Artificial intelligence

COS 116, Spring 2012

Adam Finkelstein

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

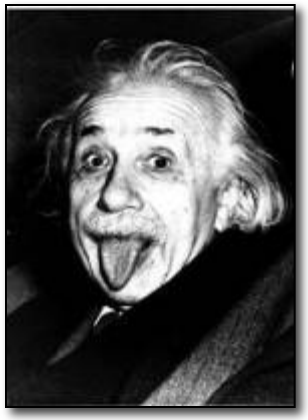
Intelligence in animal world

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 

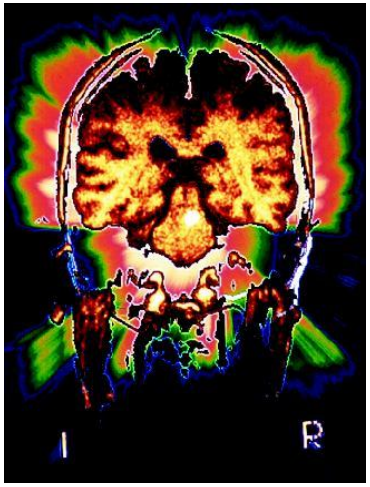
A crude first explanation: Behaviorism [Pavlov 1890's, Skinner 1930's]

- 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”.

More thoughts on behaviorism

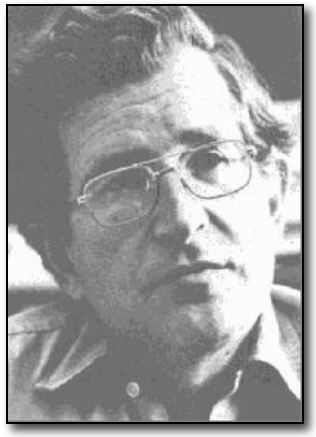
Original motivation: Cannot look inside the working brain, so theory that assumes anything about its working is not scientific or testable.



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."

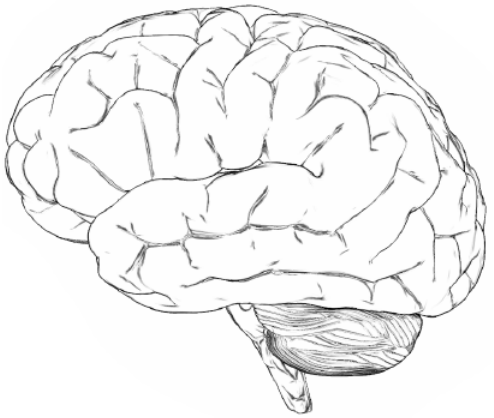
  Evidence: universal linguistic rules ("Chomsky grammars"); "self-correction" in language learning, ability to appreciate puns.

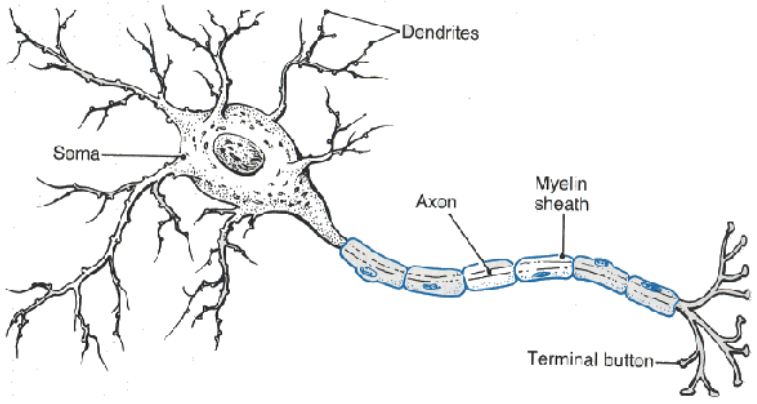
1. Brain is "prewired" for language.
2. Must understand mental structures to understand behavior

Presenting:

