The Chinese Room

Introduction to Cognitive Science

The Chinese Room Thought Experiment

- Suppose that there is a room in which there is a person who doesn't speak or understand Chinese.
- However, the room is set up so that this person can have a productive conversation in Chinese with any Chinese speaking people that are outside the room:
  - The conversation takes place by passing (through a slot in the wall) pieces of paper on which Chinese expressions are written.
  - The person inside the room has a big rule book that the person can consult to transform Chinese expressions into other Chinese expressions.
  - The rules in the rule book are such that the resulting expressions are 'intelligent' responses to whatever it was that the Chinese speaking people on the outside were asking or saying.


I'm just manipulating squiggles and squoggles to produce Chinese language behavior. But I don't understand Chinese. This rule book is in English.
在這屋裡的任何人或物一定懂中文。
[Whoever or whatever is in that room is an intelligent Chinese speaker!]
Take a squiggle-squiggle sign from basket number 1 and put it next to a squoggle-squoggle sign from basket number 2.
1
2
3

Chinese Room and the Turing Test

- The Chinese Room thought experiment can be used to argue against the Turing Test as a test for intelligence (or as a justification for the attribution of intelligence and various other mental states):
  - The Chinese Room scenario shows that someone can pass the Turing Test without having any understanding of whatever conversations are taking place
  - But this understanding (this grasp of meaning; of what the conversations were about) seems to be an important part of intelligence
  - Therefore, the Chinese Room scenario shows that something can pass the Turing Test without being intelligent (or at least: without having any intelligence or understanding regarding whatever the Chinese conversation is about)!

Chinese Room and Computationalism

- The Chinese Room argument seems to be a sophisticated version of the following common objection to computationalism (and the possibility of thinking machines):
  - "Machines *just* crunch numbers or (better put) just manipulate symbols. They don't understand what the symbols actually mean!

