# Philosophy & AI

## Chapter 23

Xavier Parent

# Definition of AI

- “[The automation of] activities that we associate with human thinking, activities such as decision-making, problem-solving, search, ... » (Bellman, 1978)

# Goals in AI

- Computer Scientist
  - To solve real-world problems. Build systems that exhibit intelligent behavior.
  - To understand what kind of computational mechanisms are needed for modeling intelligent behavior
- Philosopher
  - To understand human mind by duplicating its functionality

# Introduction

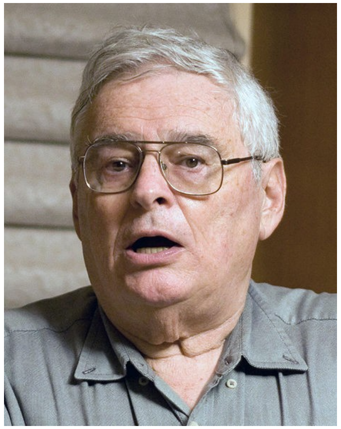
- My goal today
  - Introduce some of the philosophical debates that have been accompanying AI since its inception in the 1950s
- Relationship not necessarily a one-way street
  - AI needs philosophy in order to help clarify goals, methods and concepts (?)

# Outline

- Computational theory of the mind
- Can machines think?
  - Turing test
  - Searle's chinese room argument
- The mind-body problem

# 1
# Computational theory of the mind

# Computational Theory of the Mind (CTM)



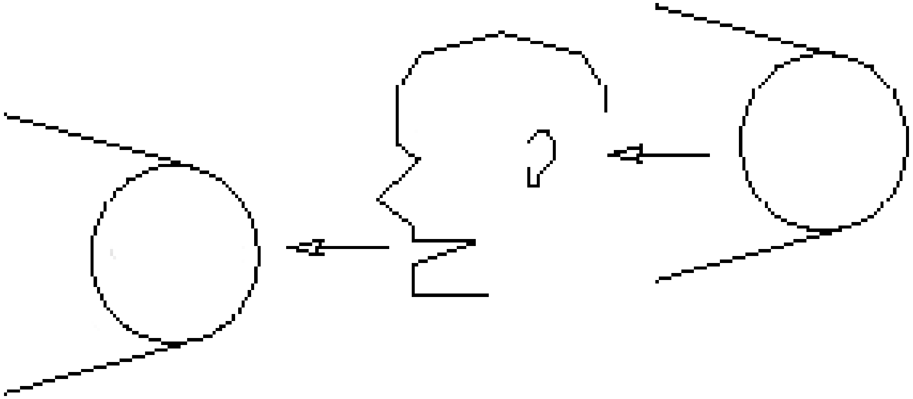
Fodor

Computer

Data structures + algorithms = running programs

Mind

Mental representations + computations = thinking



Propositional attitudes

Language of thought  (LOT) hypothesis:

Thinking done in a mental language (*mentalese*)

# Arguments for CTM

- Language acquisition
- Support from cognitive science
- Thought is productive

- What the theory claims
  - There are certain aspects of the mind that follow step by step processes to compute representations of the world.
    - Search, planning, concept learning,...
- What the theory does not claim
  - All mental states (e.g., anxiety) are suitable for a computational treatment («Qualia »)
  - Computation is sufficient for thought

# 2
# Can machines think?

# Ascribing mental attitudes to machines

- Not unusual in ordinary life
  - « Recently it was too hot upstairs, the plumber came and found the trouble. It reported that the thermostat mistakenly believed it was too cold upstairs »
  - « I'm purchasing a flight ticket. The on-line booking system wants me to provide my credit card details to complete the reservation »
  - « If the server intends to respond with a failure message, it may delay for an implementation dependent time before sending to the client »
- Should this be taken literally?

