Computation as a Foundation for Cognitive Science

Frank Jäkel

TU Darmstadt

Lecture at Peking University, September 2018

Cognition

What is Cognitive Science?

- The science of the mind (and brain)

Why Do We Need Cognitive Science?

- Isn't there already a science of the mind? (Psychology)
- Isn't there already a science of the brain? (Neuroscience)
- How can we study the mind scientifically?
- Shouldn't we study the brain instead?
- What is the mind anyway?

Examples for Mental Phenomena

- Seeing
- Hearing
- Feeling
- Paying attention
- Learning
- Remembering
- Reasoning
- Deciding
- Solving problems
- Understanding language
- ...

What is Cognition?

How we perceive, think, and generally get to know things are questions that are central to cognitive science—hence the name: The word *cognition* is of Latin origin and means *to (get to) know*.

Obstacles to Studying the Mind

Behaviorism

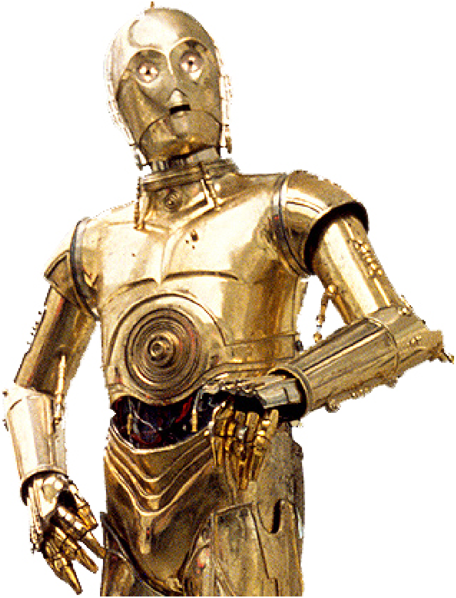
- Minds are only accessible subjectively
- Behavior, in contrast, is objectively observable

"Brainism"

- Physical world is causally closed
- Stimulus $\rightarrow$ Brain $\rightarrow$ Response
- The mind cannot be causally effective and is hence irrelevant (mind-body problem)

$\Longrightarrow$ Study brains and behavior, not minds!

The Robot Challenge



How to Build a Robot?

Hardware

- Body (metal, plastic, etc.)
- Sensors (cameras, microphones, etc.)
- Effectors (motors, pumps, etc.)
- Controllers (wires, electronics, RAM, processors, etc.)

Software

- Logic of control (independent of the hardware)
- Representation of the environment
- Representation of goals
- Internal states and memory
- Planning and reasoning capabilities
- Artificial intelligence (!)

The Computer Metaphor

$$\begin{array}{rcl} \textit{mind} & \longleftrightarrow & \textit{body} \\ \textit{software} & \longleftrightarrow & \textit{hardware} \end{array}$$

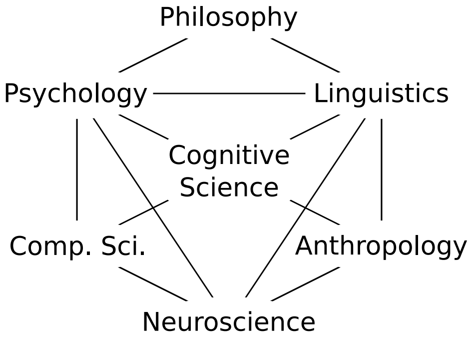
The computer metaphor makes the mind appear less mystical.
With the rise of computers there was a serious scientific alternative to behaviorism and brainism.
"[...] the mind is not the brain but what the brain does [...]" [Pinker, 1997, p. 24]

What is Cognition?[1]

- cognition = representation + computation
- cognitive processing is information processing

[1] the classical view

The Sloan Report (1978)



[Miller, 2003, Gardner, 1987]

Cognitive Science

- Psychology, computer science, neuroscience, anthropology, linguistics, and philosophy study the mind
- There is considerable overlap of interests and (partly) methodology
- Boundaries between disciplines are historical and methodological
- Each discipline also has aspects that are peripheral or irrelevant to basic cognitive science
- Cognitive science is topic-centered
- Cognitive science uses methods from all relevant disciplines
- Cognitive science is often described as being interdisciplinary
- but should be considered transdisciplinary

