Philosophy of AI (WBMV05003)

Janneke van Lith


Universiteit Utrecht

2019-2020 period 3

February 10, 2020

Today's topics

Challenges to Strong AI

- Heideggerian arguments
- The China Brain Argument
- The Chinese Room Argument

Architectures

- GOFAI
- Connectionism
- SED

Representations

- Representation as central to cognition
- Questions about representation

Challenges to Strong AI

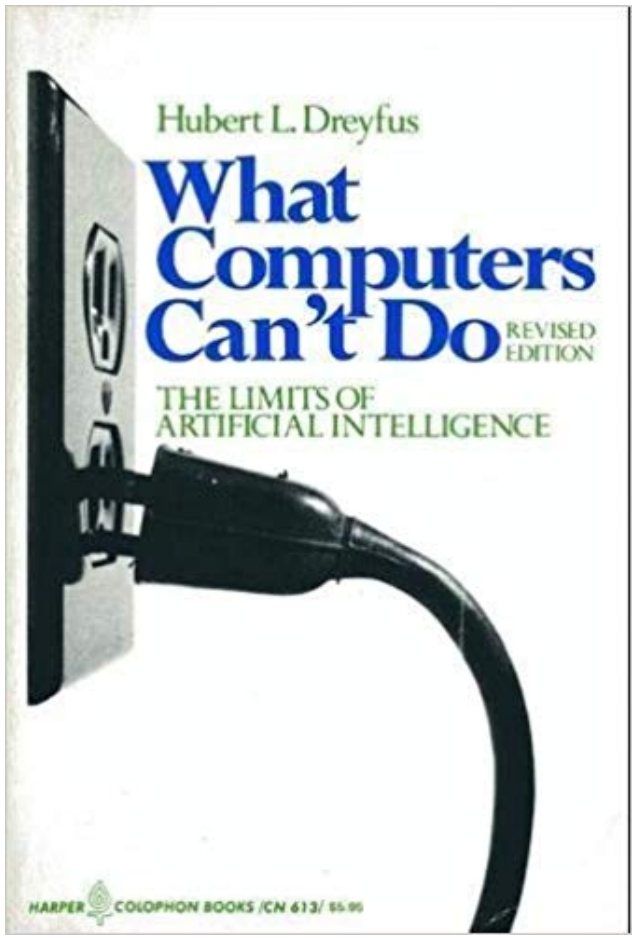
- Heideggerian arguments (Dreyfus; see Ch. 2, 3 and 4)
- The China Brain Argument (Block; see Ch. 2)
- The Chinese Room Argument (Searle; see Ch. 2, 3 and 4)
- Gödelian arguments (Lucas, Penrose; see Ch. 3)

Note: some of these arguments are aimed specifically at GOFAI.

Dreyfus' arguments against AI

Inspired by continental philosophy (Heidegger, Merleau-Ponty)

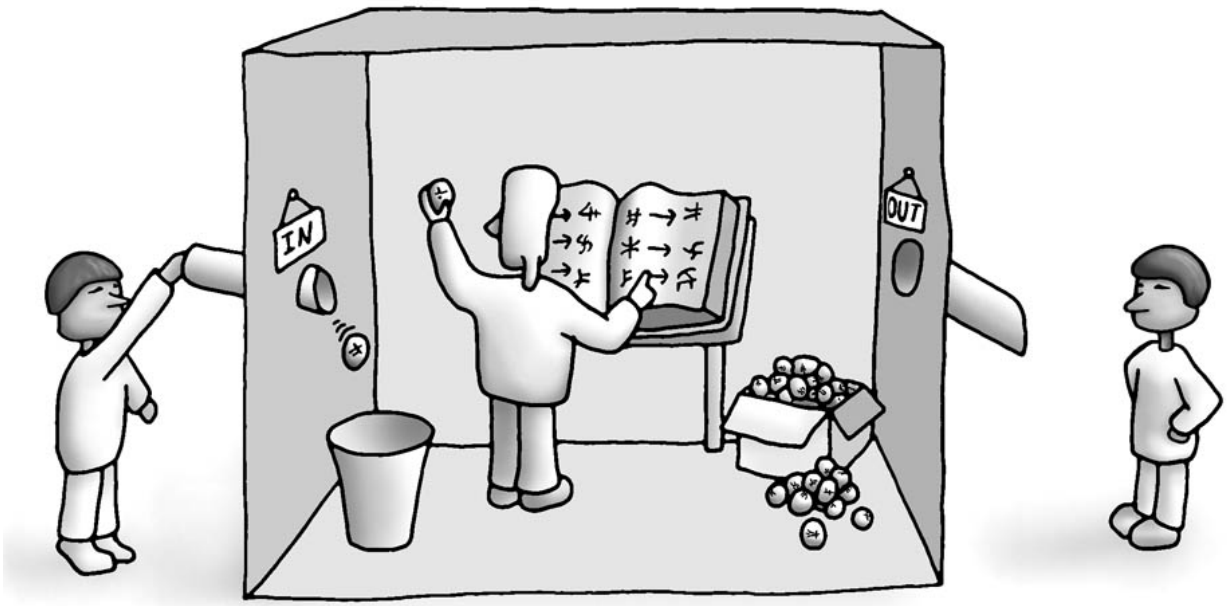
- rules
- relevance
- know-how
- situatedness