Your brain

The brain

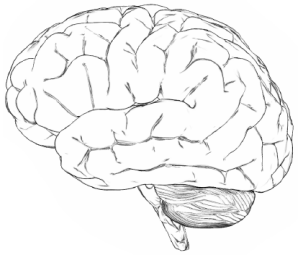




- Network of 100 billion neurons
- Evidence of timing mechanisms (“clock”)
- About 100 firings per second
  - Total of 10^{13} firings (“operations”) per second
  - Number of operations per sec in fast desktop PC: 10^{10}
  - Kurzweil predicts PC will match brain computationally by 2020

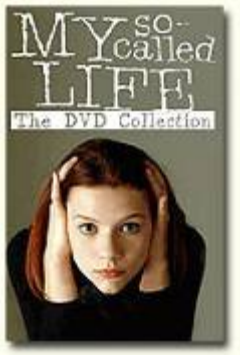
A comparison

Your brain



10^{11} neurons

Your life on a DVD



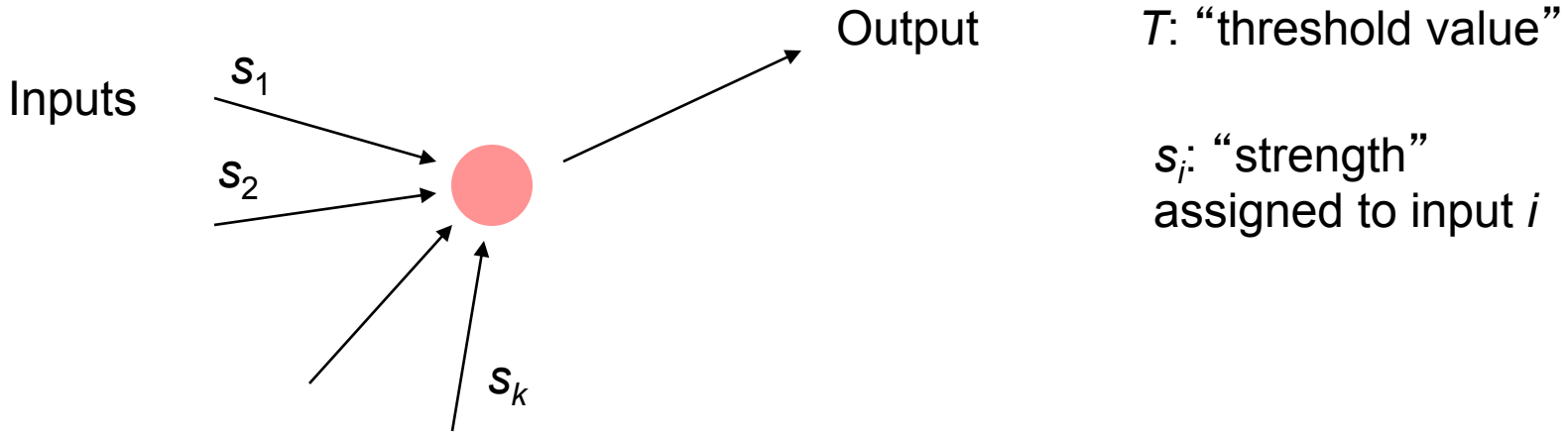
4.3 Gb for 3 hours

> 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"



