CMPSCI 383
Introduction to Artificial Intelligence

Lecture 3: Philosophical Foundations

Instructor: Andy Barto
Teaching Assistant: Phil Thomas
Grader: Andy Bristol

http://www-anw.cs.umass.edu/~barto/383-Fall2011

The Big Questions

A Few of Them

- How can minds work?
- How do human minds work?
- Mind = Brain?
- Can non-humans have minds?
- Can a machine have a mind?
- Can a machine be conscious?
- Can a machine think?
- Does it matter what a machine is made of?

Ethics and Risks of AI

- What about jobs for humans?
- What if AI is used for bad things?
- What if AI takes over?
- Can we control AI?
- Can we design AI to be "friendly" to humans?
- ....

Weak AI or Strong AI

- The Weak AI position: Machines can be made to act as if they had intelligence, as if they had minds, as if they were conscious.
- The Strong AI position: Machines can be made that are intelligent, that have minds, that are conscious.

Is it only about words?

- Can submarines swim?
- Can machines fly?

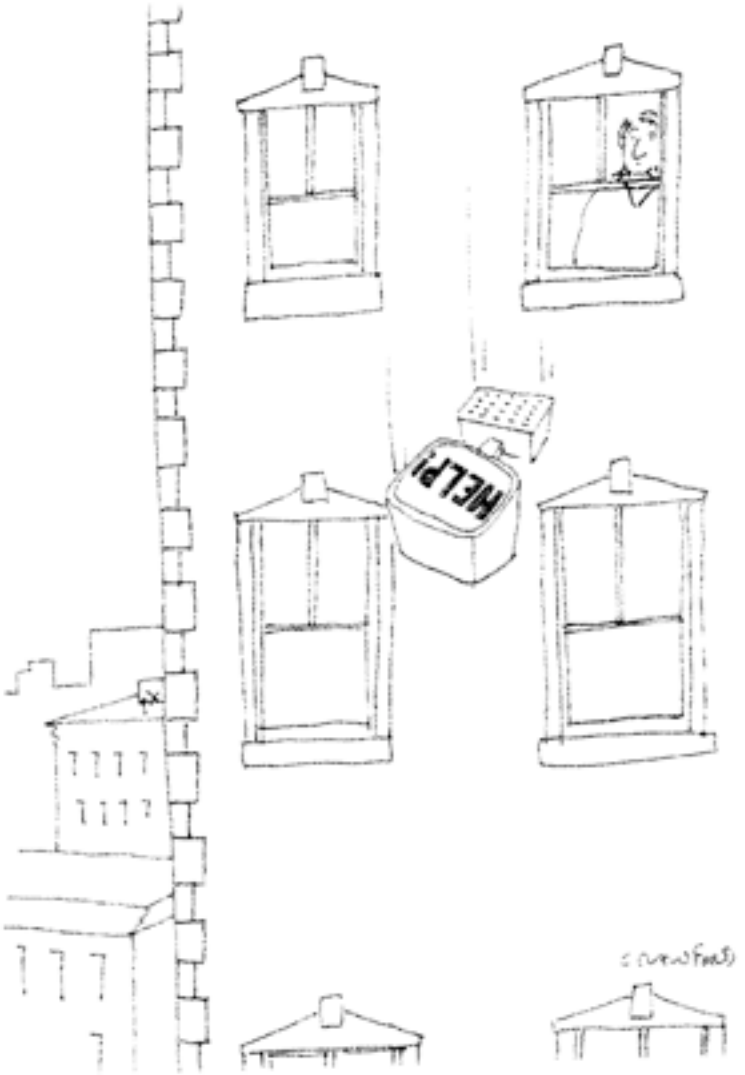
Turing Test


TURING TEST EXTRA CREDIT:
CONVINCE THE EXAMINER THAT HE'S A COMPUTER.
YOU KNOW, YOU MAKE SOME REALLY GOOD POINTS.
I'M ... NOT EVEN SURE WHO I AM ANYMORE.

