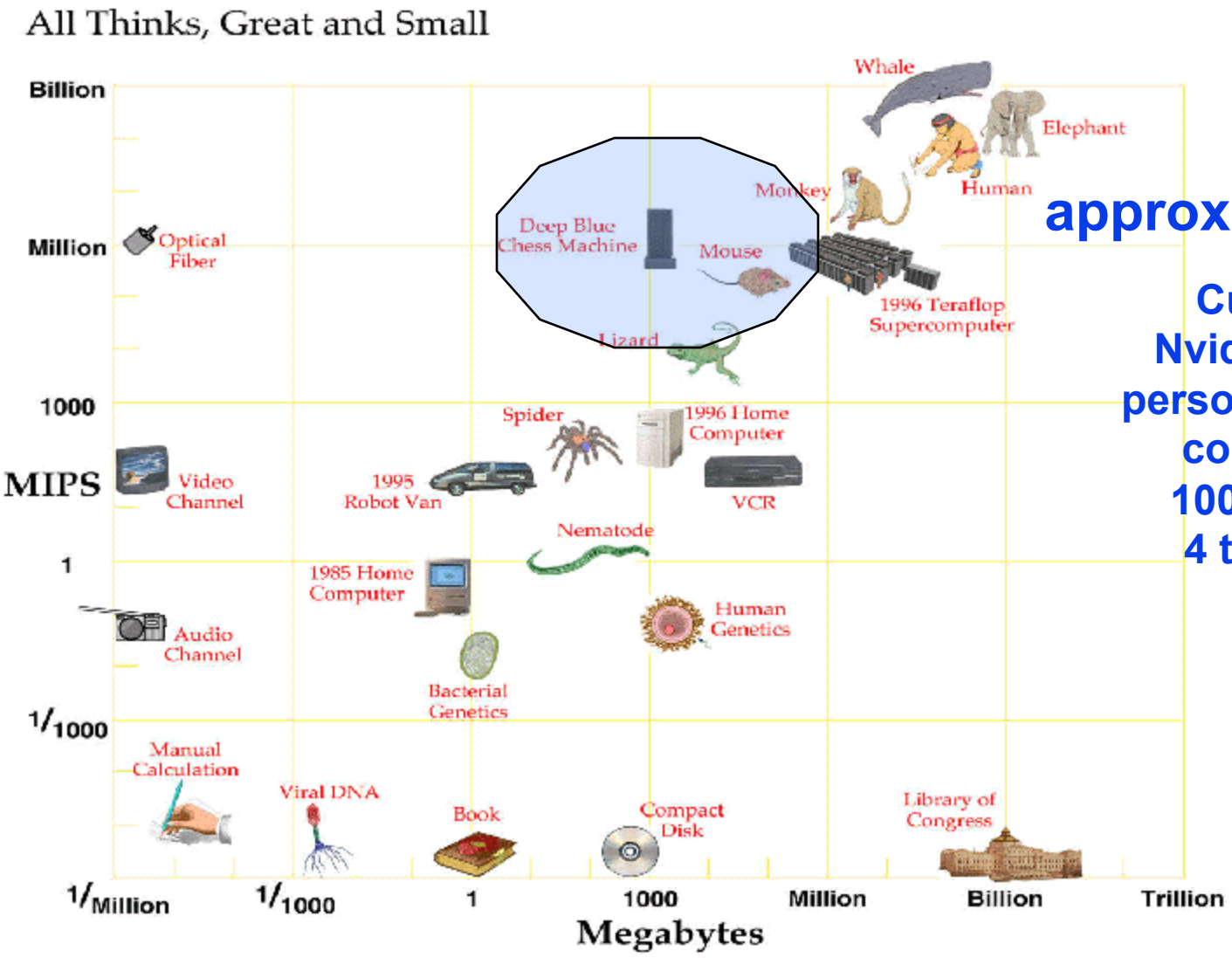
- a neuron, or nerve cell, is the basic information
- processing unit (10^11)
- many more synapses (10^14) connect the neurons
- cycle time: 10^(-3) seconds (1 millisecond)

How complex can we make computers?