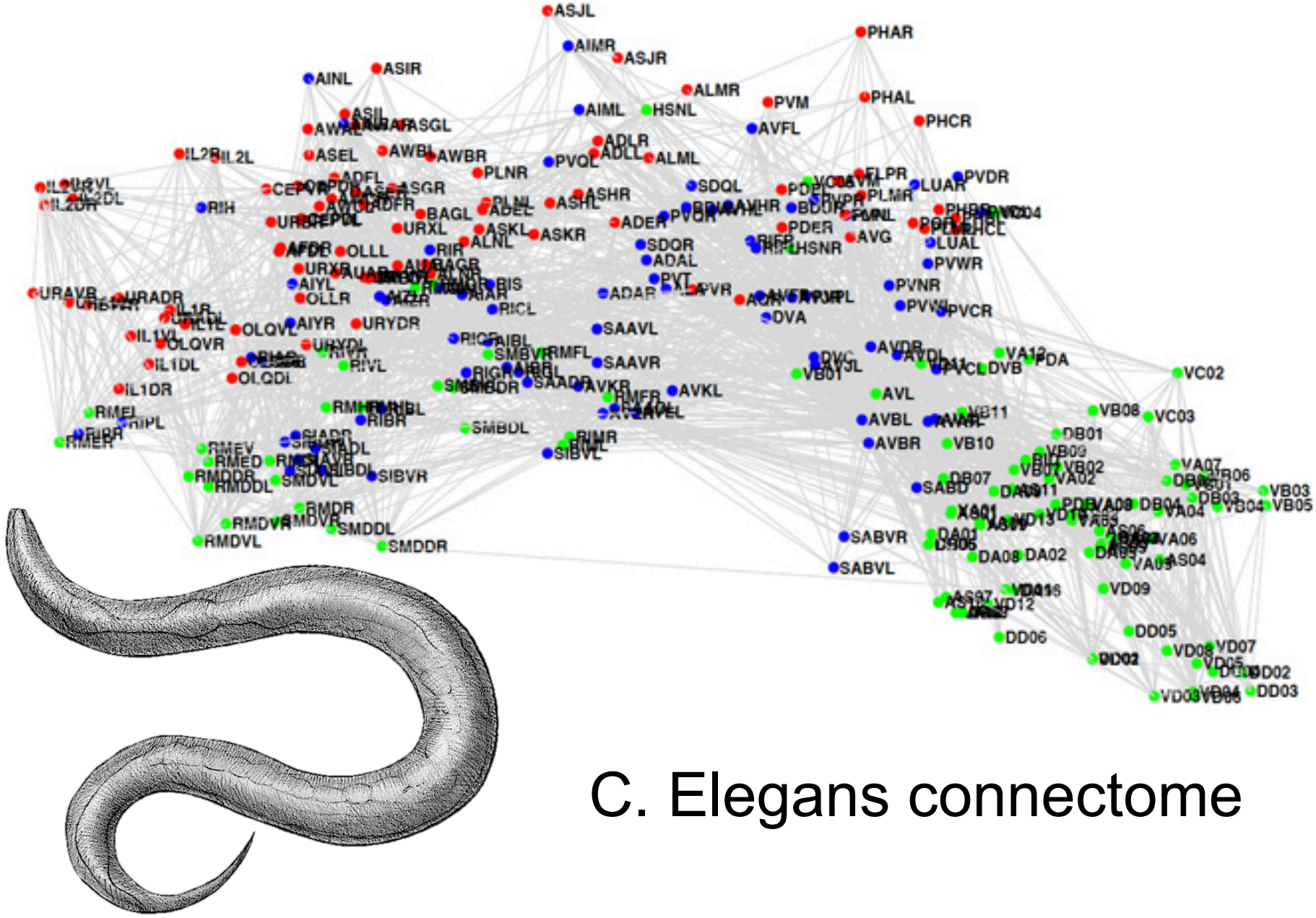
- Take the sum of strengths of all neighbors that are firing
- If sum of inputs > T, fire output

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 compute clusters
- Practical difficulty: How to figure out properties (threshold value, s_i) of each of 10^{10} neurons, the intricate wiring and chemistry

