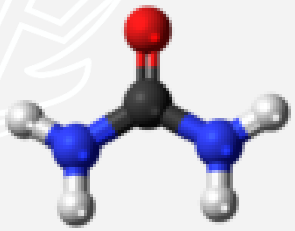
Philosophical Underpinnings

Stephen G. Ware

CSCI 4525 / 5525

Artificial Urea

- Urea is an organic compound used by animals to metabolize nitrogen.
- 1727 – Urea discovered by Herman Boerhaave
- 1828 – Artificial urea synthesized from silver cyanate and ammonium chloride
- The synthesis of organic compounds from inorganic materials contradicted *Vitalism*.



Recognizing Intelligence

- How do I know that I am intelligent?
- How do I know that someone else is intelligent?
- Is the human brain a machine?
- Can an electromechanical machine that *processes* like the brain have a mind?
- Can an electromechanical machine that *acts* like a brain have a mind?
- Can a machine have consciousness?

Various –ism's

- **Dualism**: Bodies are made of physical substance; minds are made of ethereal mental substance.
- **Physicalism**: Only physical substance exists.
- **Logical Positivism**: Only statements which can be empirically or logically verified are meaningful; all other statements are linguistic nonsense.

e.g. What does it mean to have free will?

Various –ism's

- **Mentalism**: Minds should be described by their internal states, such as thoughts and beliefs.
- **Behaviorism**: Because we cannot directly observe a mind, we should only concern ourselves with observable behaviors.

Various –ism's

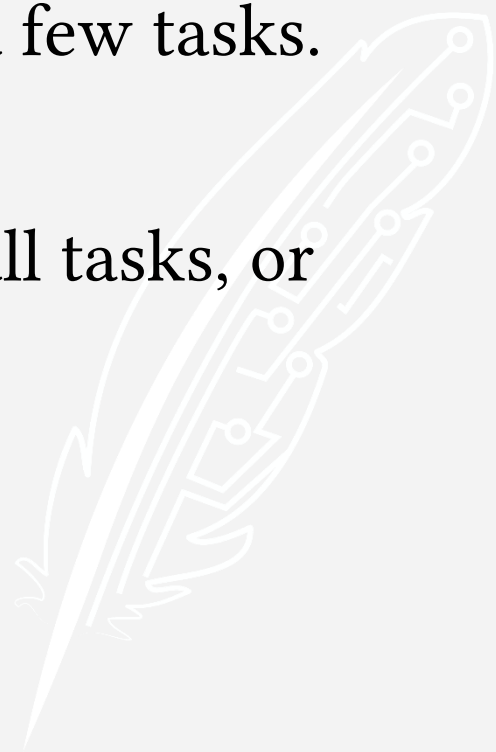
- **Epiphenomenalism**: Mental events are caused by physical events, but have no physical effects.
- **Biological Naturalism**: Mental phenomena are the emergent results of simpler, lower-level neurological processes.
- **Functionalism**: Mental states are constituted only by their functional role (that is, their relationship to sensory input and behavioral output).

