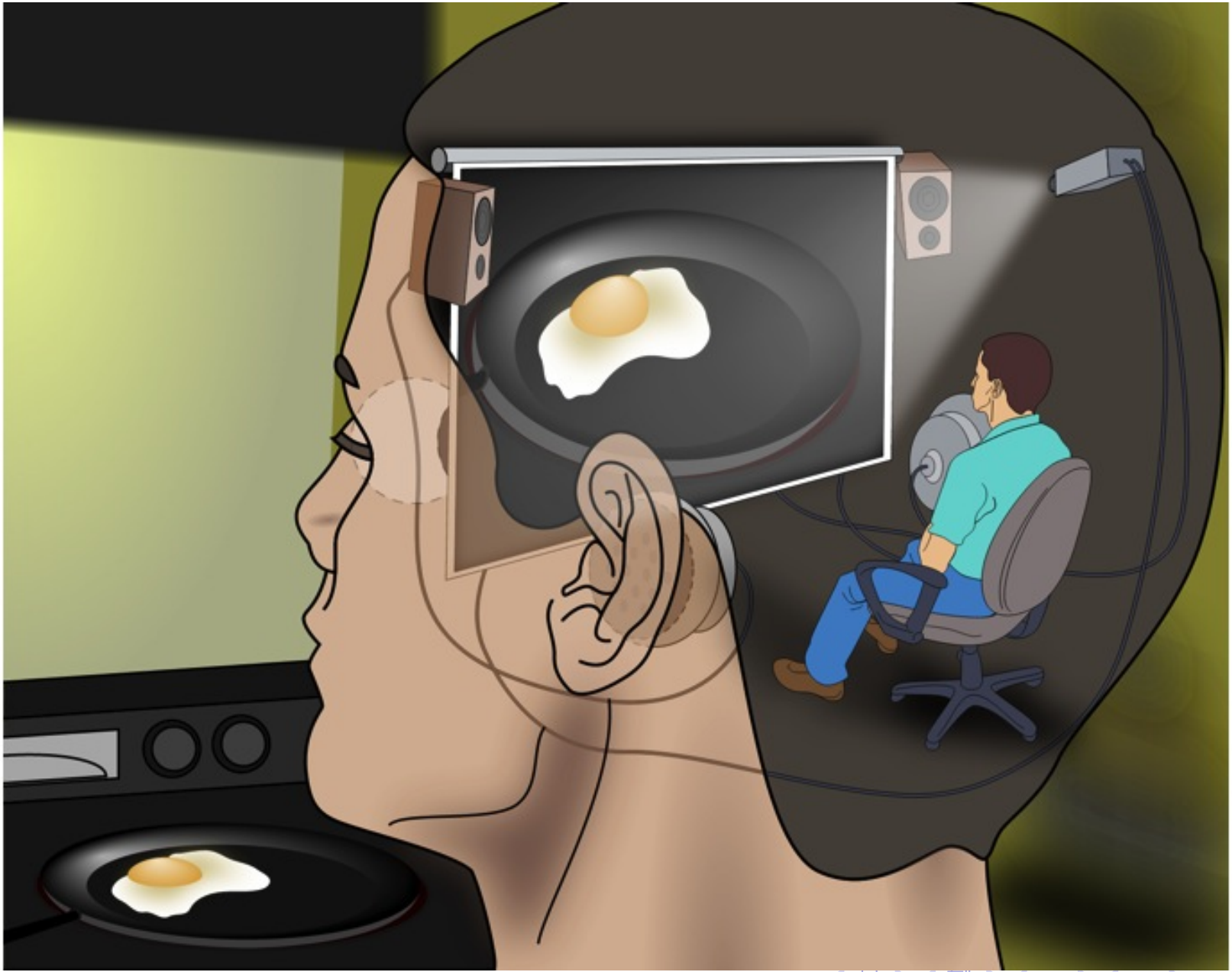
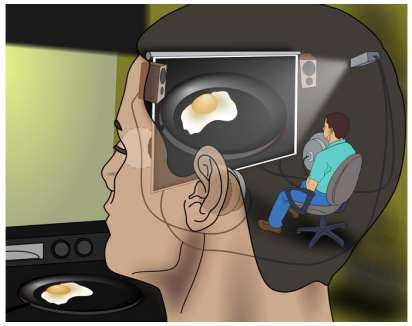
Section 1

Theories of Mind

Dualism



Dualism

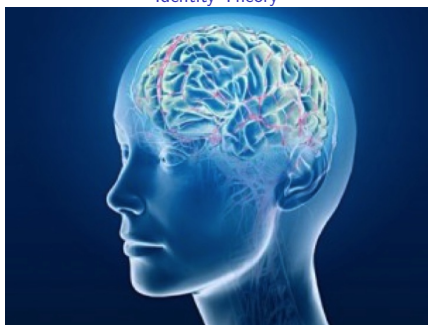


- Dualism says that states and processes of the mind are separate from states and processes of the brain.
- Consider an experience of pain, or of seeing something, or of having a mental image. Dualism says that these experiences are not identical to states and processes of the brain, even if they are correlated with them.

Behaviorism

- Logical behaviorism says that each mental state is identical to some specific set of behaviors and physiological responses.
- Consider an experience of pain, or of seeing something, or of having a mental image. Behaviorists says that these experiences are identical to some set of behaviors and physiologic responses. Notice that pain is not the cause of wincing, it is identical to it.

Identity Theory



- The identity theory of mind says that states and processes of the mind are identical to states and processes of the brain.
- Consider an experience of pain, or of seeing something, or of having a mental image. The identity theory of mind says that these experiences are identical to brain states and/or processes, not merely correlated with them.

Functionalism

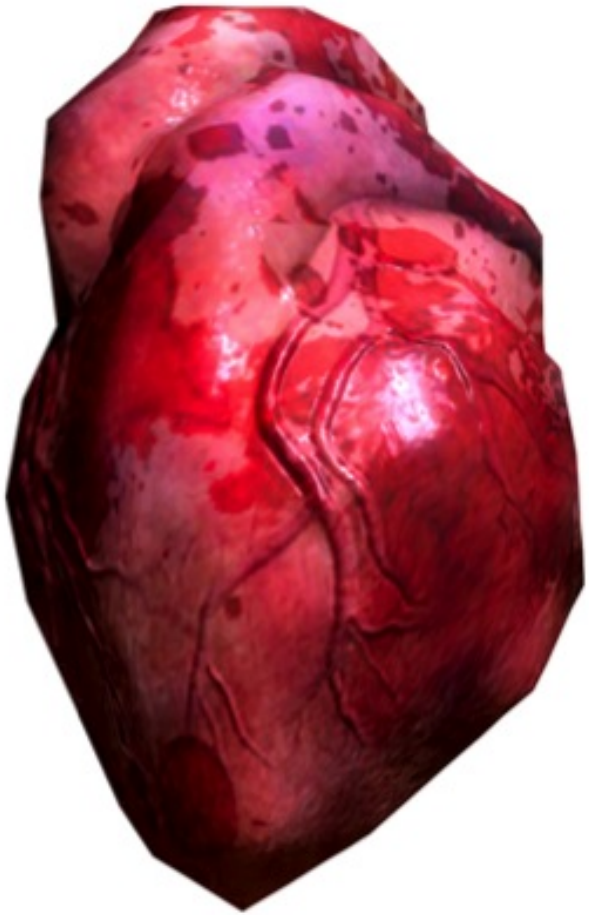
Who is the President?