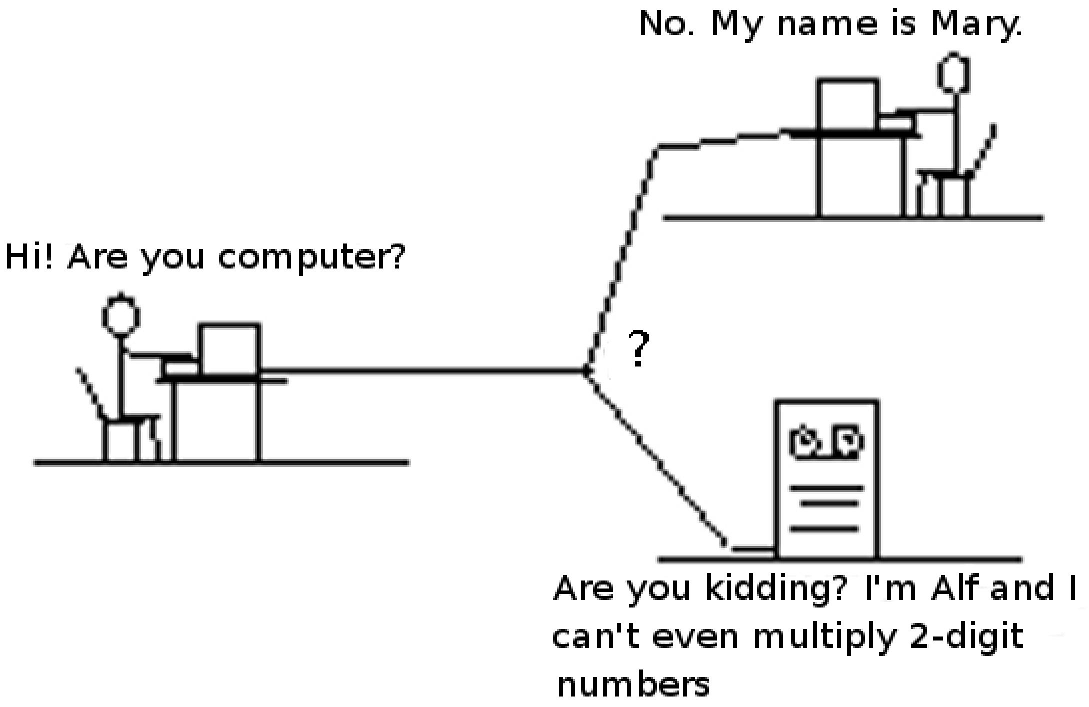
# Animal thoughts

- Similarly for animals
  - « Pet dogs regard their owners as a substitute family » (is it belief?)
  - « Some dogs start barking 'cos they just do not want you to go out »
- Can animals think?
  - No, they don't speak (Descartes, Davidson)
  - Not so sure the argument applies to computer

# Varieties of AI

- Searle 1980s):
  - Strong AI (mainstream till the 80's)
    - An AI machine really *is* capable of thought
  - Weak AI (nowadays mainstream)
    - An AI machine can only *appear* to think

# Turing test (1950)


No. My name is Mary.
Hi! Are you computer?
?
Are you kidding? I'm Alf and I can't even multiply 2-digit numbers

If a computer could pass for human in an on-line chat, it should be counted as intelligent.

# Turing test

- Turing predicted: by 2000, 30% chance for a machine to fool a human for 5 min
- An early success at fooling people: Eliza, computer psychotherapist
- Try it!

http://www.manifestation.com/neurotoys/eliza.php3

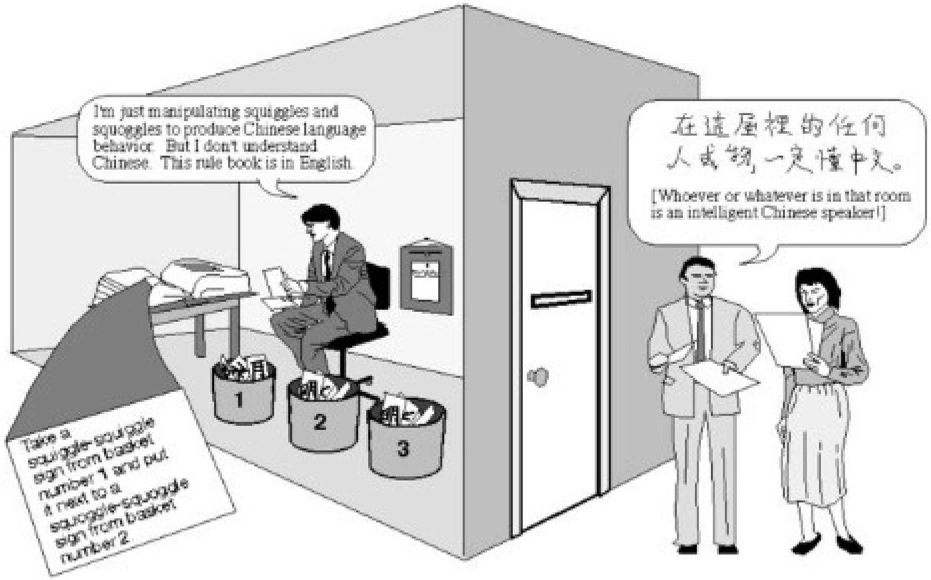
http://www.manifestation.com/neurotoys/eliza.php3