Weak vs. Strong AI

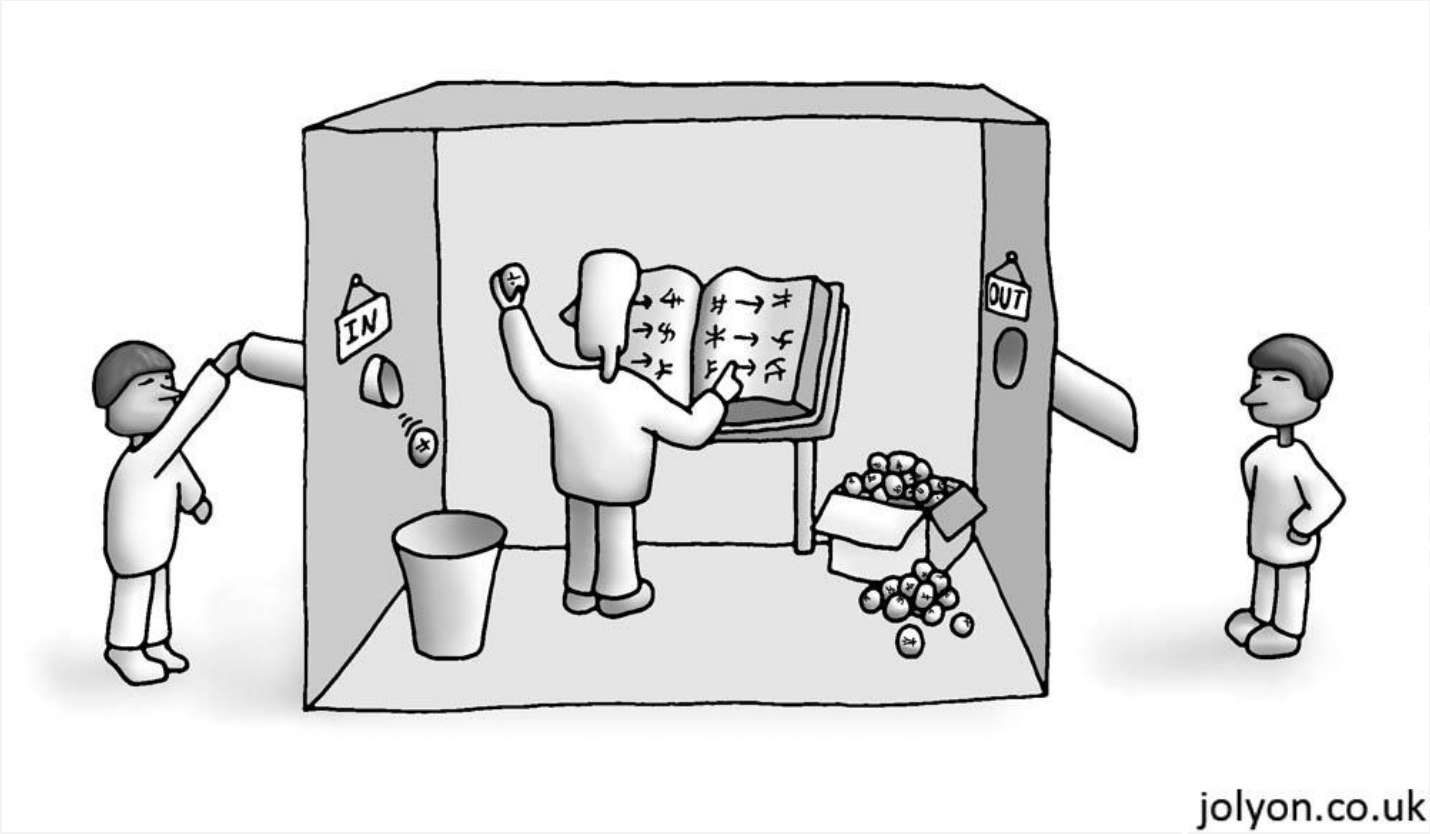
- **Weak AI**: Machines can act as if they are intelligent.

  Alternatively, a machine good at a few tasks.
- **Strong AI**: Machine can be intelligent.

  Alternatively, a machine good at all tasks, or an *artificial general intelligence*.



The Chinese Room



The Chinese Room

- Proposed by John Searle in 1980
- The input/output is the same as it would be for a human speaker who understands Chinese.
- Does the person in the room understand Chinese?
- Does the rule book understand Chinese?
- Does the room understand Chinese?
- Searle says brains cause minds, and while other machines might cause minds too, programs do not.

Can Machines Think?

"The question of whether *Machines Can Think*... is about as relevant as the question of whether *Submarines Can Swim*."

- Edsger Dijkstra

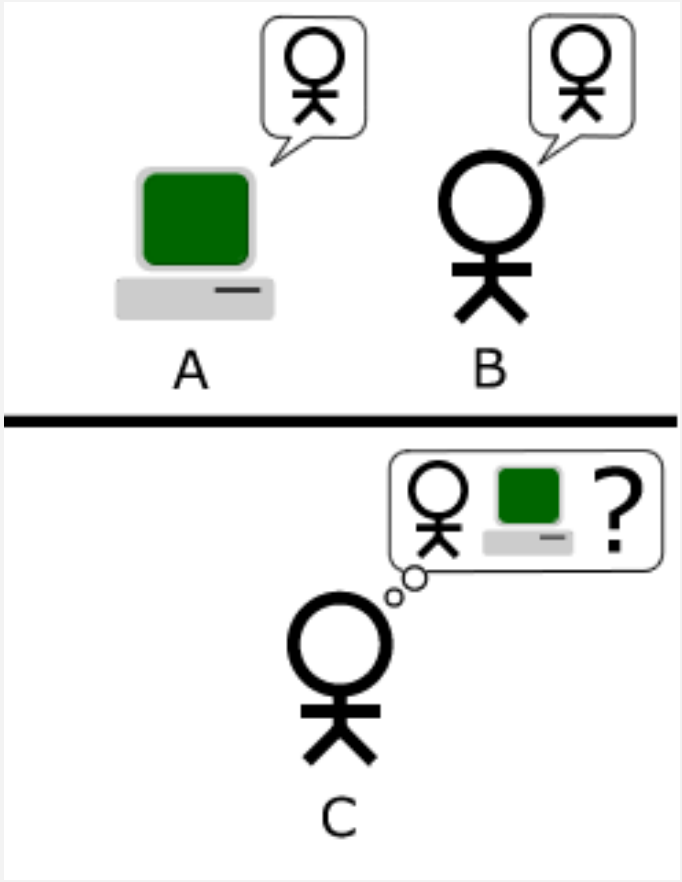
Purview of AI

"A problem is only an AI problem until it is solved."