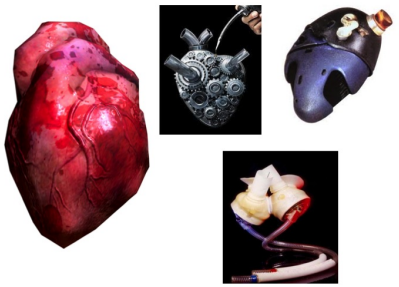
Office

Office-Holder

- Functionalists says that states and processes of the mind are defined by the causal roles they play in the cognitive system as a whole, i.e., what it is to be a particular mental state is to cause some result given some input.
- Pain is that mental state which disposes you to respond in certain ways given certain inputs, e.g. it disposes you to say 'ouch' when you scrap your knee.
- The pain state exists and it is defined by the job that it does, namely, disposing you to behave in a certain way given certain inputs. So it is distinct from the behavior it causes. It is also distinct from whatever physical state might happen to play the role of pain just as the office of the president is distinct from anyone who happens to play the role.

An Analogy





- Functionalists say that for every functionally defined mental state, there is something which realizes that state, i.e., plays the role given in the state.
- Very different kinds of entities might play the very same role.
- Consider the heart. What it is to be a heart is to pump oxygenated blood to the issues. Different kinds of entities made from different materials, having different structures, and even operating in different ways may all still count as being hearts because they each pump oxygenated blood to the tissues.

Does he have a heart?

Does he have a heart?

The function of the heart is to pump oxygenated blood to the tissues. To determine, then whether Tin Man has a heart, we have to determine whether this job is being performed in his body, i.e., determine whether any blood is being pumped around his body. If there is blood being pumped, then we can infer that he must have a heart—the thing that causes the pumping. Similarly, if mental states are functional states, we can determine whether an entity has a particular mental state by determining whether or not the job associated with that mental state is being performed in that entity.

Section 2

Functionalism and AI

Weak vs. Strong AI

Weak Artificial Intelligence (WAI)

- Computers give us a powerful tool to study the mind. Thinking may be modeled by formal symbol systems, such as computer programs.

Strong Artificial Intelligence (SAI)

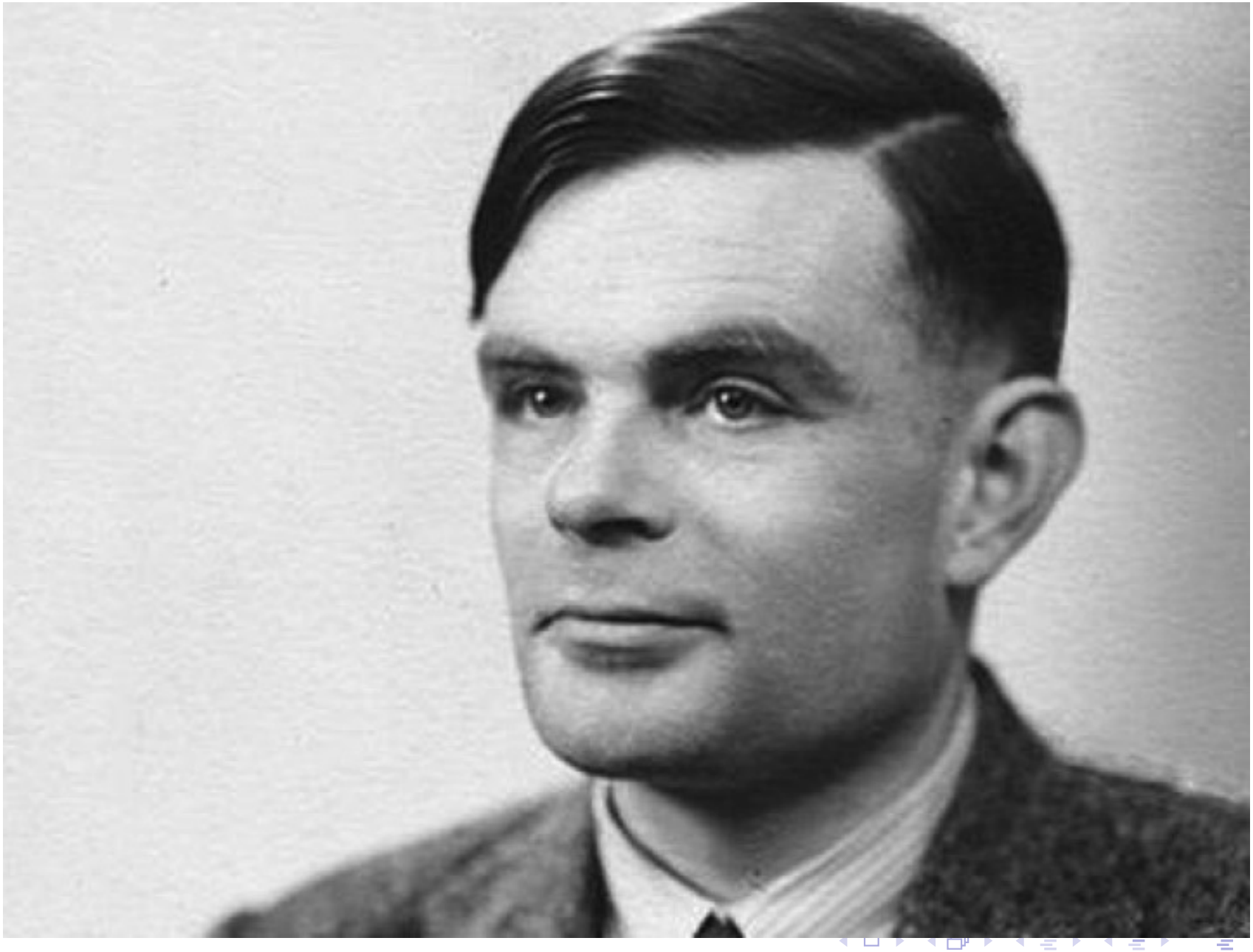
- Thinking is constituted by the manipulation of formal symbols, such as occurs in a computer program.

- Understanding a sentence is a mental state that plays a functional role, namely, disposes the entity who possess that state to respond in various ways given certain inputs.
- Given certain inputs, the computer performs in the way specified in the functional definition of understanding.
- Thus an appropriately programmed computer is a mind, where a computer is a machine whose operation is defined as an instantiation of a computer program.


ASIMO
HONDA

https://www.youtube.com/watch?v=n1AjtIAje3o

Alan Turing



Biography

- Cambridge mathematician. Father of Artificial Intelligence.
- Worked on Government Code and Cypher School in Bletchley Park and was key to breaking German naval Enigma.
- Prosecuted for homosexuality in 1952, chemically castrated, died in 1954 by suicide, and given a posthumous pardon in 2013.

Section 3

The Turing Test

- Can machines do what we as thinking beings can do?
- Can machine cause an interrogator to wrongly identify it to the same degree that an ordinary man or woman could?
- Are there *imaginable digital* computers that would do well in the game?

  If the interrogator decides wrongly as often when the game is played with the computer as he does when the game is played between a man and a woman, it may be argued that the computer is intelligent.

