



Foundations of AI

- 1. Is Weak and Strong AI possible?***
- 2. Chinese Room Experiment***



Introduction

Philosophers have formed two hypothesis:

a. Weak hypothesis.

Can machines act intelligently?

b. Strong hypothesis.

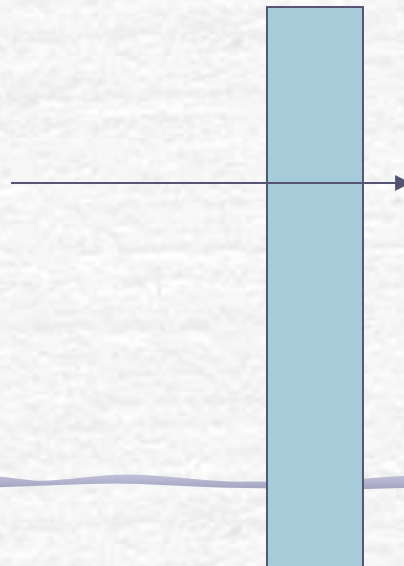
Do machines actually think?

Or do they simulate thinking?



Weak Hypothesis

Alan Turing in his paper: “Computing Machinery and Intelligence” suggested a test to decide if a computer can think or not.



Person?
Computer?
(test lasts 5 minutes;
success: 30% of the
time the interrogator is fooled).

First Objection Against “Weak AI”

1. Disability. Machines can’t do X.

Machines can do many things:

Play chess, inspect parts on assembly lines, check spelling, steer cars and helicopters, diagnose diseases, create new music, and even create art (see <http://www.ramos.nl/yyfire.html>)

Second Objection

2. Mathematical Objection.

Machines are limited by the
“Incompleteness Theorem”.

Let F be a formal axiomatic system powerful
enough to do arithmetic.

Second Objection (continued)

“Powerful Formal Systems are limited by *Gödel’s Incompleteness Theorem*”:

We can construct a Gödel sentence $G(F)$ such that:

- a. $G(F)$ is a sentence of F , but cannot be proved by F .
- b. If F is consistent, then $G(F)$ is true.

Consequence: Whatever formal system we choose there are things that cannot be proven by the formal system although they are true!

Counter Argument: It is not clear why there are not similar limitations on what humans can prove; moreover, the argument only applies to formal systems.

Controversy includes the books by Roger Penrose.

Soft Computing

Researchers might agree
and claim that this
is a problem of “*old AI*”

Third Objection

. Informality.


*“We cannot capture human behavior as a set
of logical rules”*

Known as the “Qualification Problem” in AI.

Hubert Dreyfus:

“What Computers Can’t do”

“Mind over Machine” (but no suggestions are
made on how to solve the problem)



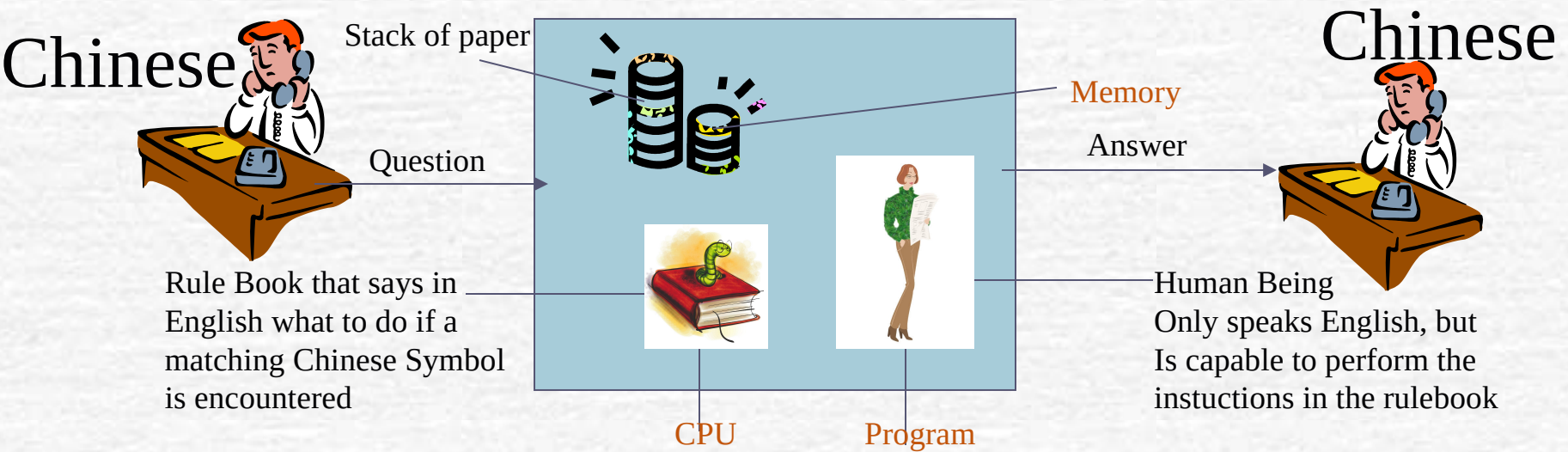
Strong Hypothesis --- Machines can “really think”

Attack against AI: machines do not actually think but only *simulate* thinking.

A. Turing: we are interested in creating machines that behave intelligently.

Chinese Room Thought Experiment

John Searle (1980): “*Running the right program does not necessarily create understanding*”



Searle argues neither the rule book nor the human being understands Chinese, therefore:
running the right program does not necessarily create understanding
 → refuting “strong AI”; argument does not deny that “*machines have minds*”