Approaches to AI

- Acting Humanly
- Thinking Humanly
- Thinking Rationally
- Acting Rationally

Acting Humanly



- **Turing Test**: Humans converse with either another human or a computer and must guess which it is.
- Value has been questioned (e.g. typing errors)
- 33% of judges convinced in 2012 Loebner competition.

Thinking Humanly

- Aims to develop accurate computational models of how humans think.

  e.g. evolution of Neural Networks

- Symbiotic relationship to cognitive science.

Note: My research on computational models of narrative falls into this category.

Thinking Rationally

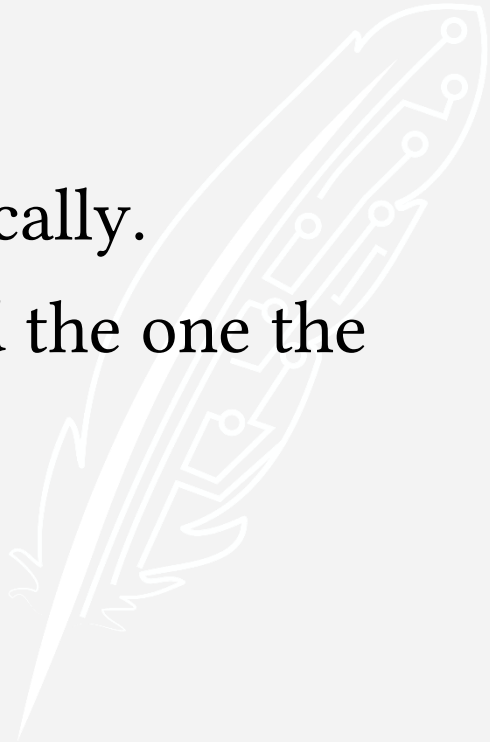
- Concerned with following provably correct procedures for gaining new knowledge.

  e.g. Logical deduction

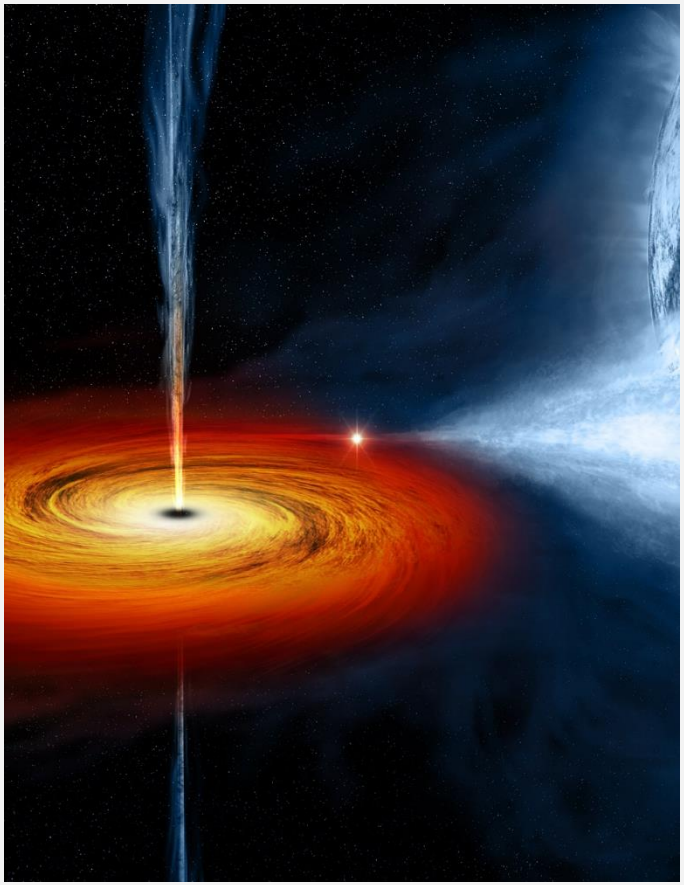
- Computer programs have authored new mathematical proofs.
- The real world is messy and hard to formalize.
- Formal models are often large and exploring them is computationally intractable.

Acting Rationally

- Given some performance measure, a **rational agent** acts to maximize it.
- Agnostic vis a vis method of thinking.
- Easier to demonstrate and test scientifically.
- This is the most common approach and the one the book is mainly concerned with.



The Singularity



- Where the measures of gravity (or technology) become infinite
- When matter passes the event horizon of a black hole, it can no longer be observed.
- The same may happen to technology if it continues exponential growth.

Ethical Concerns

- Loss of jobs
- Loss of human uniqueness
- AI systems used for evil
- Loss of accountability
- Human obsolescence / destruction
- Virtual agent and robot rights

One Final –ism

- **Transhumanism**: A fundamental change in human nature brought about through technological augmentation.

Roots of AI

- Logic
- Mathematics
- Economics and Game Theory
- Neuroscience
- Psychology and Cognitive Science
- Control Theory and Cybernetics
- Linguistics