I believe that in about fifty years' time it will be possible to programme computers...to make them play the imitation game so well that an average interrogator will not have more than 70 percent chance of making the right identification after five minutes of questioning...I believe that at the end of the century the use of words and general educated opinion will have altered so much that one will be able to speak of machines thinking without expecting to be contradicted. (Alan Turing)

Digital Computer

- A Digital Computer is a machine intended to carry out any operation which could be done by a human computer. It has three parts:

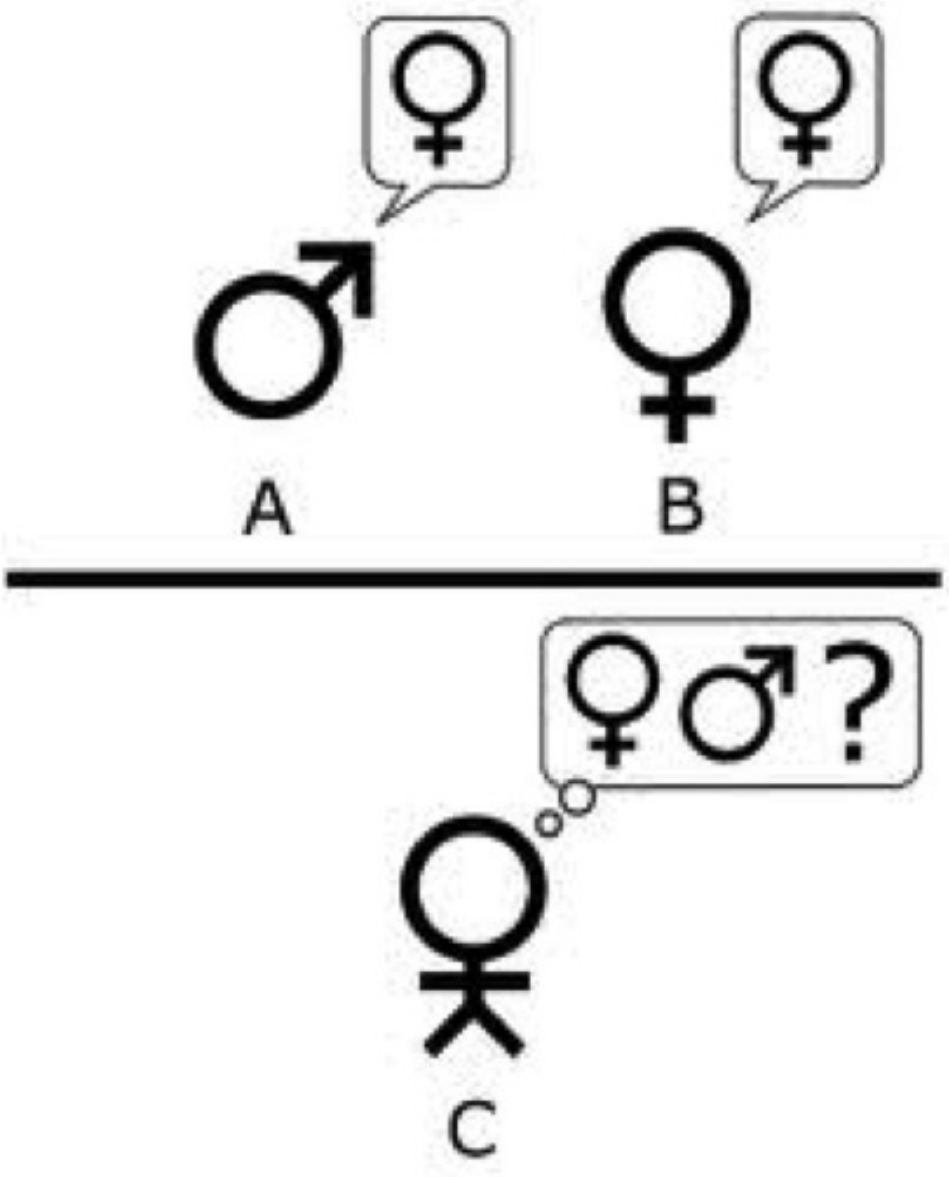
  Store is a store of information. It's like the paper or book that a human computer uses when doing their calculations.

  Executive unit carries out the various operations.

  Control is like the rules or instructions for carrying out the calculation. It's job is to ensure that instructions are obeyed correctly and in the right order.

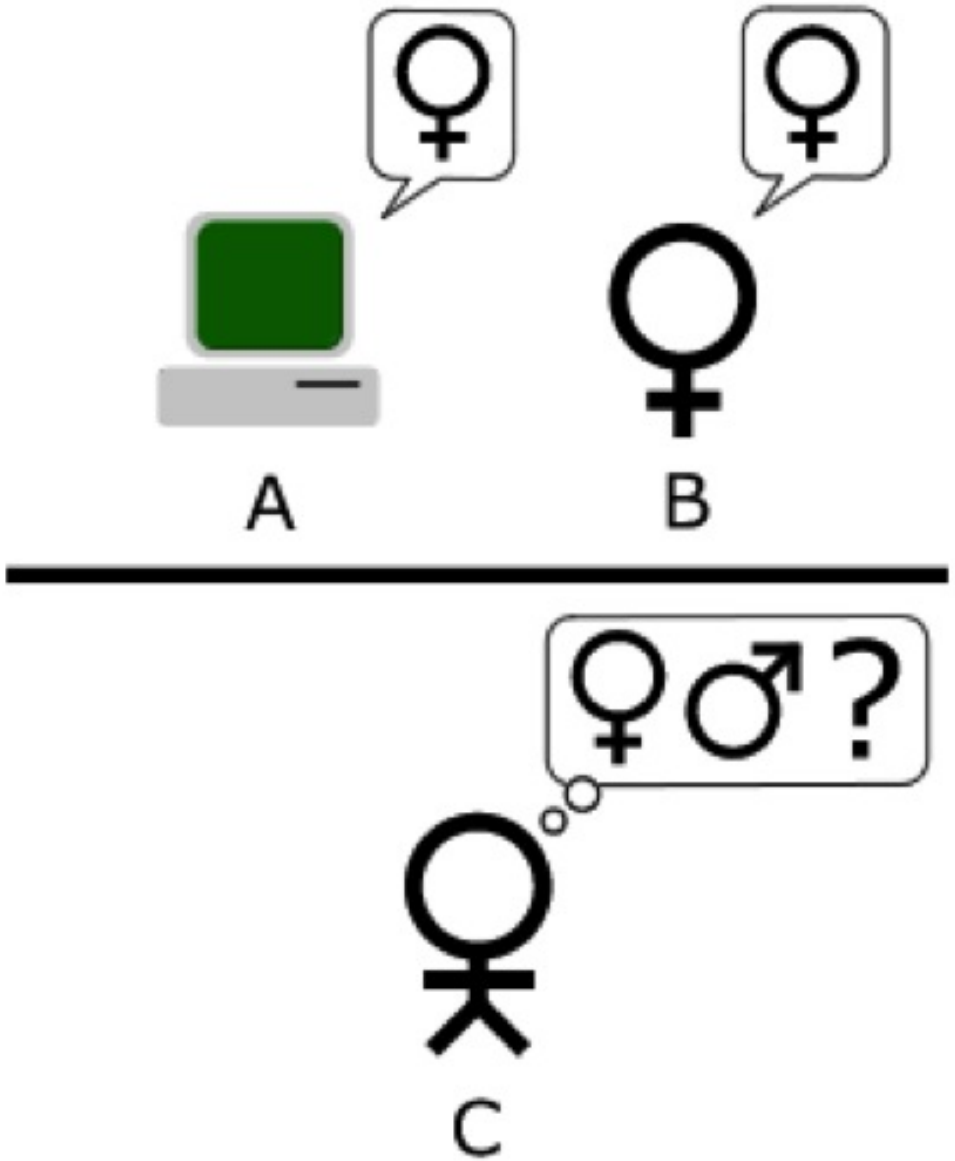
We can dramatically increase the size and power of these three parts. So even if we don't yet have a powerful enough computer that passes the Turing Test, we might build one by increasing the power and size of these parts.