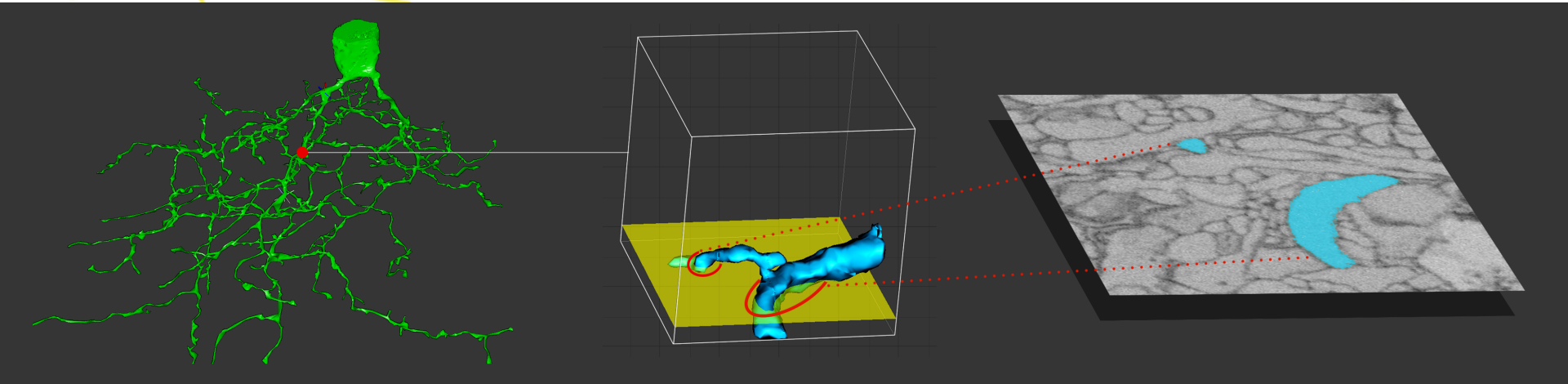
Hope

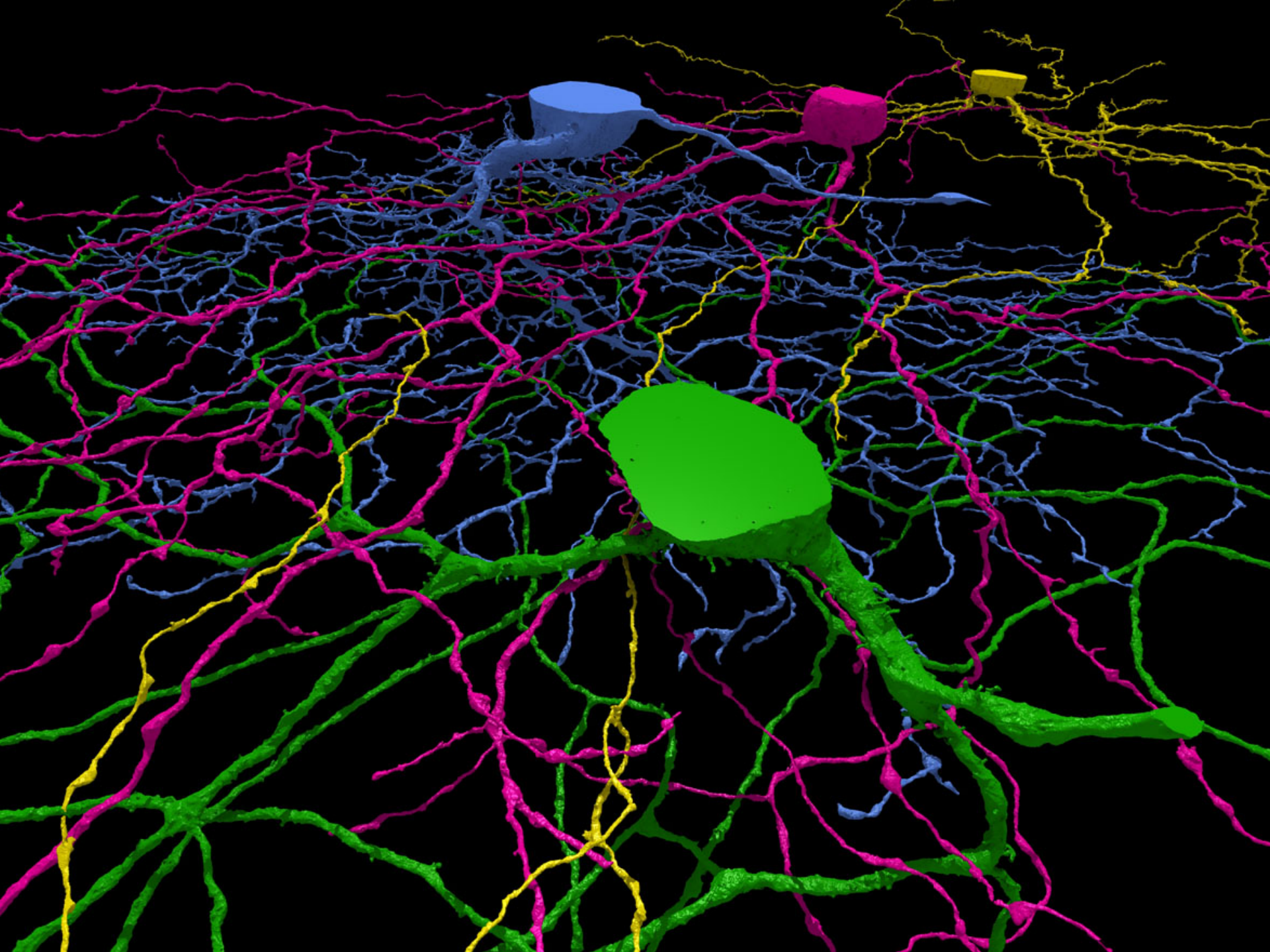
Maybe the brain is organized around simpler principles.