- 10^9 or more transistors per CPU
- Ten of thousands of cores, 10^10 bits of RAM
- cycle times: order of 10^(-9) seconds

Numbers are getting close! Hardware will surpass human brain within next 20 yrs.

Computer vs. Brain



approx. 2025

Current:
Nvidia: tesla personal super-computer
1000 cores
4 teraflop

Aside: Whale vs. human brain

So,

- In near future, we can have computers with as many processing elements as our brain, but:

  far fewer interconnections (wires or synapses)

  then again, much faster updates.

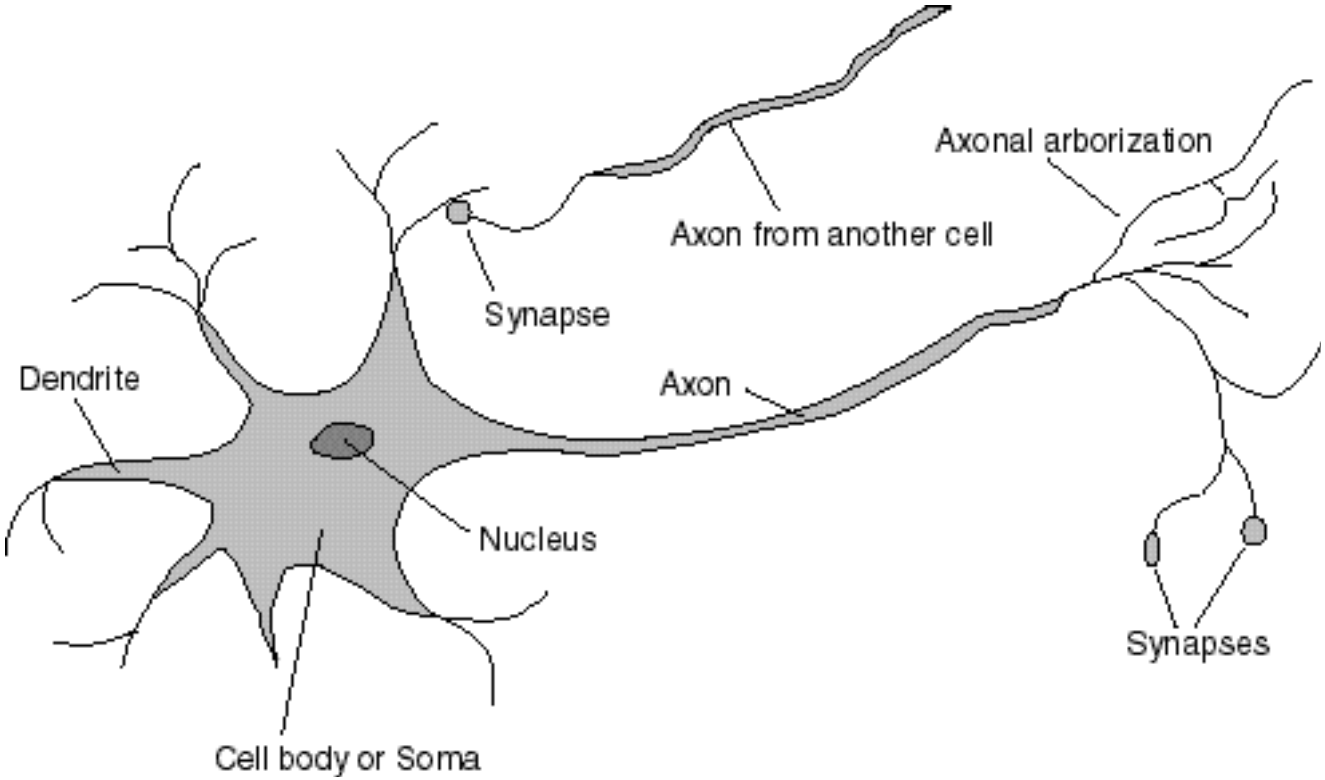
Fundamentally different hardware may

require fundamentally different algorithms!