# Objections to Turing's test

- Gunderson
  - Intelligence requires more than passing just the test
- Davidson
  - We have not proved that the machine has grasp of semantics ...
- Searle
  - ... or intentionality

# Searle's chinese room (1980)

- Searle: even if computer could pass the Turing test, it would not exhibit « thinking »
- Chinese room thought experiment:


I'm just manipulating squiggles and squoggles to produce Chinese language behavior. But I don't understand Chinese. This rule book is in English.
在這屋裡的任何人或物，一定懂中文。
[Whoever or whatever is in that room is an intelligent Chinese speaker!]
Take a squiggle-squiggle sign from basket number 1 and put it next to a squoggle-squoggle sign from basket number 2
1
2
3

A man in a room

Does not understand Chinese

Book with instructions for manipulating Chinese symbols

Chinese goes in, he processes instructions, Chinese goes out

# Formal arguments

- Ax 1: Program are formals (syntactic)
- Ax 2: Minds have mental contents (semantics)
- Ax 3: Syntax by itself is neither constitutive of nor sufficient for semantics
  - Chinese room scenario
- Conclusion: Programs are not sufficient for minds

What do you think?

# The robot reply

- What if the whole system was put in a robot?
- The robot would interact with the world
- This would create understanding

# The complexity reply

- Our intuitions about what a complex systems can be are highly unreliable
- Computers at the most basic level are just switches that flip from 0 to 1 can play chess and beat the worlds' best human players
- If you didn't know this could be done, then you would not beleive it
- Maybe symbolic manipulation of sufficient complexity can create semantics, ie. produce understanding

# The problem of other minds

- No clear consensius among philosophers on what « understanding » (Chinese, etc.) involves
- How to justify in the first place our belief that other people (not computers) have minds as well?
  - We only see their behavior, not what's in their head
- We know other people understand Chinese by their behavior
- Why not do the same for computers?

# 3
# Mind-body problem

# Mind-body dichotomy

Problem statement

What is a mind, and what is its relation to body, or to the physical in general?

Dualism (Descartes)

Mind is (ontologically) distinct from body

- Minds are immaterial,and utterly non-spatial

Dogma of the “ghost in the machine” (Ryle)

How can something immaterial interact causally with physical objects in space?

# Identity theory

Classic way out: Make the mental entirely physical

- Mental state identical to (reducible to) brain state
- Causal rôle of mental phenomena derived from their physical substrats

Problem: **Multiple realizability** thesis (Putnam)

- A mental state can correspond to, or is at least correlated with, completely different physical states of the nervous system in different organisms
  - My dog and I experience the same mental state of « being in pain »

# Functionalism

- AI-inspired solution:

(Putnam, Fodor)

$$\frac{\text{mind}}{\text{body}} = \frac{\text{software}}{\text{hardware}}$$

Slogan: « The mind is the software of the brain »

What does this mean? Certainly not that the mind is analogous to software, and the brain to hardware

# Functionalism (con't)

- Functional approach
  - Mental state M is the state that is preconceived by P and causes Q.
    - P and Q= physical + mental states
- Turing machine
  - Each state defined exclusively in terms of its relations to the other states as well as inputs and outputs.

