The Philosophy of Artificial Intelligence

George Konidaris
gdk@cs.brown.edu

Fall 2018
AI: The Very Idea

For as long as people have made machines, they have wondered whether machines could be made intelligent.

(pictures: Wikipedia)
The Mind-Body Problem
What Are Minds Made Of?

**Dualism**
- Mind and body are two different things
- Mind non-physical
What Are Minds Made Of?

**Materialism**
- Only material objects exist
- Mental states are physical states

“Brains cause minds”
- Because of what they are: *Biological naturalism*
- Because of what they do: *Functionalism*
What is computation?

Information processing process.

• Symbols and rules for operating on symbols.
• The symbols and the rules matter.
• The physical substrate doesn’t matter.
Core AI Hypothesis

“The brain is a computer.”

(picture: Wikipedia)
The Weak AI Hypothesis

Weak AI Hypothesis:

• A computer program can be built that acts as intelligently as a human.

• This is about behavior.
• It makes no claim about minds or even thought.

• Not even about practicality: asks whether it can be done in principle.

**c.f. rational agent.**
Turing


“Can machines think?”

(picture: Wikipedia)
Turing Test
Objections

The Theological Objection

The Head-In-The-Sand Objection

The Mathematical Objection

The Argument from Consciousness

The Argument from Various Disabilities
The Strong AI Hypothesis

Strong AI Hypothesis:

• A machine that behaves as intelligently as a human has a mind and is thinking.

• Takes the Weak AI hypothesis as a given.
• This is a claim about minds and thought.
The Chinese Room

Where’s the mind / understanding / consciousness?

![Diagram](image)
The Brain Replacement Experiment
Embodied Intelligence

- Being There: Putting Brain, Body, and World Together Again
  - Andy Clark

- how the body shapes the way we think
  - A new view of intelligence
  - Ralph Pfeifer and Josh Bengard
  - Foreword by Rodney Brooks
Questions

What do we mean by:

- Mind
- Consciousness
- Thought
- Intelligence

How can we explain or define subjective experience? Is it susceptible to objective analysis or definition?

Can it be reproduced by a computer program?
Highly Recommended (but dated)