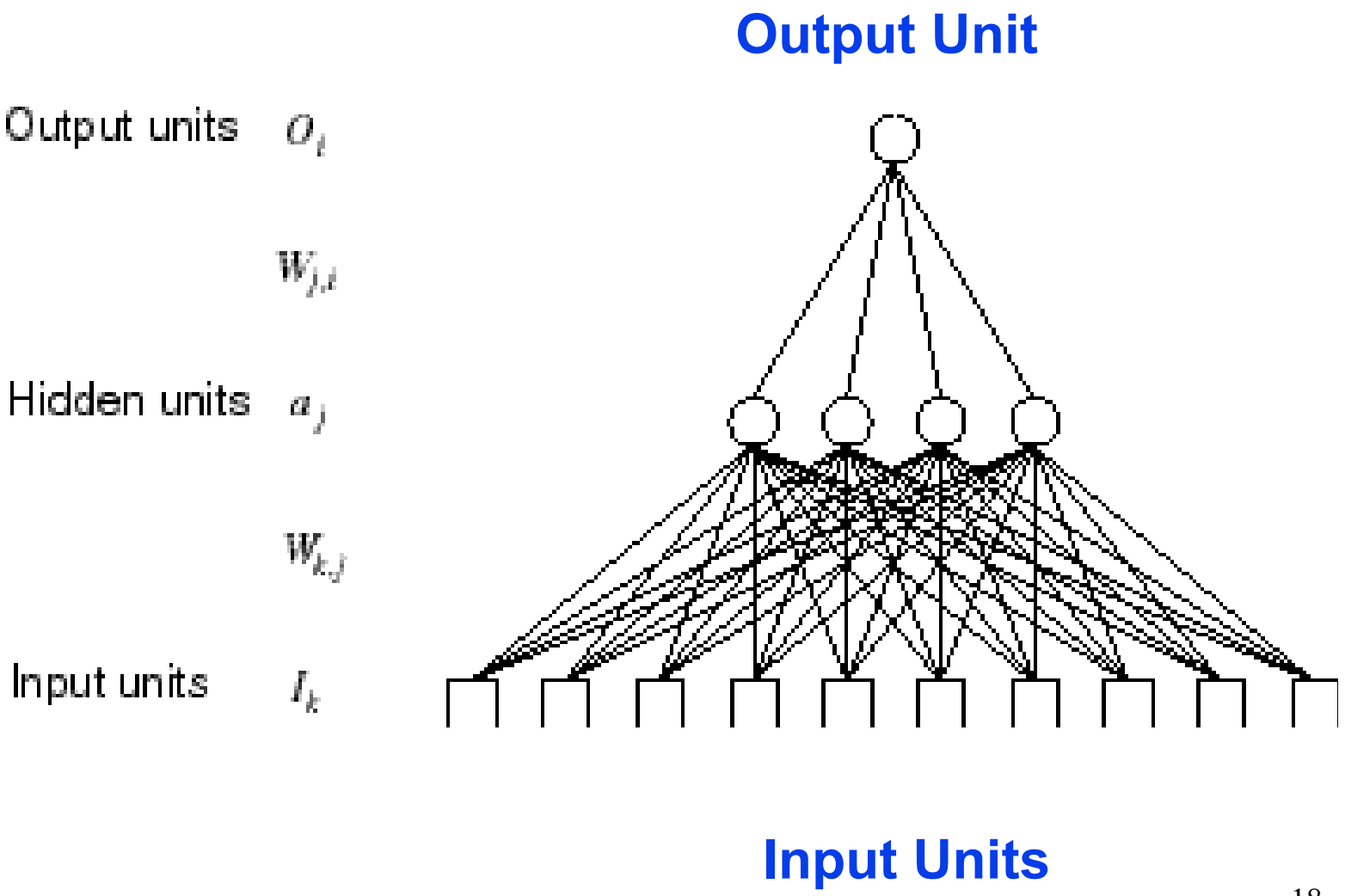
- Still an open question.
- Neural net research.
- Can a digital computer simulate our brain?