Control Case

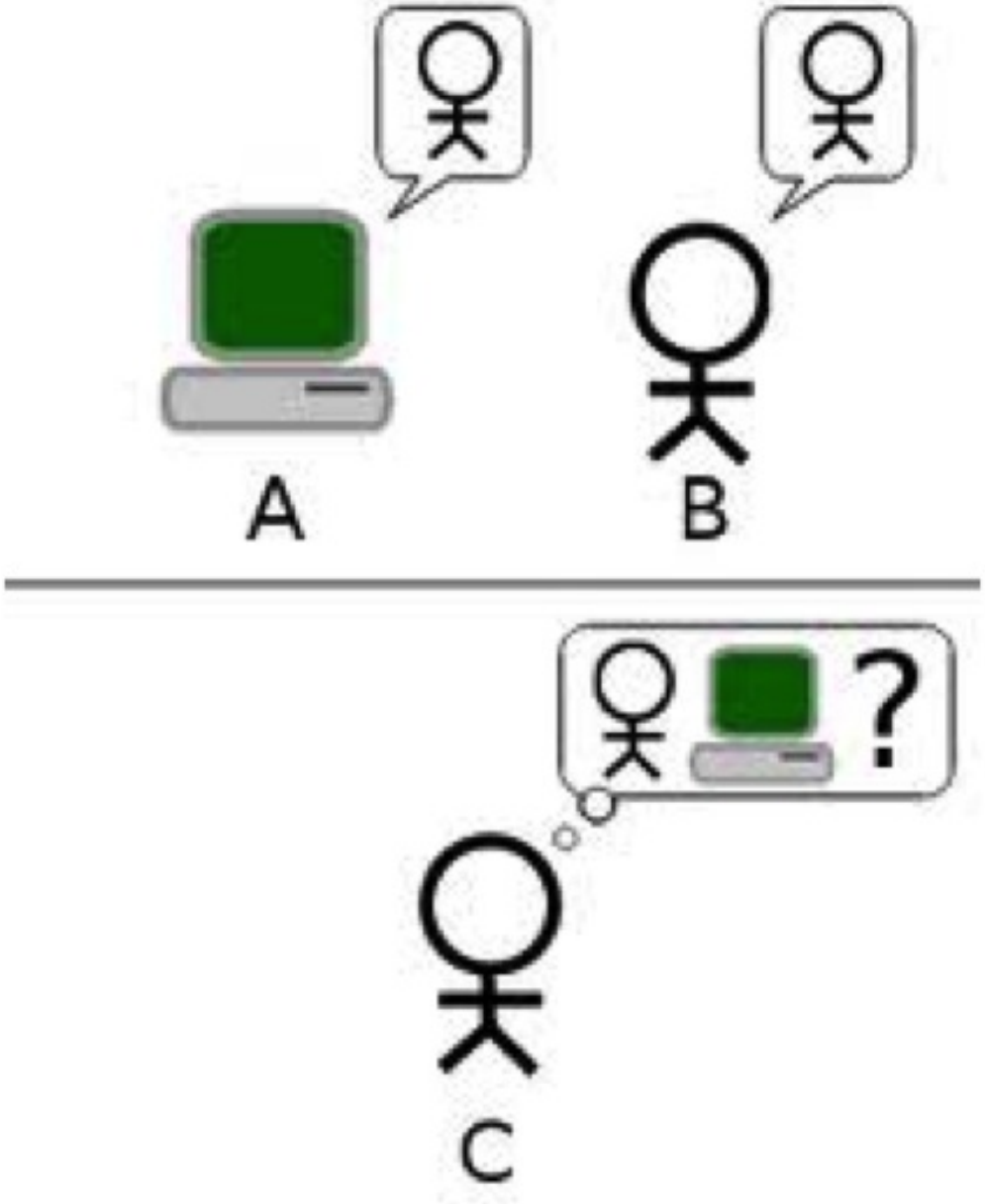


- Decision, convincing, and deception via teletype connection.
- Interrogator asks questions in written natural language.
- Receives answers in written natural language.
- Questions can be on any subject imaginable, from mathematics to poetry, from the weather to chess.
    - what's your name
    - what's your favorite color?
    - what does the smell of freshly cut grass remind you of?

Man or Woman?



Human or Computer?



Clarifying the Turing Test

Our Question

- Is passing the Turing Test necessary for intelligence?
- Is passing the Turing Test sufficient for intelligence?

Necessary and Sufficient Conditions

P is necessary for Q =

- Q cannot be true unless P is true.
- Q entails P.

P is sufficient for Q =

- P entails Q.

Examples

The Olympics

- Having access to a snowboard is necessary for competing in the snowboard contests at the Olympics, but it is is not sufficient for doing so.
- Winning a medal in a snowboard contest at the Olympics is sufficient for becoming an Olympic medal winner, but it is not necessary for becoming one.

Give an example of conditions that satisfy the following descriptions:

- Necessary, but not sufficient.
- Sufficient, but not necessary.
- Necessary and sufficient.

Our Question

- Is passing the Turing Test necessary for intelligence?
- Is passing the Turing Test sufficient for intelligence?

Turing believes:

- The question whether machines can think is meaningless.
- 'Being able to think' just means 'being able to pass the test'.
- Being able to pass the Test is then constitutive of being intelligent, i.e., both necessary and sufficient.

Summary

1. For some arbitrary time period, there may be no discernible difference between the linguistic behavior of a person and that of a machine.
2. If there is no discernible difference in linguistic behavior between man and machine, then there is no reason to think that there is any underlying difference in the causes of that behavior.
3. Thus if we are willing to say that it is intelligent thought that is the cause of the linguistic behavior in the person we should be willing to say the same thing about the machine.

Turing's Defense of the Imitation Game

Turing defends his claim that passing the test is constitutive of being intelligent by responding to several objections.

1. Argument from Extrasensory Perception
2. Argument from Various Disabilities
3. The Argument from Consciousness
4. The Mathematical Objection
5. Argument from the Informality of Behavior
6. Lady Lovelace's Objection

ESP Objection

Objection

If the human participant in the game was telepathic, then the interrogator, who is also human, could exploit this fact in order to determine the identity of the machine.

Response

Turing proposes that the competitors should be housed in a "telepathy-proof room."

Argument from Various Disabilities

Objection

Machines cannot do many things that humans can, e.g.,be kind, be resourceful, be beautiful, be friendly, have initiative, have a sense of humor, tell right from wrong, make mistakes, fall in love, enjoy strawberries and cream, make someone fall in love with it, learn from experience, use words properly, be the subject of its own thought, have as much diversity of behavior as a man, do something really new.