Simulation does not equal Reality

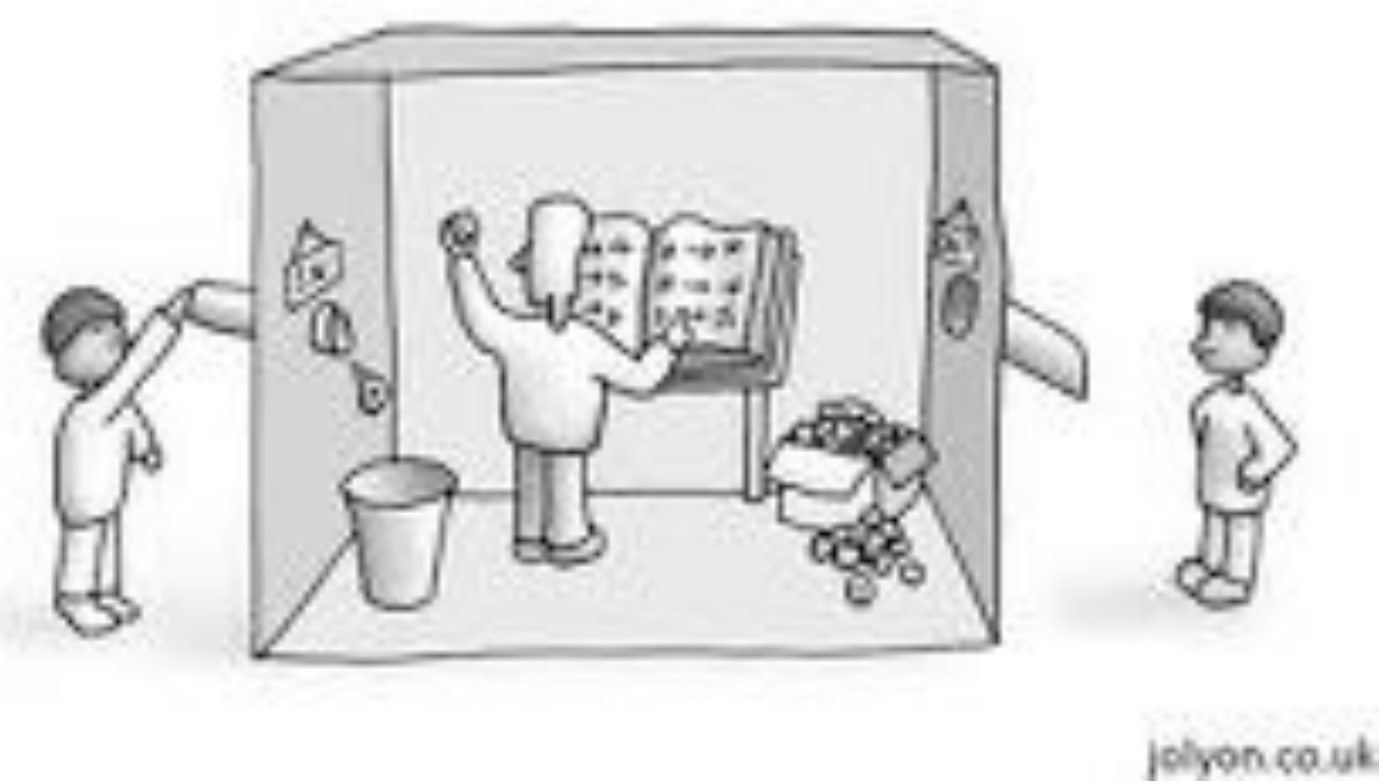
"No one supposes that a computer simulation of a storm will leave us all wet."

"formal symbols have no physical, causal powers."



How could anyone have supposed that a computer simulation of a mental process must be the real thing?

John Searle's Chinese Room



What is Searle trying to refute?