C. Elegans connectome

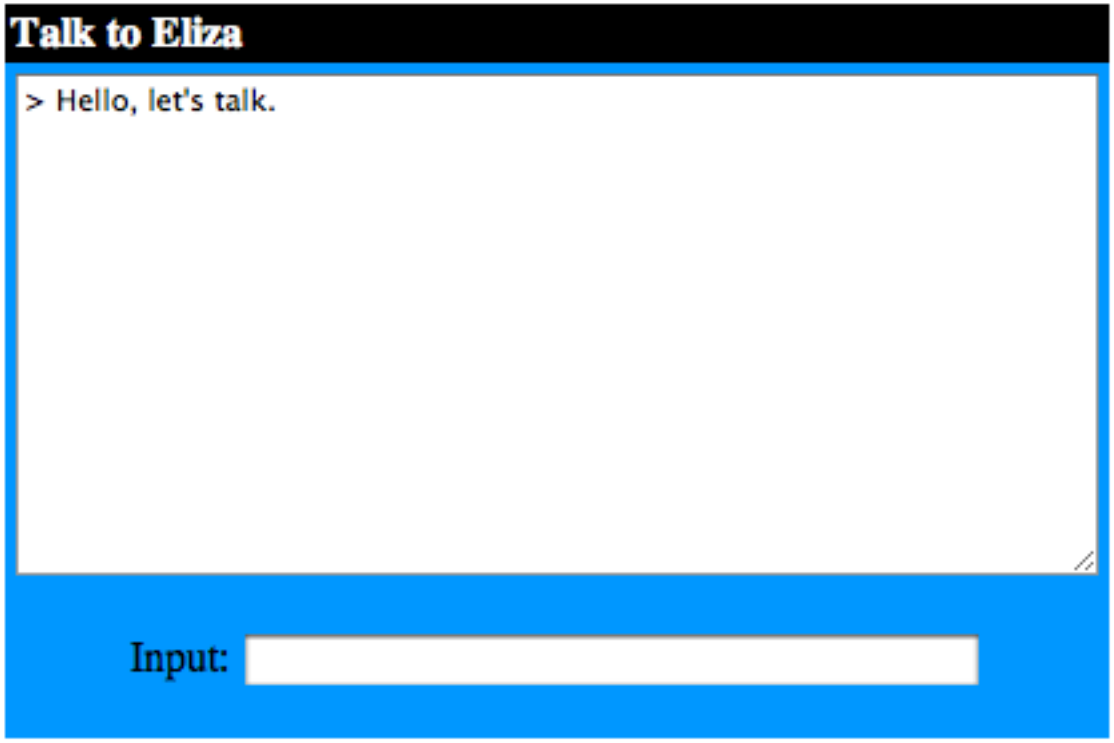
[Sebastian Seung]