Landmark works

- Computing Machinery and Intelligence. Turing (1950).
- Magical number 7 ± 2. Miller (1956).
- A Study of Thinking. Bruner, Goodnow, & Austin (1956).
- Syntactic Structures. Chomsky (1957).
- Cognitive Psychology. Neisser (1967).
- Human Problem Solving. Newell & Simon (1972).
- Vision. Marr (1982).
- Parallel Distributed Processing. Rumelhart & McClelland (1986).

Institutionalization

- 1956, Symposium on Information Theory (Miller, Newell, Simon, Chomsky)
- 1956, Meeting at Dartmouth College (McCarthy, Minsky, Newell, Simon)
- 1960, Center for Cognitive Studies at Harvard (Miller, Bruner)
- 1976, Initiative of the Sloan Foundation (Miller)
- 1977, Cognitive Science journal (Schank, Charniak, Collins)
- 1979, Cognitive Science Society (Collins, Schank, Norman)

Summary

- Cognitive science is the study of the mind (but also the brain)
- Cognitive science takes the mind seriously (reaction against behaviorism and brainism)
- Cognitive science seeks a computational-representational understanding of the mind[2] (rise of computer science, cybernetics, and information theory)
- Cognitive science is transdisciplinary (realization that many fields study the mind with different methods)

[2] at least traditionally

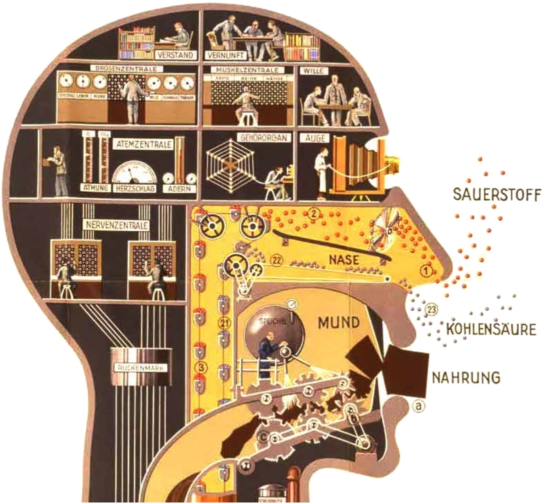
Recommended Reading

A fun book to start with is *How the Mind Works* [Pinker, 1997]. The first two chapters give you a good feel for what cognitive science is about and are highly recommended. Another pop science book that I enjoyed is *The Universe Within* [Hunt, 1982]. It is a bit dated now but conveys a lot of the excitement of early research in cognitive science. You can probably order a very cheap used copy online. The same is true for *The Mind's New Science* [Gardner, 1987]. I find this book a little dry compared to the other two but it will be very useful if you're interested in the history of cognitive science.

For foundational questions in cognitive science I would recommend a book that places a strong emphasis on computation. A book that is a little old but still good is [Johnson-Laird, 1988]. A more recent book is [Edelman, 2008]. This book is very engaging but has to be taken with a grain of salt. Both books have useful introductory chapters. The remaining chapters are roughly also what I want to cover in class. [Thagard, 2005] has a good introductory chapter and surveys cognitive science by theoretical approaches.

Computers

Der Mensch als Industriepalast



Aus Fritz Kahn, Das Leben des Menschen (1926), Franck'sche Verlagshandlung, Stuttgart [NIH]

The Mechanization of Thought

- What exactly is happening between stimulus and response?
- Can we describe these processes—including thinking—in a mechanistic way?
- Can we build (mechanical, electronic, cybernetic, bionic) machines that think?