The version of Strong AI:

Thinking can be done by the manipulation of formal symbols (which is what computers do)

But:

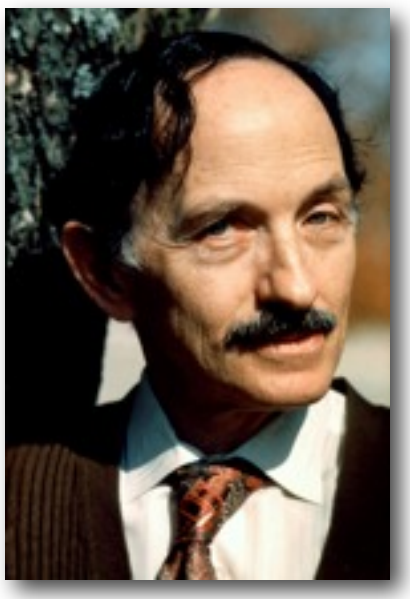
Computers might still be able to think, but they can't do it merely by running the right program.

A system does not have to be biological to think.

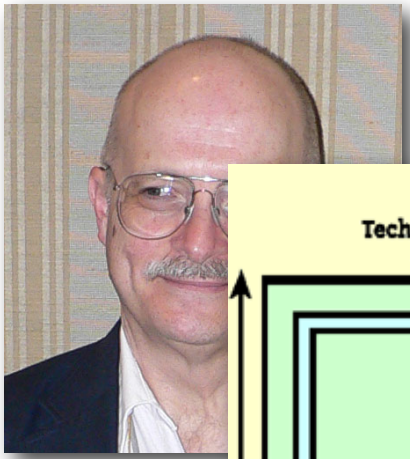
"Biological Naturalism"