# Functionalism (con't)

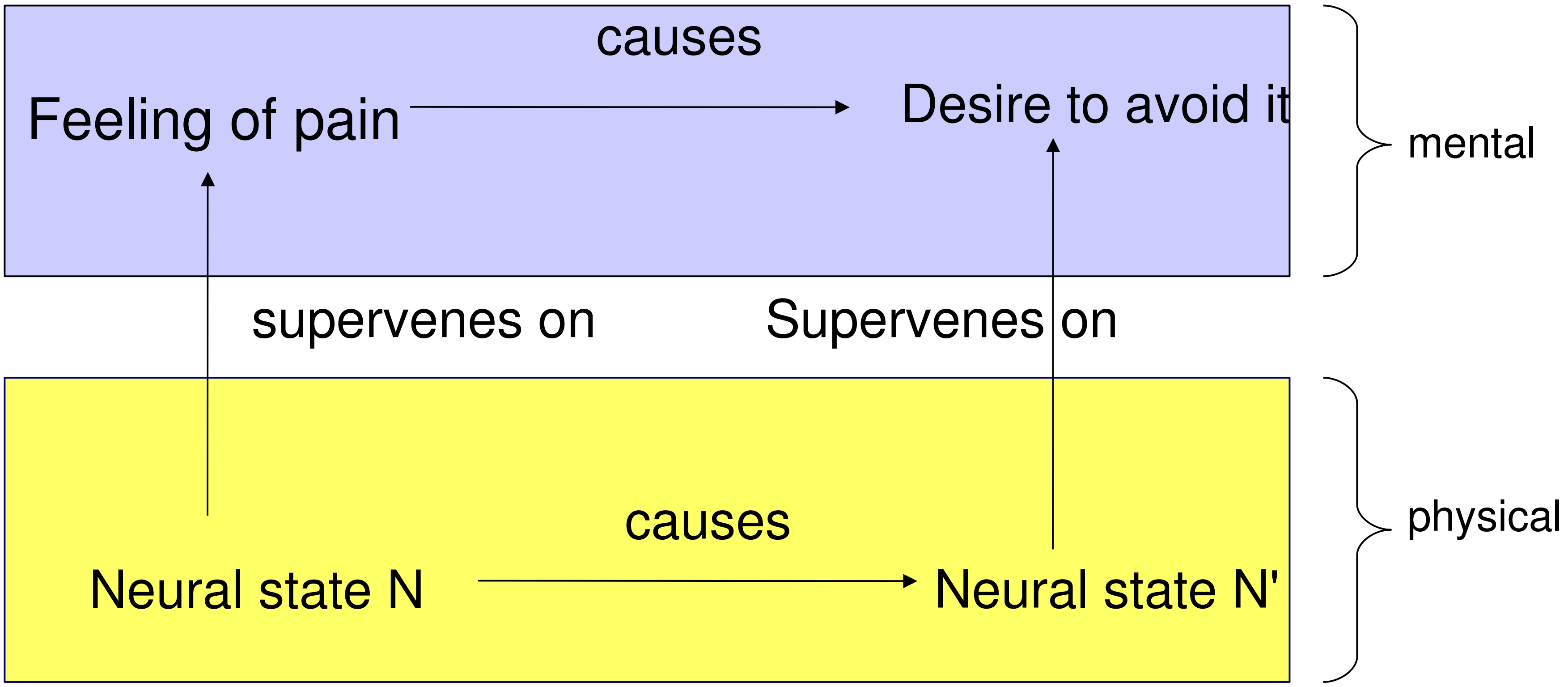
- Software is a collection of programs performing a task; hardware are the physical components of the system
- Is software the same as hardware?
  - Software (non-physical ») is **realized** in hardware (physical), and **multiply realizable**
- Is the mind the same as the brain?
  - Mental processes are **realized** in brains, **multiply realizable**

# Functionalism (con't)

- Against the dualism of substances: Minds are not distinct immaterial substances causally related to bodies
- Talk of minds is merely talk of material systems at a « higher » level of abstraction
  - There is just **one class of events**, which can be **described** in both mental terms (« thinking », « desiring », etc.) or physical terms (a pattern, a neural firing in your brain)

# Functionalism (con't)

Example



causes

Feeling of pain

Desire to avoid it

mental

supervenes on

Supervenes on

causes

Neural state N

Neural state N'

physical

# Functionalism (con't)

- Objection
  - Problems over qualias and consciousness
    - What it feels like to be in a mental state of such-and-such sort?

# Conclusion

- This illustrates how concepts from AI can be used to bring insights into old philosophical problems