The Difference between Men and Brutes

"And here I specially stayed to show that, were there such machines exactly resembling organs and outward form an ape or any other irrational animal, we could have no means of knowing that they were in any respect of a different nature from these animals; but if there were machines bearing the image of our bodies, and capable of imitating our actions as far as it is morally possible, there would still remain two most certain tests whereby to know that they were not therefore really men. [...] "

[Descartes, 1637][Gutenberg]

Descartes' First Test

"[...] Of these the first is that they could never use words or other signs arranged in such a manner as is competent to us in order to declare our thoughts to others: for we may easily conceive a machine to be so constructed that it emits vocables, and even that it emits some correspondent to the action upon it of external objects which cause a change in its organs; for example, if touched in a particular place it may demand what we wish to say to it; if in another it may cry out that it is hurt, and such like; but not that it should arrange them variously so as appositely to reply to what is said in its presence, as men of the lowest grade of intellect can do. [...]"

[Descartes, 1637][Gutenberg]

Descartes' Second Test

" [...] The second test is, that although such machines might execute many things with equal or perhaps greater perfection than any of us, they would, without doubt, fail in certain others from which it could be discovered that they did not act from knowledge, but solely from the disposition of their organs: for while reason is an universal instrument that is alike available on every occasion, these organs, on the contrary, need a particular arrangement for each particular action; whence it must be morally impossible that there should exist in any machine a diversity of organs sufficient to enable it to act in all the occurrences of life, in the way in which our reason enables us to act. Again, by means of these two tests we may likewise know the difference between men and brutes."

[Descartes, 1637][Gutenberg]

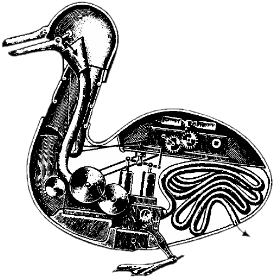
The Difference between Men and Brutes

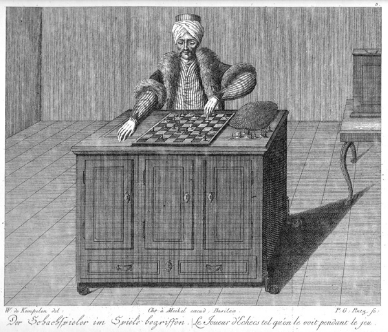
Animals are merely machines. Men have language and reason (that cannot be implemented in machines).

Cartesian Dualism

mind	$\longleftrightarrow$	*body*
res cogitans	$\longleftrightarrow$	*res extensa*

Automata





- Duck of Jacques de Vaucanson (18th century) [Wikipedia]
- Der Schachtürke of Wolfgang von Kempelen (18th century hoax) [Wikipedia]

Reasoning as Calculation

- If we perceive the body as a machine we will be able to describe the mechanisms behind arms, lungs, hearts, etc.
- Can we also describe the (non-material?) "mechanisms" behind thinking?
- Can we find the rules behind reasoning?
- Leibniz (born 1646) dreamt of such a project:
- "it is plain that men make use in reasoning of several axioms which are not yet quite certain"
- "this language will be the greatest instrument of reason"
- "when there are disputes among persons, we can simply say: Let us calculate, without further ado, and see who is right"

[Kulstad and Carlin, 2007][SEP]

The Laws of Thought (aka Logic)

- "The design of the following treatise is to investigate the fundamental laws of those operations of the mind by which reasoning is performed; to give expression to them in the symbolical language of a Calculus, and upon this foundation to establish the science of Logic and construct its method;" [Boole, 1854]
- Boole discovered the analogy between logic and algebra
- *false* $\iff$ 0 and *true* $\iff$ 1
- Then *and* becomes multiplication and *or* becomes addition (almost)
- Reasoning literally becomes calculation

Boolean Logic

$\times$	0	1
0	0	0
1	0	1

$+$	0	1
0	0	1
1	1	1

$-$	0	1
0	0	1
1	0	0

$$A \wedge B \iff a \times b$$
$$A \vee B \iff a + b$$
$$\neg A \iff 1 - a$$

$$A \wedge (B \vee C) \equiv (A \wedge B) \vee (A \wedge C) \iff a \times (b + c) = a \times b + a \times c$$

[Boole, 1854, Dewdney, 1989, 3 Systems of Logic]