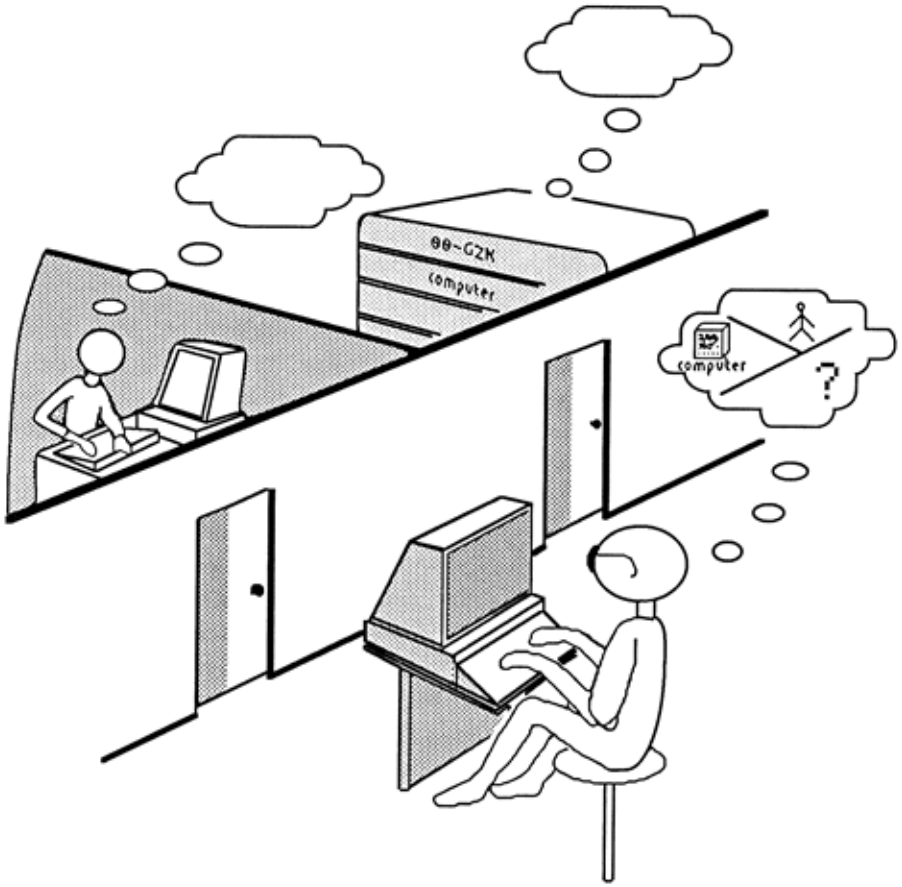
Eliza, the Computer Therapist

http://cyberpsych.org/eliza



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

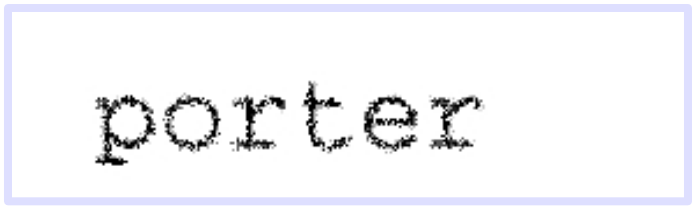
- Not reducible to formula
- No obvious way to cheat
- Customizable to different topics
- Behavioral/ black box.

Weaknesses

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

Text Captchas

EZ-Gimpy [2001]



Google [present]

Beaten by Mori & Mailk [2002]

Beaten by hackers? [NYT 2008]

Image Captchas

Strong AI (Searle)

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,…