Artificial intelligence (and Searle's objection)

COS 116: 4/29/2008

Sanjeev Arora

Artificial Intelligence

- Definition of AI (Merriam-Webster):
  - The capability of a machine to imitate intelligent human behavior
  - Branch of computer science dealing with the simulation of intelligent behavior in computers
- Learning:
  - To gain knowledge or understanding of or skill in by study, instruction, or experience
  - Machine learning (last lecture) - branch of AI

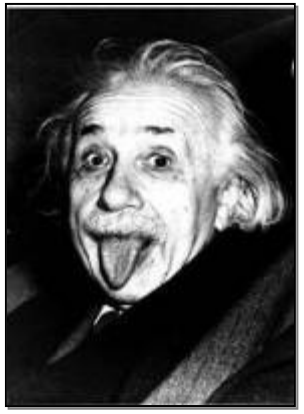
Intelligence in animal world

Is an ant intelligent?

- Build huge, well-structured colonies organized using chemical-based messaging (“Super-organism”)

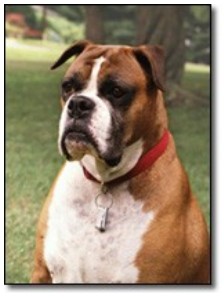
What about dogs?

Deep mystery: How do higher animals (including humans) learn?

How does  become 

A crude first explanation: Behaviorism [Pavlov 1890's, Skinner 1930's]

- Animals and humans can be understood in a "black box" way as a sum total of all direct conditioning events
- Bell → "Food is coming" →Salivate

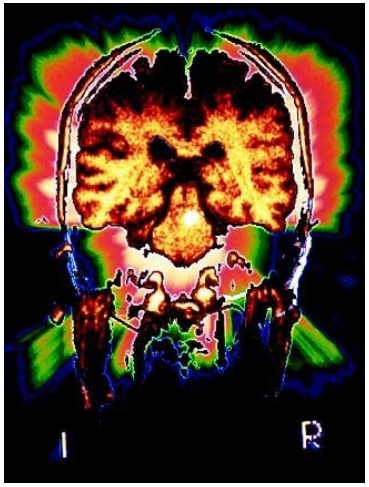


- "This person likes me more if I call her "Mama" and that one likes me more if I call him "Papa".

Aside: What does behaviorism imply for societal organization?

More thoughts on behaviorism

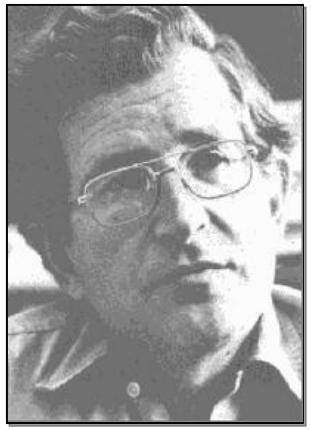
Original motivation: Cannot look inside the working brain anyway, so theory that assumes anything about its working is not scientific or testable.



Today

Little insight into how to design machines with intelligence. How did dogs, rats, humans sort through sensory experiences to understand reward/punishment?

Chomsky's influential critique of Behaviorism [1957]



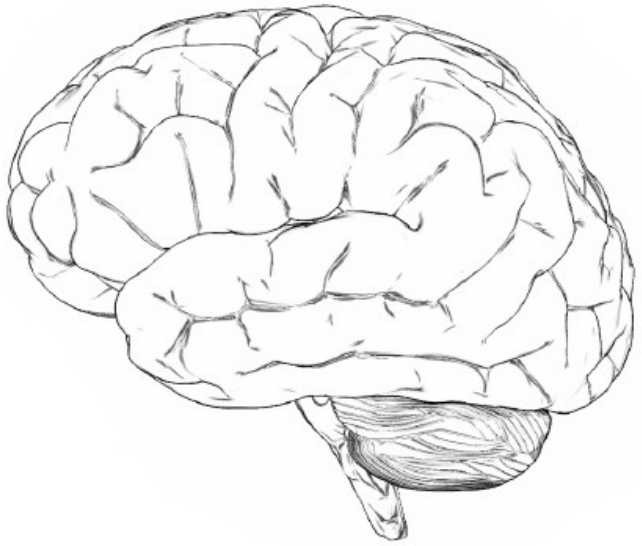
- "Internal mental structures crucial for learning."