Artificial Intelligence

- Boolean Logic is essential for the design of the circuits in a modern computer
- When the first computers became available scientists were quick to try and make the machines do intelligent things
- John McCarthy coined the term *AI* for the Dartmouth conference (1956)
- A typical example for early AI is the Logic Theorist that could prove theorems [Newell and Simon, 1956]

Artificial Intelligence Today

- In 1997 Deep Blue beat the chess world-champion Kasparov
- Darpa challenge: Build an autonomous car to drive around a parcours in the desert
- In 2004 no car finished the parcours. In 2005 the Stanford Racing Team won:

[Wikipedia]

Mechanization of Thought?

- Computers are machines that *appear* to think (attribution)
- Before, only humans could calculate, prove theorems, and play chess—activities that seem to require thinking
- Although, computers don't necessarily do it the same way that we do it
- Calculation, e.g., they can do a lot better and faster than we can but they are also a lot less flexible (Descartes' second test)
- Observation: What's hard for us is easy for them, what's hard for them is easy for us

The Computer Metaphor

$$\begin{array}{rcl} \textit{mind} & \longleftrightarrow & \textit{body} \\ \textit{software} & \longleftrightarrow & \textit{hardware} \end{array}$$

What is Computation?

- Prototypical example: Calculation
- Store: A piece of paper
- Executive unit: Reading and writing symbols, moving pen
- Control: Set of instructions that tells you when to write what where
- Purely “mechanical” or “automatic” (no magic allowed!)

A Concrete Proposal

$$\begin{array}{ccc} \textit{mind} & \longleftrightarrow & \textit{body} \\ \textit{logic} & \longleftrightarrow & \textit{neurons} \end{array}$$

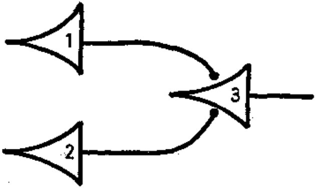
A Logical Calculus of Nervous Activity

- Neurons are all-or-none (0 or 1), they fire at time t or not
- Time is discrete, all axons have the same delay, synapses are excitatory or inhibitory, neurons have a threshold 1.5
- Neurons represent logical propositions and Boolean algebra is done by networks of neurons

$$a_2(t) = a_1(t-1)$$

[McCulloch and Pitts, 1943, Piccinini, 2004, Dewdney, 1989, 12 Boolean Logic]

A Logical Calculus of Nervous Activity

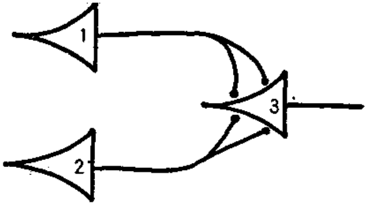


$$a_3(t) = a_1(t-1) \times a_2(t-1)$$

×	0	1
0	0	0
1	0	1

[McCulloch and Pitts, 1943, Piccinini, 2004, Dewdney, 1989, 12 Boolean Logic]

A Logical Calculus of Nervous Activity

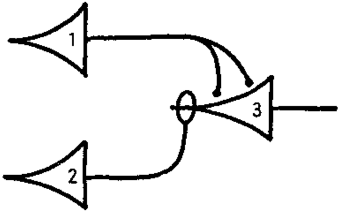


$$a_3(t) = a_1(t-1) + a_2(t-1)$$

+	0	1
0	0	1
1	1	1

[McCulloch and Pitts, 1943, Piccinini, 2004, Dewdney, 1989, 12 Boolean Logic]

A Logical Calculus of Nervous Activity



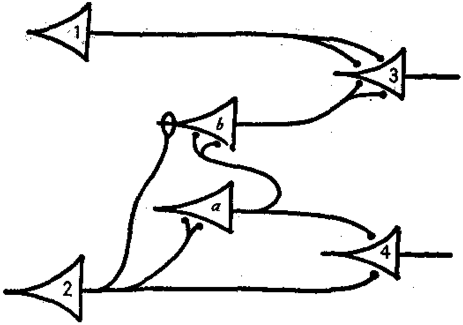
$$a_3(t) = a_1(t-1) - a_2(t-1)$$

$-$	0	1
0	0	1
1	0	0

[McCulloch and Pitts, 1943, Piccinini, 2004, Dewdney, 1989, 12 Boolean Logic]