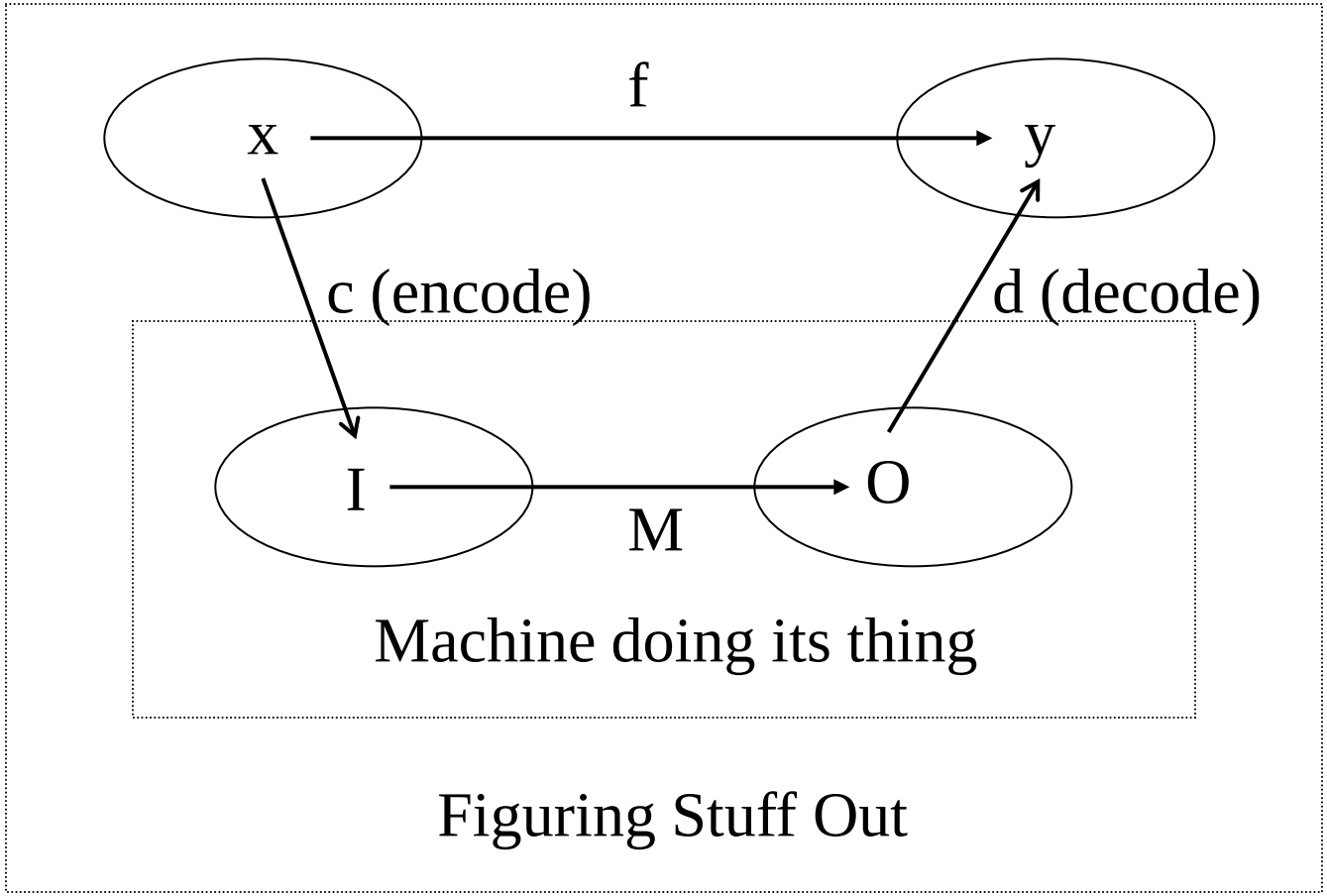
Calculators as Tools

- Does a calculator understand what it is doing?
- Or is a calculator just a *tool*? Something that is the automation of various symbol-manipulation algorithms that *we* can use to figure stuff out?

Computation, Meaning, and Understanding

- In other words, *I* am figuring something out, with the *help* of some computation.
- So, *I* understand what is being computed, as to *me*, the symbols are meaningful.
- However, the computation is just an automated algorithm of symbol-manipulation; *it* understands nothing, since to *it*, the symbols aren't meaningful at all.

Machines and Figuring Stuff Out



The Ambiguity of ‘Computation’

- You have been using the word ‘computation’ in two different ways:
  - Syntactic computation: the process of manipulating symbols in accordance to some algorithm (M)
  - Semantic computation: the process of figuring something (f) out with the help of a syntactic computation (c,M,d)
- So, syntactic computations can only become meaningful information-processing (semantic computation) if they are interpreted by some agent ... that is already cognitive!

The Look-Up Table Set-Up

- The animations suggest a ‘Look-Up’ Table Set-up
  - The person inside the room simply matches the input with a long list of input-output pairs
- But, this will never suffice to convince the outside Chinese speaking people that the person inside the room understands the conversation
  - There can be an infinite number of questions asked or things said.
  - Algorithms (i.e. sequences of operations, with loops and branchings) are needed to be able to possibly cope with this!
  - So, scrap paper is needed to perform the symbol manipulation operations as demanded by the algorithms.

‘Scrap Paper’??

- Searle’s original paper acknowledges that the person in the room will have to have ‘scrap paper’
- But again, in order to be convincing, there needs to be some sort of ‘record’ or ‘history’ of the conversation. I.e. you need memory!
  - Also, without memory, the same input will always lead to the exact same output.
  - And no learning could take place either.
- Well, you can use the paper as memory!
- But it is no longer ‘scrap’ paper!

Argument Against AI (and Against Computationalism)

- But now we see the true power of the Chinese Room thought experiment: given that the rules of the rule book can be arbitrarily complex, and given that the scrap paper can be used as memory, *any* computation can be captured by the Chinese Room scenario.
- Indeed, Searle uses the Chinese Room scenario to argue that Artificial Intelligence is, in principle, impossible: intelligence cannot be obtained by computations, because whatever computation you think would lead to intelligence can be implemented in a Chinese Room scenario in which no understanding, and hence no intelligence (of the relevant kind) is present.

Objections to Searle's Argument

- "There *is* Intentionality" Reply
- Person Can Learn Reply
- Brain Reply
- Reductio Ad Absurdum Reply
- Robot Reply
- System Reply

There *is* Intentionality Reply

- The person inside the room is plenty intentional … and intelligent:
  - Doesn't the person need to be intelligent in order to understand and execute the instructions in the rulebook?!
- Not a good objection: The person is intentional, sure, but the person doesn't have intentionality of the right kind: the person doesn't know what the Chinese symbols mean, and that's what's relevant.

The Person can Learn Objection

- But after doing this for a while, wouldn't the person gain some understanding of the symbols, just by observing certain patterns, and relating how maybe similar patterns would occur were the conversation in English?
- Again, not a good objection: According to computationalism, intentionality would have to be there at the very start.