  Evidence: universal linguistic rules ("Chomsky grammars"); "self-correction" in language learning, ability to appreciate puns.

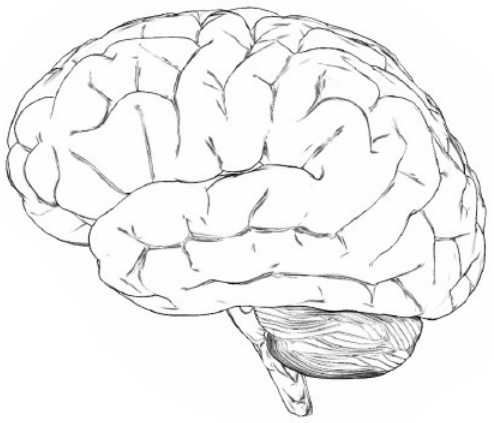
1. Brain is "prewired" for language.
2. Must understand mental structures to understand behavior

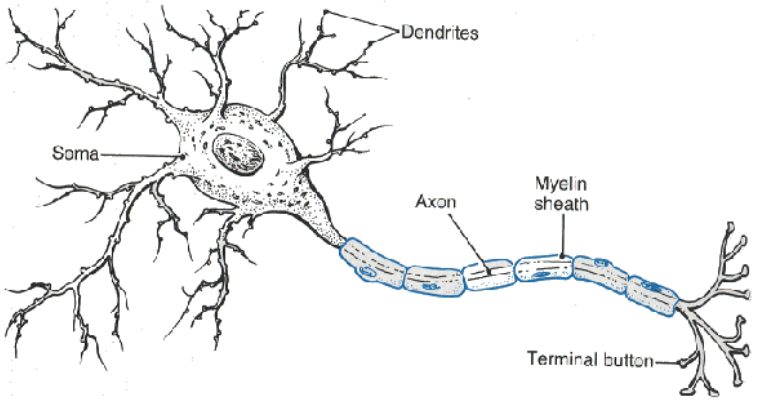
Presenting:

Your brain



The brain

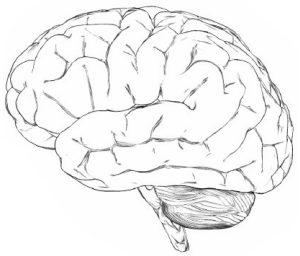




- Network of 100 billion neurons
- Evidence of timing mechanisms ("clock")
- About 100 firings per second
  - Total of 10^{13} firings ("operations") per second
  - Number of operations per sec in fast desktop PC: 10^{10}
  - Kurzweil predicts PC will match brain computationally by 2020