Response

- The objection relies on an inductive inference. You know that particular machines have various disabilities and then infer that all present and future machines will have various disabilities.
- Induction can fail!

Induction

Explication

Induction is a sort of non-deductive inference. We infer that since some instances of a certain group have a certain attribute that the next members of that group will have that attribute.

Examples

- The sun rose every day that I remember, hence, the sun will rise tomorrow.
- Every lion I have seen was a golden yellow color, hence every lion I will see will be a golden yellow color.

The Problem with Induction

Inductive inferences are not deductive valid. They are also difficult to justify.

The Argument from Consciousness

Objection

> *"Not until a machine can write a sonnet or compose a concerto because of thoughts and emotions felt, and not by the chance fall of symbols, could we agree that machine equals brain—that is, not only write it but know that it had written in it."*

1. A being can think only if it can know that it thinks.
2. Computers cannot know that they think.
3. Computers cannot think. (From 1 and 2)

Response: Reject 2

1. The evidence we could have that a being is self-aware is either from their reports or from being that being.
2. We cannot be another being.
3. So the only evidence we have that another being is self-aware is their reports that they are self-aware. (From 1&2)
4. Computers can report that they are self-aware. (From performance in Turing Test)
5. So we have evidence that they are self-aware. (From 3&4)

In addition, if the machine gave sufficiently complex answers about how it wrote poetry, we would likely think that it is, in fact, conscious.

The Mathematical Objection

Objection

1. Within any consistent formal system there are statements that can neither be proved nor disproved within that system.
2. Suppose that a computing machine is a formal system that operates by proving or disproving various statements put to it, i.e., by answering 'yes' or 'no' to various questions put to it.
3. There will be questions a computing machine gives wrong answers to (From 1 and 2; the machine, in some cases, should say that there is no correct answer.)
4. Intelligent beings are not so limited.
5. A computing machine is not intelligent. (From 3 and 4)

Response: Reject 4

1. There are questions that humans can't answer, yet they are intelligent.
2. It could be that the inability to answer every question is due to precisely the same limitations that restrict the capacity of machines.

Objection from the Informality of Behavior

Objection

''If each man had a definite set of rules of conduct by which he regulated his life he would be no better than a machine. But there are no such rules, so men cannot be machines."

Response. Distinguish

Laws of Conduct: Precepts such as "Stop if you see red lights," on which one can act, and of which one can be conscious.

Laws of Behavior: Laws of nature as applied to a man's body such as "if you pinch him he will squeak." (We do not act upon these laws, nor is it important that we are conscious of them.)

- Denies that being regulated by laws of conduct are relevant for determining if X is a machine.
- Agrees that if there is a definite set of laws of behavior that regulate X's life, then x is a machine.
- Agrees that there is no definite set of laws of conduct that regulate human life.
- Denies that there is no definite set of laws of behavior that regulate human life.

Lady Lovelace's Objection

Objection

Machines have "no pretensions to *originate anything*. It can do *whatever we know how to order it to perform*."

Response

- The objection assumes that machines cannot learn.
- But it is possible for machines to learn.

Section 4

The Chinese Room Thought Experiment

Weak vs. Strong AI

Weak Artificial Intelligence (WAI)

- Computers give us a powerful tool to study the mind. Thinking may be modeled by formal symbol systems, such as computer programs.

Strong Artificial Intelligence (SAI)

- Thinking is constituted by the manipulation of formal symbols, such as occurs in a computer program.

Argument for SAI

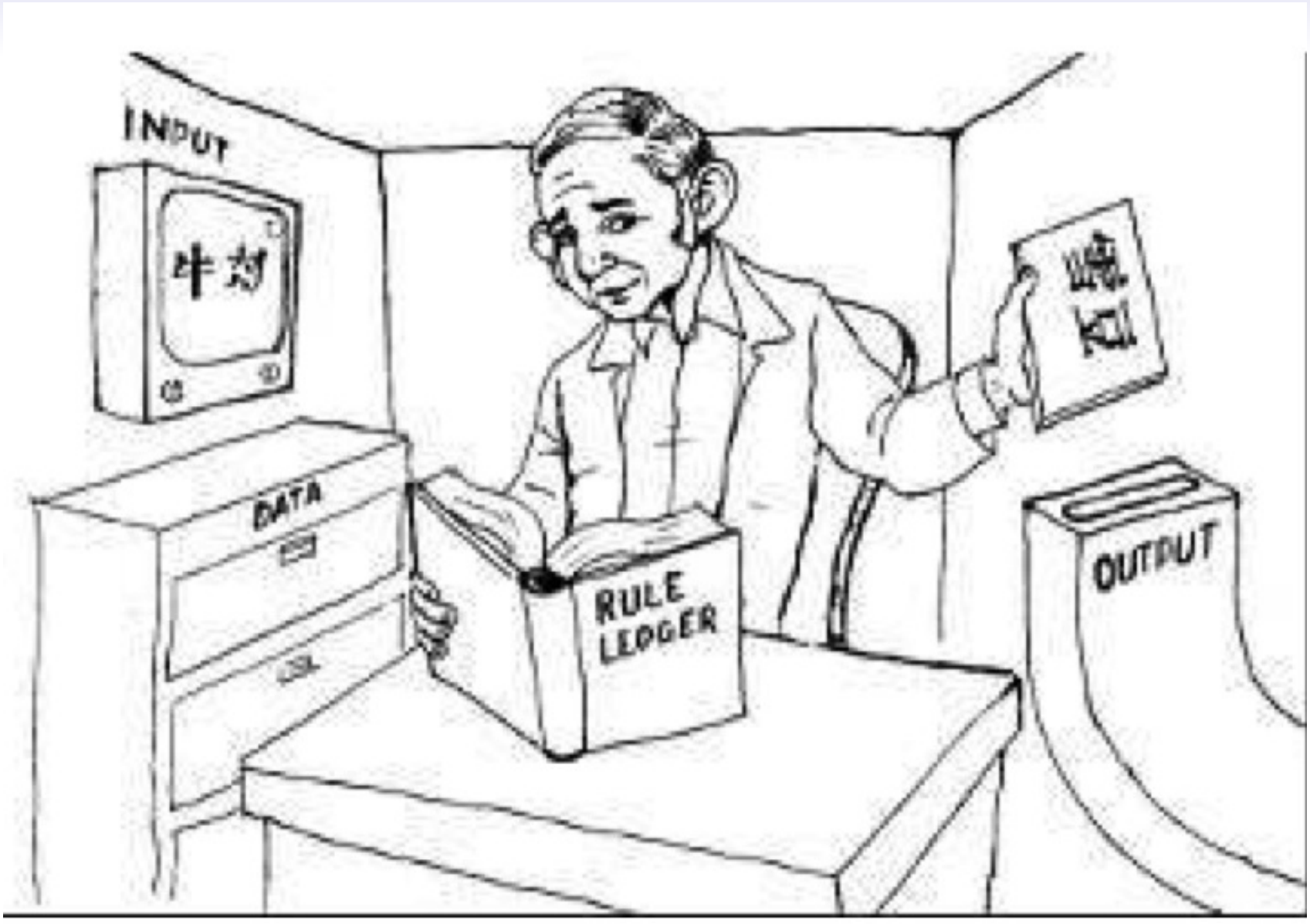
P1 Understanding a sentence is a mental state that plays a functional role, namely, disposes the entity who possess that state to respond in various ways given certain inputs.