The China Brain Argument

- Thought experiment by Ned Block (1978)
- Imagine the people of China to simulate the workings of the neurons in a brain, communicating by radio. Would such a China Brain be conscious / have mental content? Intuitively: no!
- Aimed specifically at machine functionalism, which (according to Block) would have to say yes.

The Chinese Room Argument



Searle's argument

From the abstract of John Searle's 1980 paper:

- Instantiating a computer program is never by itself a sufficient condition of intentionality
- Strong AI has little to tell us about thinking, since it is not about machines but about programs, and no program by itself is sufficient for thinking

Searle's argument

- Against the Physical Symbol Systems Hypothesis: the GOFAI recipe will not deliver a thinking machine.
- Against the possibility of Strong AI
- Against the Turing Test

The systems reply

Systems reply:

- Not only the symbol manipulator inside the room is concerned, but the system as a whole (including, for example, the rulebook).

Searle's response:

- The man in the room doesn't understand, but if we add paper and pencil he would?!? (cynically) .
- Let the man learn all rules of the program by heart; still he wouldn't understand Chinese.

Other arguments against the CRA

- Robot reply: indeed, machines that only crunch symbols lack semantics / mentality / understanding. But causally connecting the machine to the world solves it all. Robots, but also computers, may have the appropriate causal connections.
- Brain simulator reply: Searle will have to conclude that a machine simulating brain activity also doesn't yield semantics / mentality / understanding.
- Other minds reply: When attributing mentality to other people, we only have their outward behaviour to go on.
- Intuition reply: Searle appeals to intuitions, which may simply be false.

Wider issues raised by the CRA

So, perhaps the CRA is not a compelling argument. Still, it raises fundamental issues:

- Is strong AI possible?
- Is the Symbol System Hypothesis (SSH) true?
- Is syntax sufficient for semantics?
- How do machines acquire semantics?
- Can computational accounts of thinking explain consciousness?
- Is simulating understanding sufficient for real understanding?

Good Old Fashioned AI

GOFAI methodology involves programmed instructions operating on formal symbolic representations (p. 89).

GOFAI is also loosely related to:

- Computational Theory of Mind (Ch. 2, pp. 41–46)
- Physical Symbol System Hypothesis: 'A physical symbol system has the necessary and sufficient means for general intelligent action', Allen Newell and Herbert Simon (1976)
- Machine functionalism

The Computational Theory of Mind

- Mental activity involves Turing-style computation over a language of thought.
- Mental states such as beliefs and desires are attitudes towards a proposition; believing e.g. that 7 is a prime number involves having a symbolic structure with the meaning that 7 is prime in your "belief box".
- Computational processes are algorithms that act on the symbols (syntax) only.

GOFAI recipe for an intelligent system

1. Use a sufficiently expressive, inductively defined, compositional language to represent real-world objects, events, actions etc.
2. Construct an adequate representation of the world and the processes in it in a universal symbol system (USS).
3. Use suitable input devices to obtain symbolic representation of environmental stimuli.
4. Employ complex sequences of the fundamental operations of the USS to be applied to the symbol structures of the inputs and the knowledge base, yielding new symbol structures (some of these are designated as output)
5. This output is a symbolic representation of response to the input. A suitable robot body can be used to translate the symbols into real behaviour / action.

The Symbol System Hypothesis

The SSH says:

This recipe is correct!

- SSH: computers (i.e. universal symbol manipulators) can think (i.e. are massively adaptable).
- SSSH: Only computers can think. Ergo, the human mind is a computer (i.e. universal symbol manipulator)!

GOFAI: criticism

John Haugeland coined the term 'GOFAI'. His reasons for rejecting GOFAI (see Boden, Ch. 4):

- GOFAI only yields "empty programs" (intentionality; CRA)
- GOFAI lacks embodiment

Note: these points of criticism are perhaps better aimed at Strong AI and the PSSH. GOFAI needn't be committed to this (see Boden, Ch. 4, p. 97).