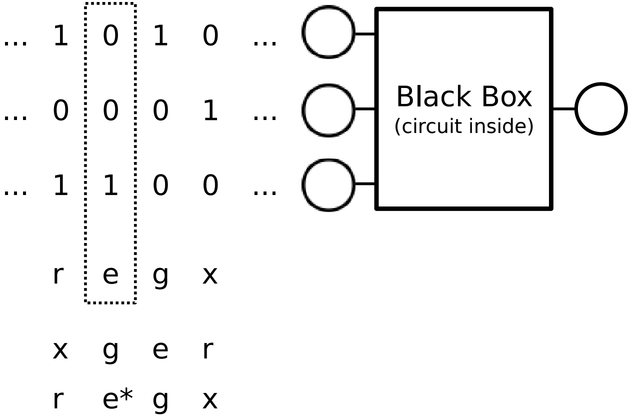
A Logical Calculus of Nervous Activity

- Simple circuits can be combined to more complicated circuits
- What does the circuit below compute?



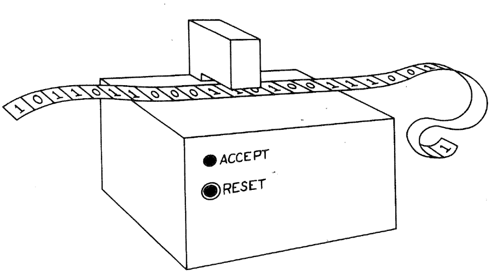
[McCulloch and Pitts, 1943, Piccinini, 2004, Dewdney, 1989, 12 Boolean Logic]

Nerve Nets and Finite Automata



[Kleene, 1951, Dewdney, 1989, 2 Finite Automata]

Formal Languages



- A machine gets a string as input and accepts it or not
- Those that are accepted are grammatical in the language
- There is a correspondence between classes of languages and classes of machines
- How can we describe natural language formally? What is the grammar of English?

[Dewdney, 1989, 7 The Chomsky Hierarchy]

Hierarchy of Computation

- Some machines are more powerful than others
- Some languages are more expressive than others

Increasing generality ⟶

Computing Model	Finite automata	Pushdown automata	Linear bounded automata	Turing machines
Language Class	Regular languages	Context-free languages	Context-sensitive languages	Recursively enumer-able

[Dewdney, 1989, 7 The Chomsky Hierarchy]

Formal Languages

- You know at least one formal language from your logic classes:
- $\neg((a \wedge b) \vee a)$ is a sentence
- $)ab\wedge)$ is not

$$S \rightarrow (S \wedge S)$$
$$S \rightarrow (S \vee S)$$
$$S \rightarrow \neg S$$
$$S \rightarrow a$$
$$S \rightarrow b$$

A Context-Free Language

Context-free: There is only one symbol on the left, i.e. the context does not matter for the application of the rule

$$S \rightarrow aSb$$
$$S \rightarrow ab$$

$$\{ab, aabb, aaabbb, aaaabbbb, \ldots\}$$

$$\{a^n b^n \mid n \geq 1\}$$

A Context-Sensitive Language

Context-sensitive: There can be more symbols on the left, i.e. the context matters for the application of the rule

$$\begin{array}{rcl} S & \rightarrow & aSBC \\ S & \rightarrow & aBC \\ CB & \rightarrow & BC \\ aB & \rightarrow & ab \\ bB & \rightarrow & bb \\ C & \rightarrow & c \end{array}$$

$$\{a^n b^n c^n \mid n \geq 1\}$$

A Context-Sensitive Language

$$
\begin{aligned}
S &\rightarrow aSBC \quad (1)\\
S &\rightarrow aBC \quad (2)\\
CB &\rightarrow BC \quad (3)\\
aB &\rightarrow ab \quad (4)\\
bB &\rightarrow bb \quad (5)\\
C &\rightarrow c \quad (6)
\end{aligned}
$$

$$
\begin{aligned}
& S\\
(1) &\Rightarrow aSBC\\
(2) &\Rightarrow aaBCBC\\
(3) &\Rightarrow aaBBCC\\
(6) &\Rightarrow aaBBcC\\
(4) &\Rightarrow aabBcC\\
(6) &\Rightarrow aabBcc\\
(5) &\Rightarrow aabbcc
\end{aligned}
$$