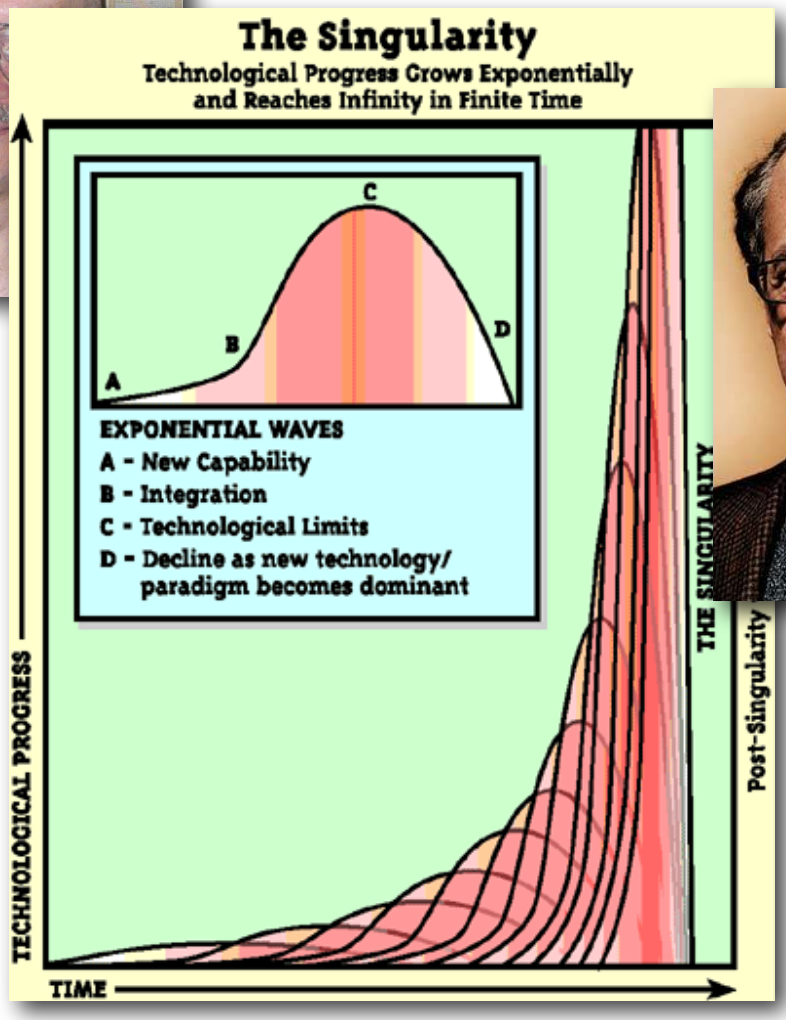
Brain Replacement Experiment

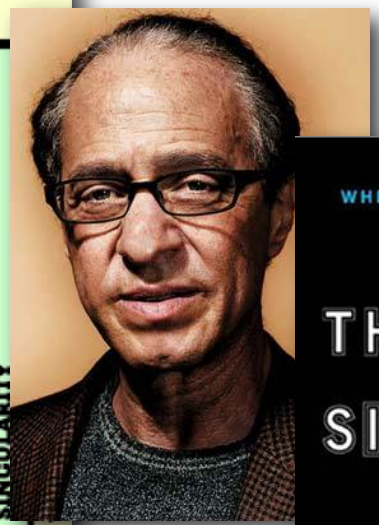
" Let an ultraintelligent machine be defined as
a machine that can far surpass all the
intellectual activities of any man however clever.
Since the design of machines is one of these intellectual activities,
an ultraintelligent machine could design even better machines;
there would then unquestionably be an 'intelligence explosion,'
and the intelligence of man would be left far behind.
Thus the first ultraintelligent machine
is the last invention that man need ever make."
— *I.J. Good (1965)*

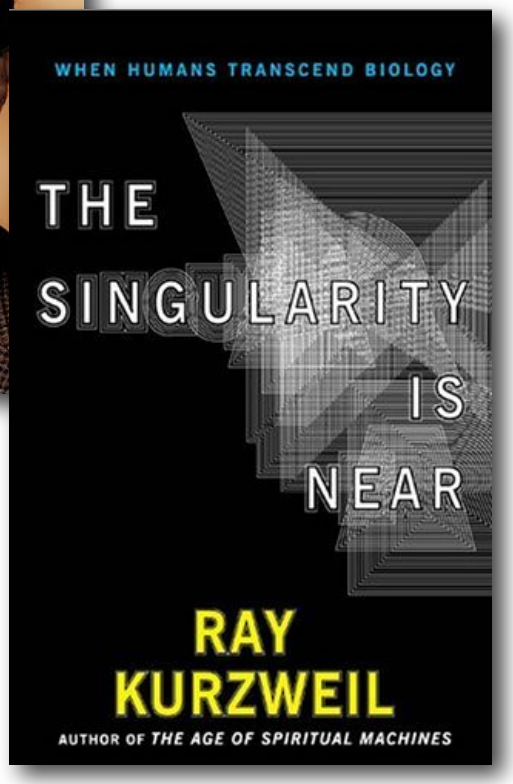


"Within thirty years, we will have the technological means
to create superhuman intelligence.
Shortly thereafter, the human era will be ended."
— *Vernor Vinge (1993)*









Bill Joy

Why the future doesn't need us.
Our most powerful 21st-century technologies - robotics, genetic engineering, and nanotech -
are threatening to make humans an endangered species.

Wired, April 2000

"Accustomed to living with almost routine scientific breakthroughs, we have yet to come to terms with the fact that the most compelling 21st-century technologies - robotics, genetic engineering, and nanotechnology - pose a different threat than the technologies that have come before. Specifically, robots, engineered organisms, and nanobots share a dangerous amplifying factor: They can self-replicate. A bomb is blown up only once - but one bot can become many, and quickly get out of control."

"The 21st-century technologies - genetics, nanotechnology, and robotics (GNR) - are so powerful that they can spawn whole new classes of accidents and abuses. Most dangerously, for the first time, these accidents and abuses are widely within the reach of individuals or small groups. They will not require large facilities or rare raw materials. Knowledge alone will enable the use of them. Thus we have the possibility not just of weapons of mass destruction but of knowledge-enabled mass destruction (KMD), this destructiveness hugely amplified by the power of self-replication."