Likely: Church-Turing Thesis
(But, might we need quantum computing?)
(Penrose; consciousness; free will)

A Neuron


Axonal arborization
Axon from another cell
Synapse
Dendrite
Axon
Nucleus
Synapses
Cell body or Soma

An Artificial Neural Network (Perceptrons)



An artificial neural network is an abstraction (well, really, a "drastic simplification") of a real neural network.

Start out with random connection weights on the links between units. Then train from input examples and environment, by changing network weights.

Recent breakthrough: Deep Learning

(automatic discovery of "deep" features by a multi-layer neural network.)

Deep learning is bringing perception (hearing & vision) within reach.

3. Thinking Rationally

	Human-like Intelligence	"Ideal" Intelligent/ Rationally
Thought/ Reasoning	Thinking humanly → Cognitive Modeling	**3. Thinking Rationally →formalizing "Laws of Thought"**
Behavior/ Actions	Acting Humanly →Turing Test	Acting Rationally

Thinking rationally: formalizing the "laws of thought"

Long and rich history!

Logic: Making the right inferences!

Remarkably effective in science, math, and engineering.

Several Greek schools developed various forms of *logic*: *notation* and *rules of derivation* for thoughts.

Aristotle: what are correct arguments/thought processes? (characterization of "right thinking").

Syllogisms
Aristotle

Socrates is a man

All men are mortal

Therefore, Socrates is mortal

Can we mechanize it? (syntactic; strip interpretation)

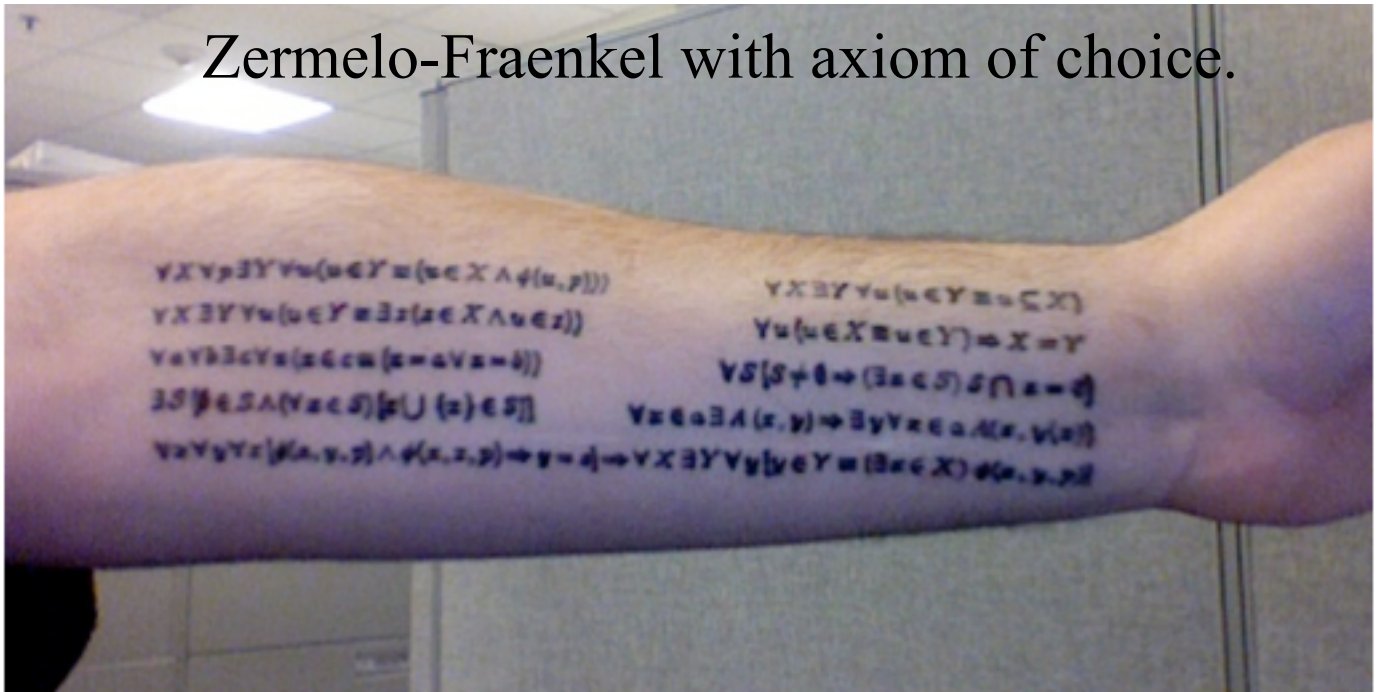
Use: legal cases, diplomacy, ethics etc. (?)