A Fragment of English

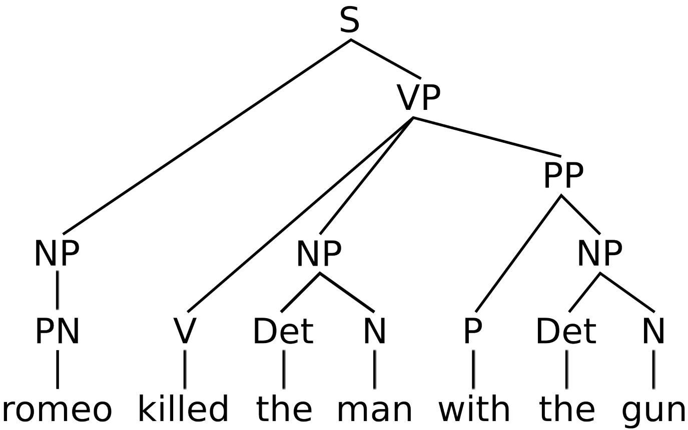
$$\begin{aligned} S &\rightarrow NP\ VP \\ NP &\rightarrow Det\ N \\ NP &\rightarrow PN \\ VP &\rightarrow V\ NP \end{aligned} \qquad \begin{aligned} Det &\rightarrow a, the \\ N &\rightarrow man, woman \\ PN &\rightarrow romeo, juliet, tybalt \\ V &\rightarrow loved, killed \end{aligned}$$

S = Sentence, NP = Noun Phrase, VP = Verb Phrase
Det = Determiner, N = Noun, PN = Proper Noun, V = Verb

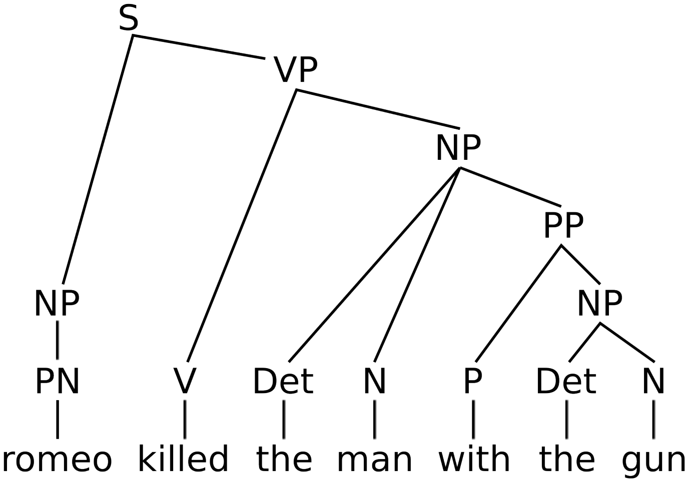
romeo loved juliet, juliet loved tybalt, romeo killed romeo, romeo killed the man, a woman loved a man, ...

Is natural language context-free?

Parsing


S
VP
PP
NP
NP
NP
PN
V
Det
N
P
Det
N
romeo
killed
the
man
with
the
gun

Parsing Ambiguity



Chomsky's Hierarchy, revisited

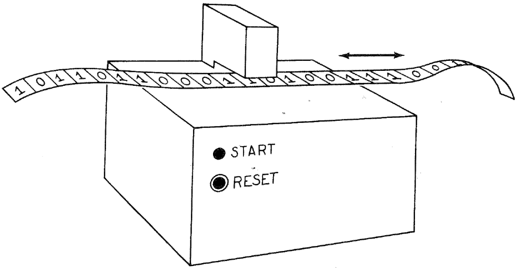
- Some languages are more expressive than others
- There is a correspondence between language and computing
- To recognize more expressive languages you need more general computers