P2 Given certain inputs, the computer performs in the way specified in the functional definition of understanding.

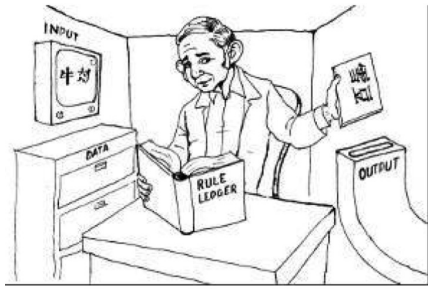
C Thus an appropriately programmed computer is a mind, where a computer is a machine whose operation is defined as an instantiation of a computer program.

Strategy

1. Searle accepts WAI, but rejects SAI.
2. He argues against P1 by describing a thought experiment, the Chinese Room Thought-Experiment, which seems to show that understanding is not merely a state which causes you to respond appropriately to certain inputs.
3. He then responds to various objections.


INPUT
牛对
DATA
RULE
LEDGER
OUTPUT

The Chinese Room Thought Experiment

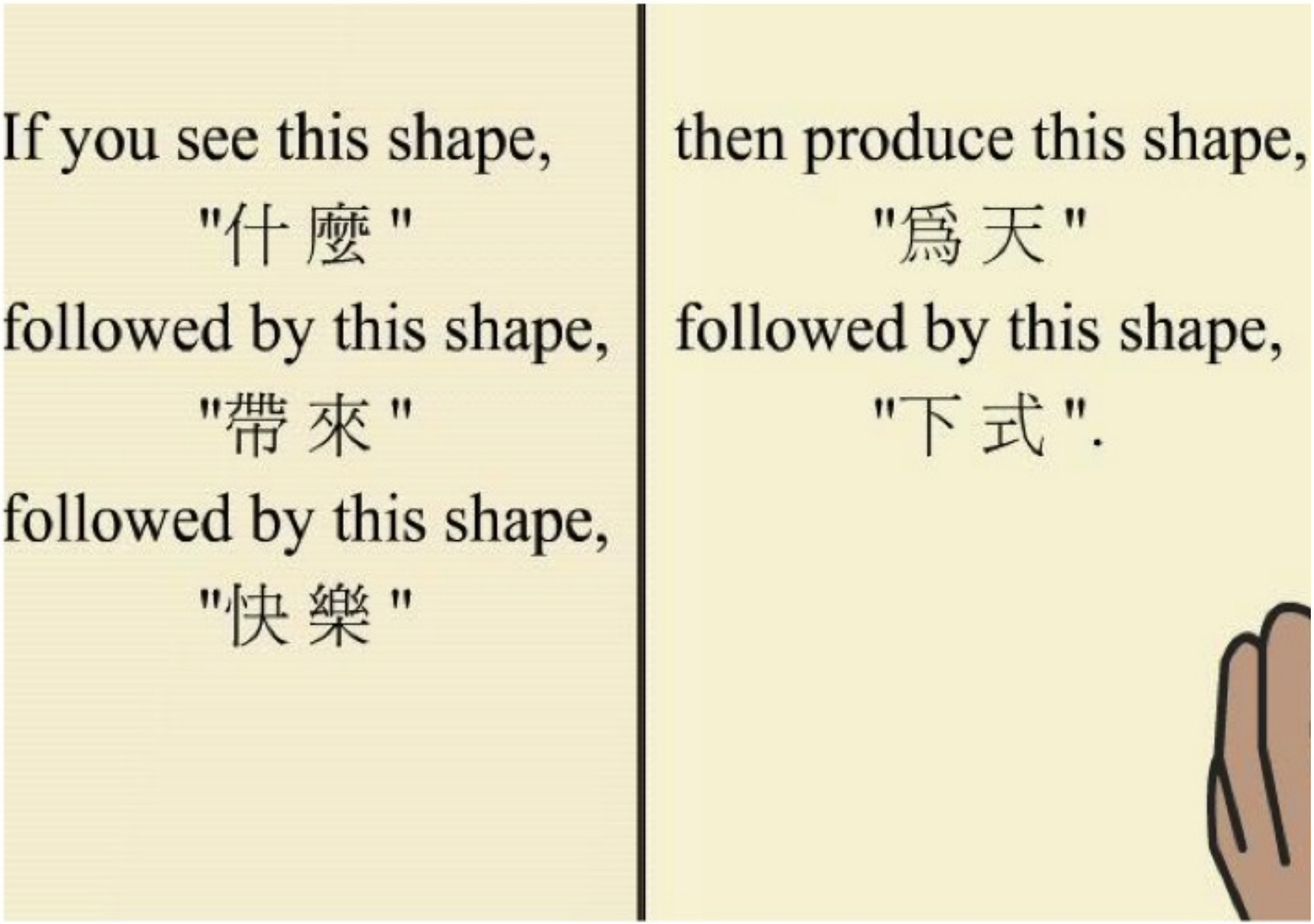


Input: Chinese symbols that unbeknown to Searle are questions in Chinese.

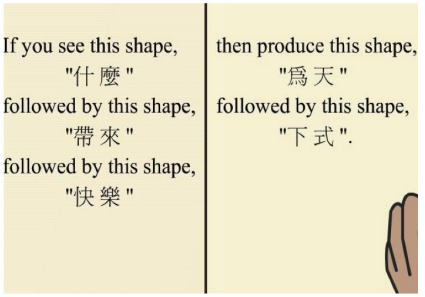
Data-Base: Boxes of Chinese symbols.

Program/Instructions: Book of instructions for manipulating the symbols.

Output: Chinese symbols that unbeknown to Searle are answers in Chinese.


If you see this shape,
"什麼"
followed by this shape,
"帶來"
followed by this shape,
"快樂"
then produce this shape,
"爲天"
followed by this shape,
"下式".