The Brain Reply

- Suppose that instead of Chinese symbols, the symbols would correspond to brain states of an actual Chinese person having the same conversation, and suppose that the operations that the man in the room goes through mimic the operations that the brain of this Chinese speaking person goes through.
- Since this Chinese person understands the conversation, such understanding should now be present in the Chinese Room scenario as well.

Reductio Ad Absurdum

- Put differently: Using Searle's logic, there would be no understanding in this scenario, hence by Searle's logic, humans couldn't have any intentionality or understanding either!
- This is absurd, so Searle's argument is absurd!

Not so Fast!

- First, pointing out that an argument makes an absurd conclusion is one thing, but pointing out why or where the argument goes wrong is much more important!
- Second, this argument assumes that understanding and intentionality is produced through the simulation of a human brain.
  - So, this objection assumes the truth of functionalism (as well as materialism (and that minds come from brains)).
  - But why should that be so? Indeed, isn't Searle arguing against exactly that?

Searle's Reply to the Brain Reply

- The same reasoning still applies: in this scenario, the person in the room still has no idea of what is going on!
- Understanding of the conversation goes beyond functional operations.
- (indeed, it looks as if Searle's argument can be used to argue against much broader functionalist positions, not just the more narrow claim of computationalism).

So Where does Searle Stand?

- Can a machine think?
  - Yes, we think and we are (meat) machines
- Can an artificially constructed machine think?
  - Sure, if we can artificially construct humans
- OK, but can a digital computer think?
  - That depends on how it is implemented.
- Can a computer (or anything) think in virtue of its program alone?
  - No.

The ‘Milk’ of Intentionality

- It is not because I am the instantiation of a computer program that I am able to understand [,but] it is because I am a certain sort of organism with a certain biological (i.e. chemical and physical) structure. … Perhaps other physical and chemical processes could produce exactly these effects; perhaps Martians also have intentionality but their brains are made of different stuff. That is an empirical question, rather like the question whether photosynthesis can be done by something with a chemistry different from that of chlorophyll.

The Robot Reply

- Suppose that we add arms and legs to the room, and equip it with cameras and other sensors. Then, as a result of the symbol manipulations of the person in the room, information gets processed in such a way that the robot will behave in a perfectly intelligent way.

The Symbol Grounding Problem

- Where does our understanding of symbols come from?
- By looking them up in the dictionary?
  - I would just get more symbols!
- Intuitively, our understanding of words and symbols would come from us interacting with a physical outside world.

Searle’s Reply to the Robot Reply

- First of all, the Robot Reply does admit that there is more to cognition than symbol manipulations alone.
- Second, the person inside the robot still has no idea what is going on, there is more to understanding than mere symbol manipulation.

The System Reply

- Just because the person in the room doesn’t understand Chinese doesn’t mean that the bigger system, as made up by the man, the rulebook, and anything else that is involved in keeping up the conversation, isn’t intelligent. Hence, Searle’s argument contains a gap.

The Person as a mere Causal Facilitator

- We can point to the Robot and Brain Reply to make this gap clear: the person in the room is merely a ‘causal facilitator’, and all the other parts of the system as a whole play a crucial role as well.

Person as CPU

- Indeed, on Searle's Chinese Room way of implementating a computer, the person inside the room is … the CPU!
- But when we say that computers can be intelligent, we are not claiming that the CPU by itself is intelligent!
- Clearly, the program (rulebook) and memory (paper) are absolutely essential here!