Increasing generality ⟶

Computing Model	Finite automata	Pushdown automata	Linear bounded automata	Turing machines
Language Class	Regular languages	Context-free languages	Context-sensitive languages	Recursively enumer-able

[Dewdney, 1989, 7 The Chomsky Hierarchy]

Turing Machines



- Simplified “calculation” with paper and pencil
- Store: An infinite tape
- Executive unit: Reading and writing zeros and ones, moving tape
- Control: Set of instructions that control when to read, write, and move

[Dewdney, 1989, 28 Turing Machines]

Church-Turing Thesis

- TMs are very general, e.g. every McCulloch-Pitts net can be emulated by a TM (but not vice versa)
- Every other mechanism that has ever been suggested as a computing mechanism can be emulated by a TM
- Thesis: The TM captures what we intuitively mean by computing and every other mechanism that will be suggested to capture this intuition will turn out to be at most computationally equivalent to the TM
- Hence, computable means computable by a TM

[Dewdney, 1989, 60 Church's Thesis]

The Computer Metaphor, Revisited

- Church-Turing Thesis: Everything that can be computed can be computed with a Turing Machine
- The Computer Metaphor: The brain is a computer (implemented in a neural network)
- Whatever the brain computes can be computed on another architecture, e.g. a Turing Machine
- There is an abstract level of description of what the brain does that is independent of the brain
- Of course, we also want to understand how the brain computes
- If you accept Church-Turing it will be the ultimate metaphor (the switchboard metaphor went away, the computer metaphor won't)
- The limits of computation are also the limits of thought!

Can Computers Think?

If we take the computer metaphor seriously it will be possible to build machines that think. How could we test whether a computer is capable of thinking?

The Turing Test

- You sit in front of a terminal
- Over the terminal you either chat with another person in another room
- Or you talk to a computer program
- Can you tell the difference?
- This question replaces the original question whether machines can think
- Similar to Descartes' tests but making it easier for the computer (no body needed)

[Turing, 1950]

ELIZA et al.

Men are all alike.
IN WHAT WAY
They're always bugging us about something or other.
CAN YOU THINK OF A SPECIFIC EXAMPLE
Well, my boyfriend made me come here.
YOUR BOYFRIEND MADE YOU COME HERE
He says I'm depressed much of the time.
I AM SORRY TO HEAR YOU ARE DEPRESSED
It's true. I'm unhappy.
DO YOU THINK COMING HERE WILL MAKE YOU NOT TO BE UNHAPPY
...

[Weizenbaum, 1966, p. 36-37]

ELIZA et al.

- ELIZA emulates a psychotherapist (Rogerian)
- ELIZA works best, i.e. a subject attributes some intelligence to the program, if the subject is instructed to assume she is talking to a therapist
- This scenario was chosen because the therapist "is free to assume the pose of knowing almost nothing of the real world"
- Weizenbaum was concerned that people could be fooled by AI: "But once a particular program is unmasked, once its inner workings are explained in language sufficiently plain to induce understanding, its magic crumbles away; [...] The object of this paper is to cause just such a reevaluation of the program about to be 'explained'. Few programs ever needed it more."

[Weizenbaum, 1966, p. 42, p. 36]

IBM's Watson

- Feb 14-16 2011 IBM's Watson participated in Jeopardy
- The computer beat two Jeopardy champions
- Watson uses a massive knowledge database (not thinkable before the internet and wikipedia)
- US Cities. What is Toronto?????

Weak vs Strong AI

- Weak AI: Computer programs are simulations of thoughts. They are useful tools to specify hypotheses about mental mechanisms (what's going on between stimulus and response). But nothing more.
- Strong AI: A simulated thought is a real thought.