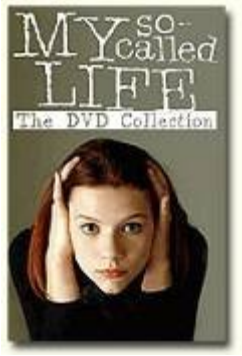
A comparison

Your brain



10^{11} neurons

Your life on a DVD



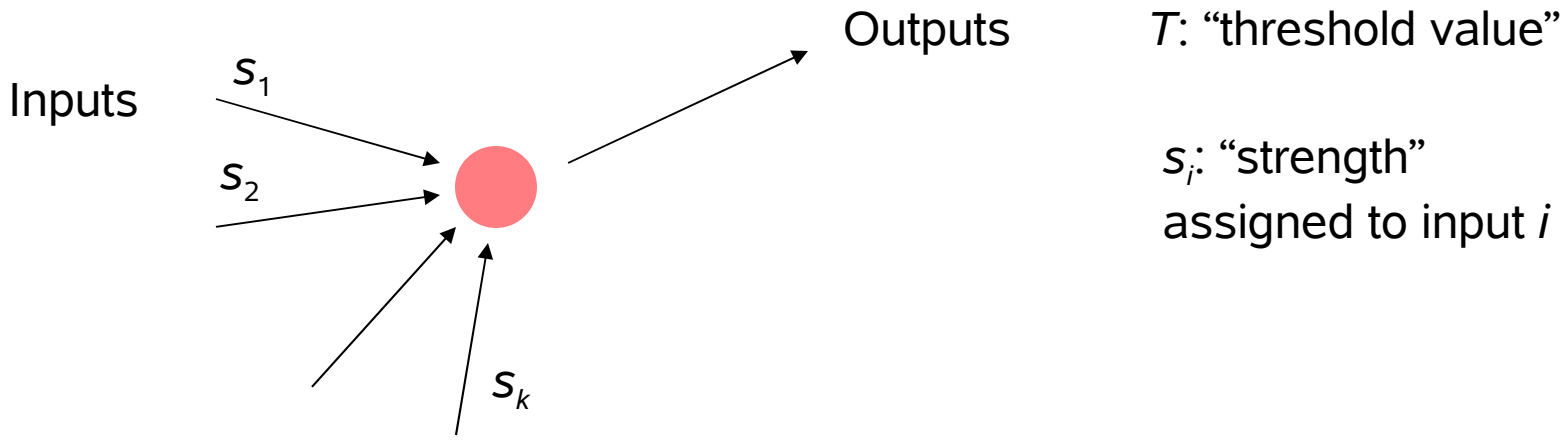
4.3 Gb for 3 hours

> 10^{17} bytes for entire life

Conclusion: Brain must contain structures that compress information and store it in an interconnected way for quick associations and retrieval

A simplistic model of neurons—Neural Net [McCulloch – Pitts 1943]

- Neuron computes "thresholds"



- Take the sum of strengths of all neighbors that are firing
- If sum > T, fire

Does a neural network model remind you of something??

Why AI is feasible in principle: the simulation argument

- Write a simulation program that simulates all 10^{11} neurons in the brain and their firings.
- For good measure, also simulates underlying chemistry, blood flow, etc.
- In principle doable on today's fastest computers
- Practical difficulty: How to figure out properties (threshold value, s_i) of each of 10^{10} neurons, the intricate chemistry