The Chinese Room Thought Experiment



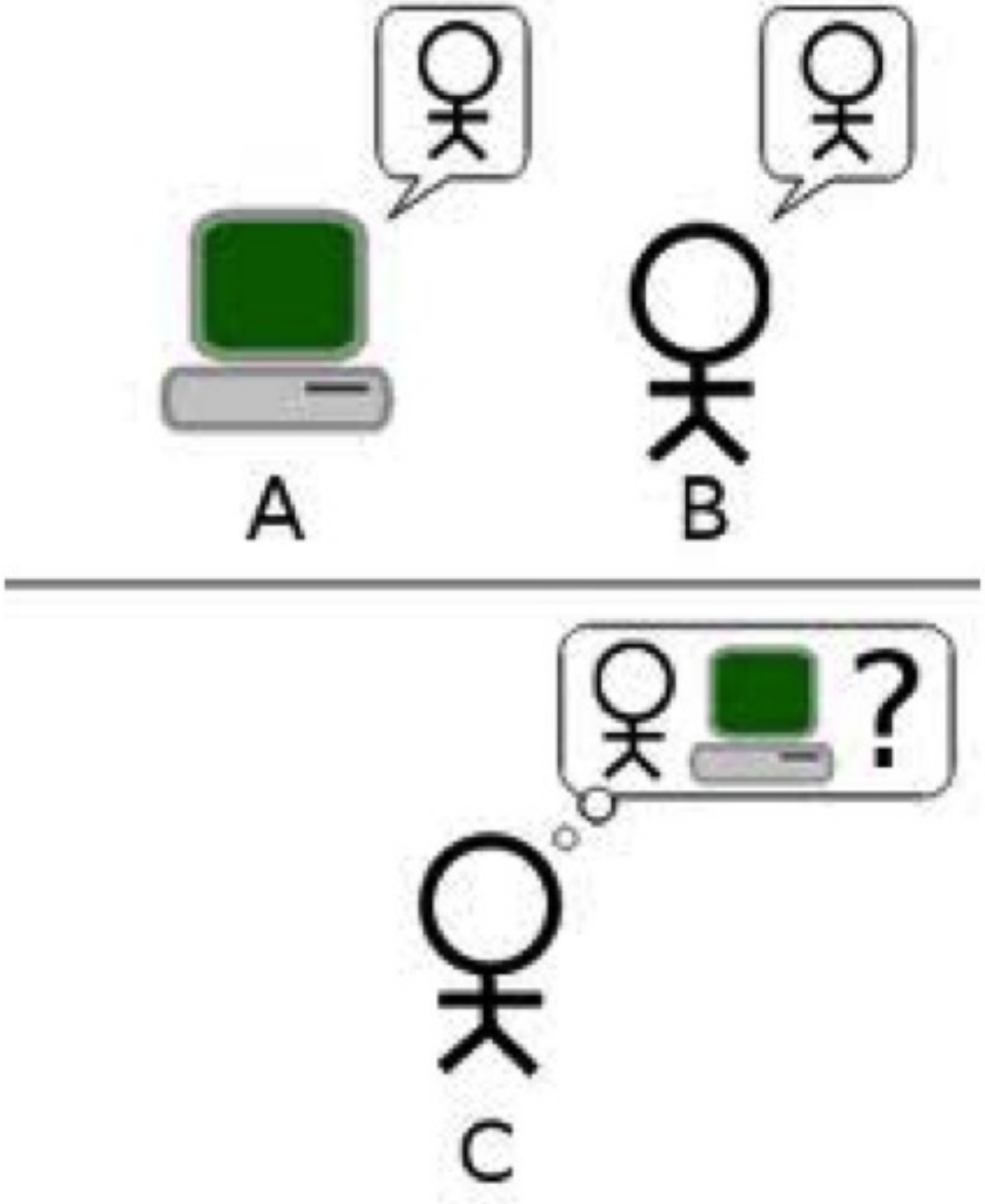
Imagine a native English speaker who knows no Chinese locked in a room full of boxes of Chinese symbols (a data base) together with a book of instructions for manipulating the symbols (the program). Imagine that people outside the room send in other Chinese symbols which, unknown to the person in the room, are questions in Chinese (the input). And imagine that by following the instructions in the program the man in the room is able to pass out Chinese symbols which are correct answers to the questions (the output). The program enables the person in the room to pass the Turing Test for understanding Chinese but he does not understand a word of Chinese.

Digestion Time

What does Searle think his Chinese Room thought-experiment shows?

- Would SAI say that Searle understands the Chinese stories? Why, why not?
- How exactly does Searle's apparent lack of understanding threaten SAI?

Relation to Turing Test



Does Searle's thought-experiment allow us to draw any conclusions about the adequacy of Turing's "imitation game," which is supposed to stand in for the question "Can machines think"?

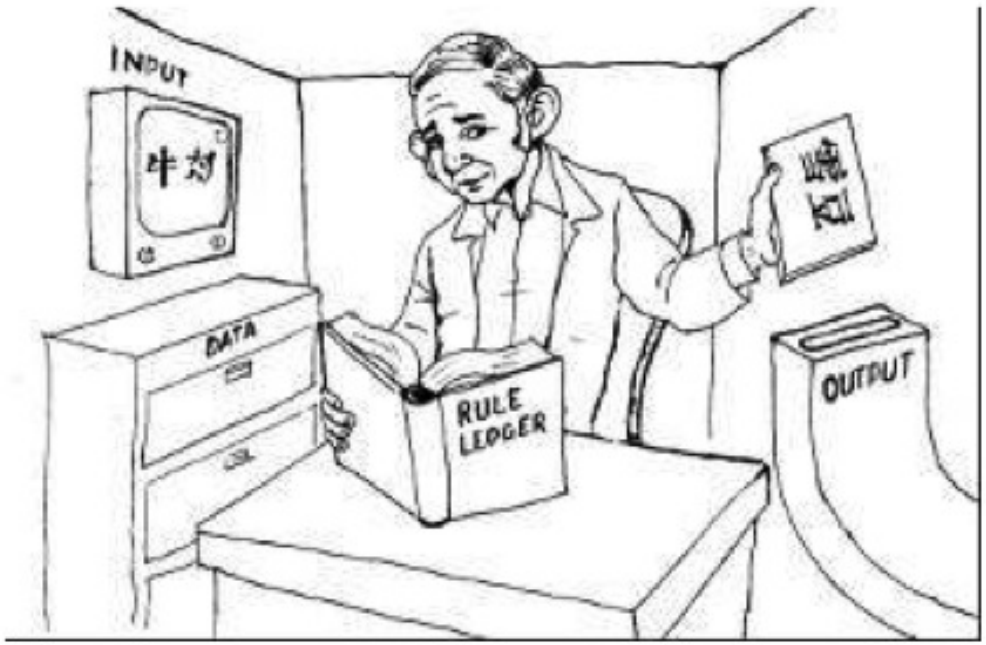
Section 5

Objections & Responses

1. The Systems Reply
2. The Robot Reply
3. The Brain Simulator Reply
4. The Combination Reply
5. The Other Minds Reply
6. The Many Mansions Reply

Systems Reply

The Whole



UNDERSTANDS

The systems reply objects that even though the human inside of the Chinese room doesn't understand Chinese, nevertheless the entire system, including books and pieces of paper, does understand Chinese.

The Fallacy of Composition

The fallacy of composition arises when one infers that something is true of the whole from the fact that it is true of some part of that whole. For example, this is a fallacy:

P1 Searle's hand weighs 2lb.

P2 Searle's hand is a part of him.

C Searle weighs 2lb.