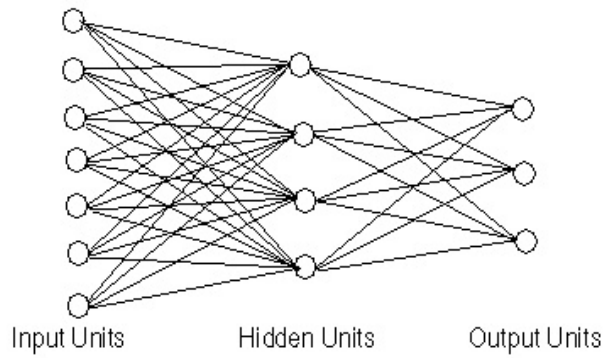
The myth of GOFAI failure

According to Margaret Boden (Ch. 4), it is false that GOFAI has failed:

- It has many successful applications (expert systems, search engines, planners, video games, . . .) – which however sometimes are invisible, or unrecognized as AI
- Computational psychology is partly successful: it helps explain topics such as reasoning, problem solving and language use
- Abstract symbol systems will not deliver Strong AI – but GOFAI might be part of truly intelligent systems

Comparison: connectionism vs. GOFAI

- Pattern recognition
- Sensitivity to context
- Learning
- Graceful degradation
- Distributed representation
- Systematicity and productivity



SED

Cognition is . . .

- Situated. An agent's environment plays a key role in its behaviour.
- Embodied. The physical aspects of an agent's body play a key role in its behaviour.
- Dynamical. Cognition takes place in real time, and cognitive processes can be modelled and explained by Dynamical Systems Theory.

Overview

Challenges to Strong AI

- Heideggerian arguments
- The China Brain Argument
- The Chinese Room Argument

Architectures

- GOFAI
- Connectionism
- SED

Representations

- Representation as central to cognition
- Questions about representation

Two slogans

- 'Intentionality is the mark of the mental' (Franz Brentano 1874)

Two slogans

- 'Intentionality is the mark of the mental' (Franz Brentano 1874)

- 'Cognition is computation over representations' (Newell and Simon 1976; Fodor and Pylyshyn 1988)

C.S. Peirce: semiotics

Peirce (1867): representations consist of signs, objects, and interpreters.

Three types of *signs*, that differ in the way in which the sign relates to the object:

- icons: similar to the represented object. Examples: pictures, maps.
- indices: influenced by the represented objects. Examples: clocks, thermometers.
- symbols: abstract entities, refer to represented object, but bear no resemblance to them.

Representations

Many types of representations: words, pictures, diagrams, numerals, data structures, tree rings, political demonstration, music, etc.

Representation consists of:

- the representing world;
- the represented world;
- the process of representing;
- the system that makes use of the representation.

Is a naturalistic account of representations possible?

Conclusions from Searle

There are really two distinct conclusions Searle argues for in his Chinese Room experiment:

- Running a program is inadequate for instantiating understanding or mentality in general. This is a rejection of strong AI: running the right program is not sufficient for having a mind, or for thinking.
- Running a program is inadequate for having real representational content. Syntax alone is not sufficient for semantics, and digital computers have a syntax alone.

Harnad: the symbol grounding problem

Stevan Harnad (1990):

> *'Suppose you had to learn Chinese as a second language and the only source of information you had was a Chinese/Chinese dictionary. The trip through the dictionary would amount to a merry-go-round, . . . never coming to a halt on what anything meant.'*

Even more difficult: learning Chinese as a first language, with only a Chinese/Chinese dictionary to go on!

Two philosophical questions about representation

- content: What are representations about? What determines the content?
- function: How does a physical state actually fulfil the role of representing in a physical or computational process?

Representational content

What determines the content of representations?

- According to causal-informational theories the content of a representation is grounded in the information that it carries about the external state of affairs that causes the representation.
- According to conceptual role semantics or functional role semantics the content of a representation is grounded in the computational or inferential role it plays in the "cognitive ergonomy" of the subject.
- According to evolutionary theories, the content of a representational state is determined by the function for which the state has been selected.

The problem of misrepresentation

Possibility of error: target and content of a representation may come apart.

Can the distinction between targets and contents fruitfully be made in naturalistic theories? This has proved to be difficult. Attempts: teleological account of targets; causal account of contents (Ruth Millikan, Fred Dretske).

Representational function

The challenge (William Ramsey): provide a description that tells us what it is for something to function as a representation in a physical system.

Do tree rings represent the age of the tree? Do gastric juices represent food? Answering yes to many of such questions leads to pan-semanticism: attributing representations to a system then loses its explanatory value.

Is indication sufficient for representation? And similarity?

Representations: a definition

According to John Haugeland (1991), a system has internal representations if the following conditions are met:

- It must coordinate its behaviors with environmental features that are not always reliably present to the system;
- It copes with such cases by having something else (in place of a signal directly received from the environment) stand in and guide behavior in its stead;
- That "something else" is part of a more general representational scheme that allows the standing in to occur systematically and allows for a variety of related representational states.