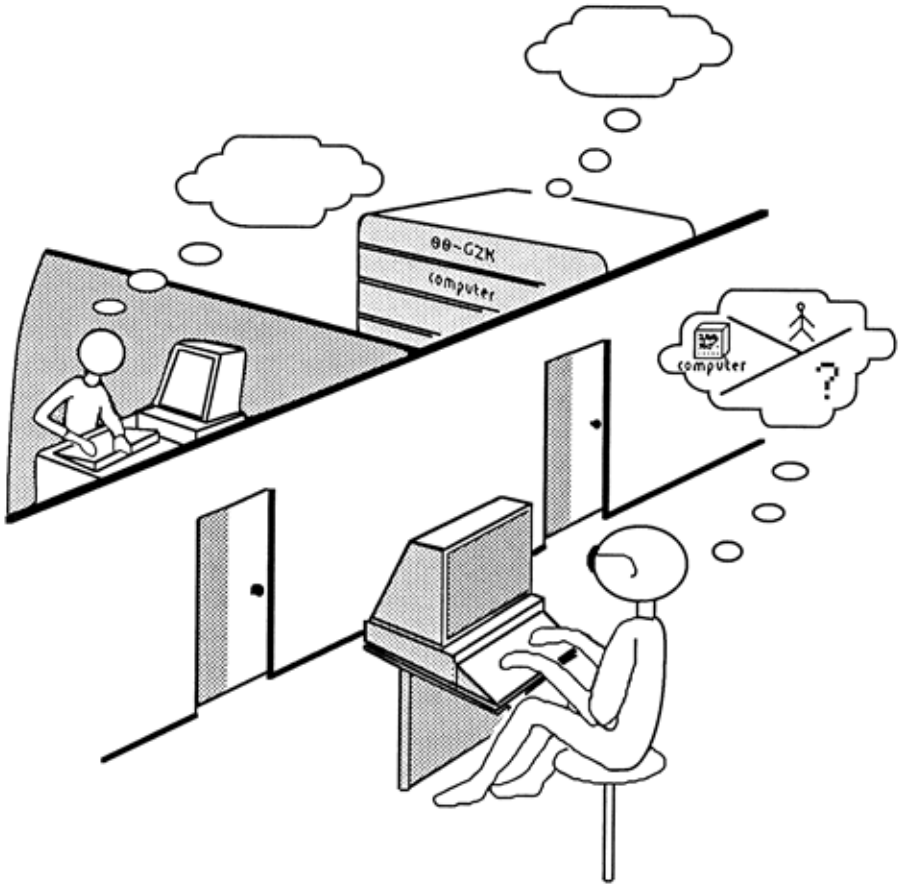
Hope

Maybe the brain is organized around simpler principles.

Simple machine learning algorithms from last lecture provide a hint?

Turing test (Turing 1950; see turinghub.com)

- You are allowed to chat with machine or a human (don't know which)
- You have to guess at the end if you were talking to a machine or human. (Machine wins if you have only 50-50 success rate.)
- Note: Impossible for machine to store answers to all possible 5-minute conversations!



What are strengths and weaknesses of the Turing test?

(Feel free to contrast with other tests, e.g. Stanford-Binet IQ, SAT)

Strengths

- Not reducible to formula
- No obvious way to cheat
- Customizable to different topics
- Behavioral/ black box.

Weaknesses

- Too subjective
- Too human-centric
- Too behaviorist.
- Tests only one kind of intelligence.

Poll: Did you like Searle's article?

(as in, interesting, thought-provoking)

Poll: Which of the following are Searle's conclusions?

1. It is impossible for a computer to pass the Turing test.
2. The Turing test is not a valid test for whether a machine can "think."
3. A computer is nothing but a rulebook applied mechanically. The rulebook doesn't understand Chinese, so neither does the computer.
4. There is a big difference between syntax and semantics. Computers deal with symbols, and hence with syntax. Thinking is about semantics.

Some background: Strong AI

A machine able to:

- reason, use strategy, solve puzzles, and make judgements under uncertainty;
- represent knowledge, including commonsense knowledge;
- plan;
- learn;
- communicate in natural language;
- perceive, and especially see;
- have social intelligence;
- be able to move and manipulate objects (robotics);
- and the ability to integrate all these skills towards common goals.

Other potentially relevant traits (unclear if necessary or even definable): consciousness, wisdom, self-awareness,…

What role does the Chinese room argument play in the article?

- explain to the average reader what a computer program is: a long rulebook (recall: Turing Post program, pseudocode)

- appeal to the “obvious” intuition that a rulebook cannot think

(Caution:His “intuition” ignores processing speed.)

Question: What does Searle think of the “Simulation Argument” for AI?

My problems with Searle's paper

- He rejects Turing test but gives no alternative definition of "thinking" or "mind."

- Scientifically speaking, no clear line between
  (a) hardware and software ("Game of life.")
  (b) syntax and semantics ("genetic code.")

- He often resorts to ridicule (a bad sign!)

- If a machine ever passes Turing test, exhibiting accurate emotions, social skills etc., this would seriously make *me* wonder if it has some kind of mind in it.

Time warp

Rene Descartes (1637) “I think therefore I am.”