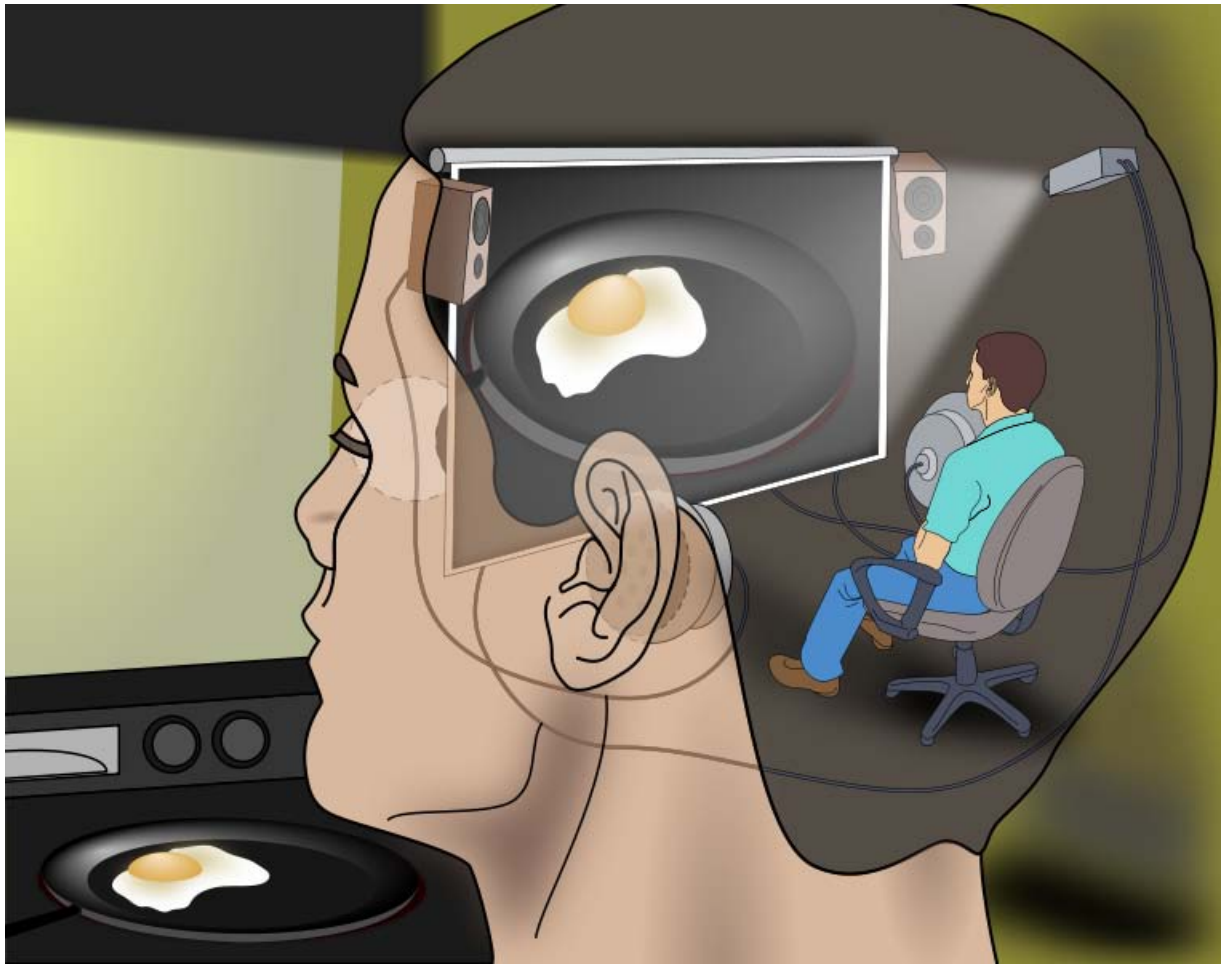
Is the Chinese Room Deceiving?

- The way that the scenario is described intuitively puts all the focus on the person inside the room … if there is any intentionality at all, it would have to be with this person.
- The rulebook, and especially the paper, get little attention.
- Also, think of the size and speed that you would need to pull this off. Imagining this ‘properly’ gives us much more the idea of a complex system.

The Homuncular Fallacy

- If we try to explain the cognitive ability of some larger system by making reference to a smaller subsystem having that very cognitive ability, we are committing a homuncular fallacy.
- We are basically saying that ‘we have cognitive ability X, because there is something inside of us (the ‘homunculus’; ‘little man inside our head’) that has cognitive ability X’
- The problem:
  - This does not explain why or how something has cognitive ability X; we still need to know how something has cognitive ability X
  - Indeed, this circular explanation leads to an infinite regress: the ‘homunculus’ would have to have its own homunculus inside it, etc.

The “Cartesian Theater”



Searle's Reply to the System Reply

- First, the only place where any kind of understanding can take place is with the person in the room, and that person doesn't understand in any of the scenarios. Adding pieces of paper, cameras, artificial limbs, etc. doesn't change anything to this lack of understanding.
- Second, the person in the room can always simply memorize all the transformations, and do the relevant transformations in his/her head. Thus, the person could even step out of the room and have a conversation with Chinese people without having any understanding regarding this conversation. And, this time there is no larger system to point to to which we can possibly attribute any understanding.