Response

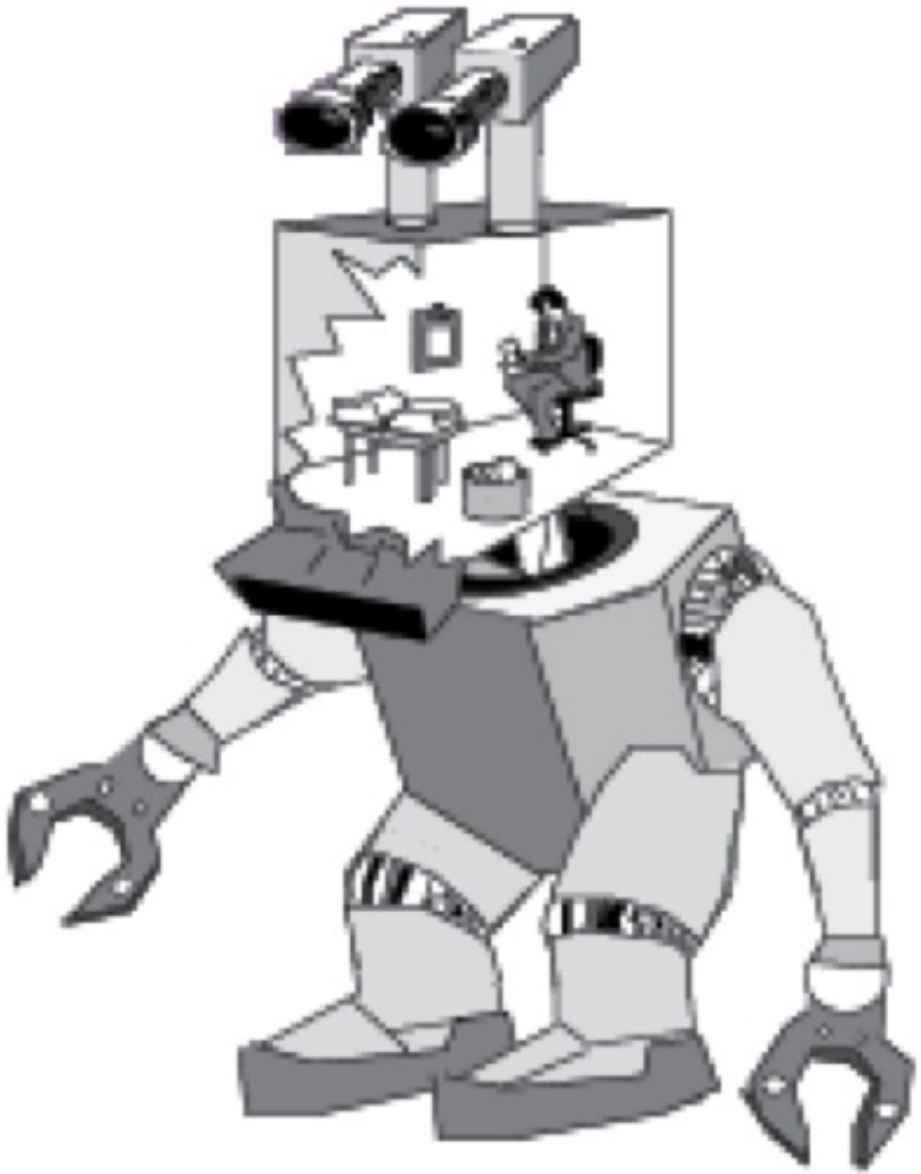
But Searle thinks that this objection, in addition to being incredible, overlooks the fact that the same thought-experiment can be run even if we assume that the human has internalized all of the books and other parts of the system.

Review

1. If Strong AI is true, then there is a program for Chinese such that if any computing system runs that program, that system thereby comes to understand Chinese.
2. I could run a program for Chinese without thereby coming to understand Chinese.
3. Therefore Strong AI is false.

The Robot Reply

Searle's Response



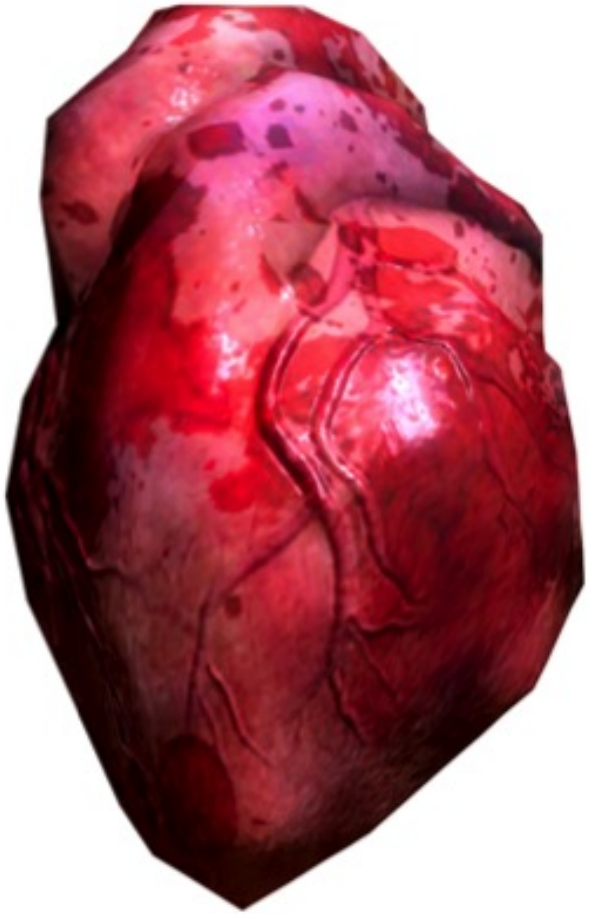
The Many Mansions Reply

Reply

Even if current computers cannot understand, we will probably be able to build machines that process in whatever way is required for understanding.

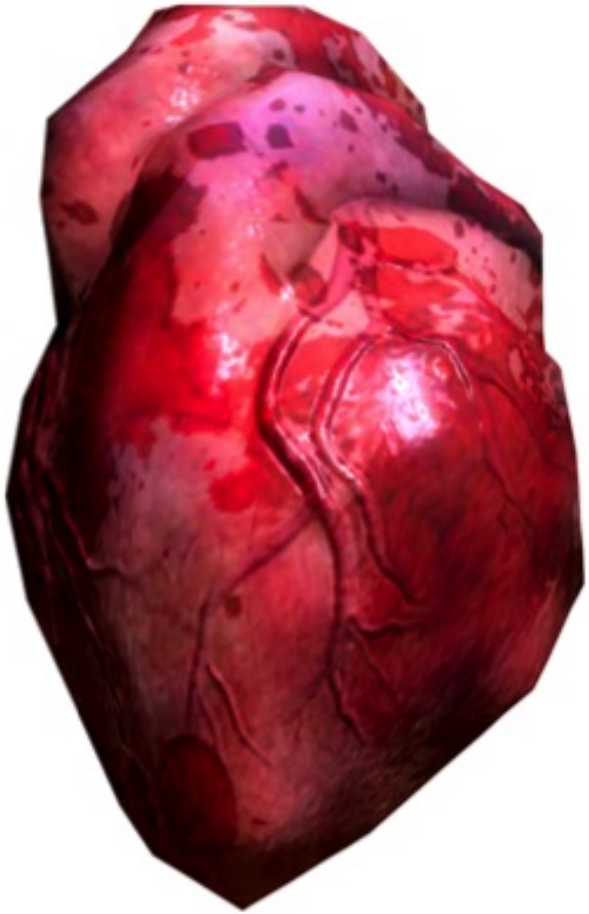
Searle's Response

Even if we could build such computers, the reply trivializes the claims of SAI because SAI is supposed to give us insight into what understanding is, i.e., that it is a functionally defined state.



Brain Simulator Reply

If we could make a computer that exactly parallels whatever goes on in a fluent Chinese speaker's brain when he or she understands Chinese, then that computer would understand Chinese.



Searle's Response

But Searle responds that even in such a case, there would be no understanding, because you could just change the original thought experiment so that the man in the room turns on and shuts off water valves that correspond to the neural firings of a Chinese speaker. In such a case, the man doesn't understand.