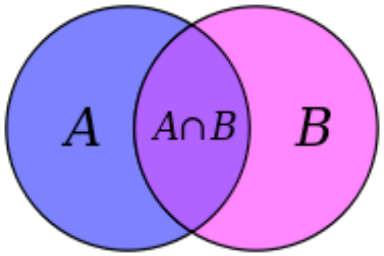
More contemporary logicians (e.g. Boole, Frege, and Tarski).
Ambition: Developing the "language of thought."
Direct line through mathematics and philosophy to modern AI.

Key notion:

Inference derives new information from stored facts.

Axioms can be very compact. E.g. much of mathematics can be derived (in principle) from the logical axioms of





Also, Godel's incompleteness.

Limitations:

- **Not all intelligent behavior is mediated by logical deliberation (much appears not…)**

- **(Logical) representation of knowledge underlying intelligence is quite non-trivial. Studied in the area of "knowledge representation." Also brings in probabilistic representations. E.g. Bayesian networks and graphical models.**

- **What is the purpose of thinking?**
- **What thoughts should I have?**

- **Seems to require some connection to "acting in the world."**
- **We ("agents") want/need to affect our environment (in part for survival).**

4. Acting Rationally

	Human-like Intelligence	"Ideal" Intelligent/ Rationally
Thought/ Reasoning	Thinking humanly → Cognitive Modeling	Thinking Rationally →formalizing "Laws of Thought"
Behavior/ Actions	Acting Humanly →Turing Test	**Acting Rationally**

Rational agents

- **An agent is an entity that perceives and acts in the world (i.e. an "autonomous system" (e.g. self-driving cars) / physical robot or software robot (e.g. an electronic trading system))**

  This course is about designing rational agents

- **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

  →design ***best*** program for given machine resources

  →"Limited rationality"

Building Intelligent Machines

I Building exact models of human cognition
view from psychology, cognitive science, and neuroscience

II Developing methods to match or exceed human performance in certain domains, possibly by very different means

Main focus of current AI.

But, I) often provides inspiration for II). Also, Neural Nets blur the separation.

Key research areas in AI

Problem solving, planning, and search --- generic problem solving architecture based on ideas from cognitive science (game playing, robotics).

Knowledge Representation – to store and manipulate information (logical and probabilistic representations)

Automated reasoning / Inference – to use the stored information to answer questions and draw new conclusions

Machine Learning – intelligence from data; to adapt to new circumstances and to detect and extrapolate patterns

Natural Language Processing – to communicate with the machine

Computer Vision --- processing visual information

Robotics --- Autonomy, manipulation, full integration of AI capabilities