The Chinese Room

- I sit in a room with a set of (English) instructions on how to manipulate Chinese characters (that I do not understand)
- Through a hatch I receive a stack of Chinese characters that I manipulate according to the instructions
- The instructions lead me to compile a stack of characters to give back through the hatch
- A native Chinese speaker is talking to me through the stacks of characters
- My instructions are so good that a Chinese speaker thinks he is speaking to a real person in Chinese (i.e., I am passing the Turing Test)
- I still don't understand a word!
- Isn't this the situation a computer program is in?

[Searle, 1980]

Summary

- Logic is (partly) an attempt to discover the laws of thought
- Boolean logic links logic with calculation
- McCulloch-Pitts nets (inspired by neural nets) can compute logical functions
- This idea is an early illustration for the computer metaphor of thought
- Computation can be carried out with various machines that have different architectures
- The Turing machine is thought to capture what it means to compute
- What does it mean to say that a computer can think?

Recommended Reading

A really nice book is the *Turing Omnibus* [Dewdney, 1989]. It has very short and readable chapters on fundamental and interesting topics in computer science. I've referenced the relevant chapters on the respective slides. A (rather long) pop-science book that has been very influential for many budding cognitive scientists of previous generations is *Gödel, Escher, Bach* [Hofstadter, 1977]. The author presents the big topics in logic, computation, and artificial intelligence in an eccentric but inspiring way. The standard textbook on AI is [Russell and Norvig, 1995]. Turing's paper is a classic [Turing, 1950].

Boole, G. (1854).
An Investigation Of The Laws of Thought.
Number 15114. Projekt Gutenberg.

Descartes, R. (1637).
Discourse on the Method of Rightly Conducting One's Reason and of Seeking Truth in the Sciences.
Number 59. Projekt Gutenberg.

Dewdney, A. K. (1989).
The Turing Omnibus.
Computer Science Press, Rockville, MD.

Edelman, S. (2008).
Computing the Mind.
Oxford University Press, Oxford.

Gardner, H. E. (1987).
The Mind's New Science: A History of the Cognitive Revolution.
Basic Books, New York.

Hofstadter, D. (1977).
Gödel, Escher, Bach.
Basic Books, New York.

Hunt, M. (1982).
The Universe Within: A New Science Explores the Human Mind.
Simon & Schuster, New York.

Johnson-Laird, P. N. (1988).
The Computer and the Mind. An introduction to Cognitive Science.
Harvard University Press, Cambridge, MA.

Kleene, S. C. (1951).
Representation of events in nerve nets and finite automata.
Technical Report RM-704, The RAND Corporation, Santa Monica, CA.

Kulstad, M. and Carlin, L. (2007).
Leibniz's philosophy of mind.
In Zalta, E. N., Nodelman, U., Allen, C., and Perry, J., editors, *Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University.

McCulloch, W. S. and Pitts, W. (1943).
A logical calculus of the ideas immanent in nervous activity.
Mathematical Biophysics, 5:115–133.

Miller, G. A. (2003).
The cognitive revolution: a historical perspective.
Trends in Cognitive Sciences, 7(3):141–144.

Newell, A. and Simon, H. A. (1956).
The logic theory machine: A complex information processing system.
Technical Report P-868, The RAND Corporation, Santa Monica, CA.

Piccinini, G. (2004).
The first computational theory of mind and brain: A close look at McCulloch and Pitts's "logical calculus of ideas immanent in nervous activity".
Synthese, 141:175–215.

Pinker, S. (1997).
How the Mind Works.
Penguin Books, London.

Russell, S. and Norvig, P. (1995).
Artificial Intelligence: A Modern Approach.
Prentice Hall, Upper Saddle River, NJ.

Searle, J. R. (1980).
Minds, brains, and programs.
Behavioral and Brain Sciences, 3(3):417–457.

Thagard, P. (2005).
Mind: Introduction to Cognitive Science.
MIT Press, Cambridge, MA, 2nd edition.

Turing, A. M. (1950).
Computing machinery and intelligence.
Mind, 59:433–460.

Weizenbaum, J. (1966).
ELIZA—A computer program for the study of natural language communication between man and machine.
Communications of the ACM, 9(1):36–45.