UPI2212 Reading Presentation

John R.Searle: “Minds, Brains,and Programs”

Presenters: Gab and Wying

01

Introduction to the Reading

Weak AI

The computer is a powerful tool that we can use to study the mind (1.1).

Strong AI

The computer can be programed to literally understand and have cognitive states (1.1).

Strong AI is no longer just a mere tool but is in itself the explanation for human cognition

Schank’s Program

AIM: Simulate the human ability to understand stories

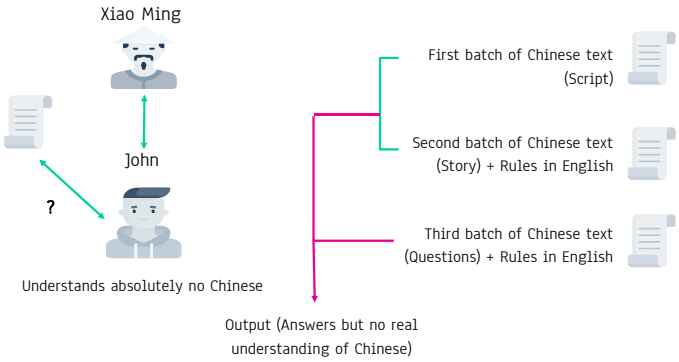
PRINCIPLE: Understanding is shown by our capacity to answer questions about the story that require inference (ideas deduced implicitly)

AI that is programed to simulate this ability can be said to:

1. Understand the story
2. Explain our ability to understand stories

02

The Chinese Room



What Searle Thinks

“Strong AI understands the stories”

Chinese Room example shows that the computer is just following a set of instructions, taking in input, processing it through a formal program and producing output without any actual understanding.

“Strong AI explains human understanding”

Strong AI does not provide sufficient conditions of understanding nor is it a necessary condition. Formal programs lack the causal properties of certain biological structures which are necessary for producing intentionality (12.1).

Formal: Involves only the outward or functional aspects of the elements that constitute the object, without the actual content

Causal: Idea of cause and effect. In context, Searle usually means that something in the brain causes intentionality

Intentionality: The ability to have mental states, i.e to think/feel **about** things external to the mind (e.g objects or things happening around us that need not be material).

03

Replying to Critics

Systems Reply @Berkeley

The individual may not understand Chinese but the whole system does. The man that does not understand is just a part of a larger system that does (5.3).

John R. Searle @AI_sucks

However, if we deconstruct the system, and internalise it into individual components, the system still wouldn't understand because the individual doesn't (5.4).

Systems Reply @Berkeley

OK wait, not so fast: Even though the man does not understand Chinese like how a native Chinese speaker would, the subsystem of the man that does the formal symbol manipulation for Chinese *really* does understand Chinese (5.4).

John R. Searle @AI_sucks

The Chinese subsystem is only a manipulation of symbols based on English rules, at best it is a Chinese subsystem as part of an English subsystem (6.1). Just because I pass the Turing test, does not mean I understand because it is already shown in the Chinese room that the man can perform as well as a native Chinese speaker without understanding Chinese (6.2).

Robot Reply @Yale

We can put a computer inside a robot! The robot can have the same sensory inputs and "act" like humans, all under the control of a computer "brain". Surely it can understand and have mental states unlike Schank's program because of its added complexities (8.1).

John R. Searle @AI_sucks

AHA! You admit that cognition is more than just formal symbol manipulation because you are adding some form of causal relationship between the robot's action and the outside world (8.2). However...

Robot Reply @Yale

oh no, here comes the rebuttal :(

John R. Searle @AI_sucks

The robot is just following more instructions programed into it. In the Chinese room example, the man is merely interpreting more input from the sensory apparatus ("perceiving") and giving output to the motor apparatus ("acting"). However, the man still understands nothing. In a way, the man is a homunculus that doesn't know what's going on. Similarly, the computer in the robot doesn't understand as well and could not have mental states (8.2).

Brain Simulator Reply @Berkeley_MIT

What if we have a program that simulates the neuron firings of a Chinese speaker. The program takes in Chinese stories and questions, processes these inputs by simulating the actual brains of native Chinese speakers processing these stories, and returns the answers in Chinese (8.3).

John R. Searle @AI_sucks

Hold on.. isn't the idea of Strong AI the notion that you don't need to know everything about the brain to know how the mind works?

Brain Simulator Reply @Berkeley_MIT

😨

John R. Searle @AI_sucks

You can't *just* model the formal structure of the brain and neurons hoping that the system will now understand. Suppose the man in the room is now operating water pipes that produce Chinese symbols. Does the man understand Chinese? No! Nor do the water pipes (8.3, 9.1).

You are missing what is crucial: the causal properties of the brain. We need to stimulate how the brain processes inputs (cause) and turns them into emotions and thoughts directed at such inputs (effect) (9.2). But is the computer capable of such simulations?

Combination Reply @Berkeley_Stanford

Let us take all three of the previous replies, combine them and build a robot that is almost exactly like us behaviorally. Furthermore, this robot works as one system instead of a computer running on a program. In such a case we'd have to ascribe intentionality to the system (9.3).

John R. Searle @AI_sucks

We are indeed inclined to ascribe intentionality to the robot because we do not know how to account for its behaviour. However, once we figure out that its behaviour is the result of a program, we would no longer say it has intentionality (9.4, 10.1 - 10.3).

Other Minds Reply @Yale

How do we know about other minds? We can only know them through behaviour. Similarly, if a computer passes all the behavioural tests we throw at it, then we can say it has cognition. Just like how we would agree that humans who passed such tests have cognition (11.2).

John R. Searle @AI_sucks

I am not even sure replying to this is worth my time, since you missed the point of my argument. I am not discussing how to determine if others have cognitive states but *what* I attribute to them when I say they have cognition. It's not simply computational outputs and their processes because these can exist without cognitive states (11.3).

Many Mansions Reply @Berkeley

Eventually, we'll be able to build AI with whatever causal properties that are essential for intentionality. So you can't say that AI cannot produce and explain cognition (11.4).

John R. Searle @AI_sucks

OK, but this trivialises strong AI by redefining it from "mental processes are computational processes over formally defined elements" to "whatever artificially produces and explains cognition" (11.5).

04

Searle's Concluding remarks

No purely formal model will ever be sufficient by itself for intentionality. Formal properties on their own do not constitute intentionality because:

1. Formal models themselves have no causal power. They **cannot cause intentionality** as the brain does. Instead, they just add another layer of structure to the machine's operation (12.2, 13.8).

2. Intentional states are beyond structure, they are **defined in terms of their content**. There can be no formal structure to a mental state (e.g a belief) because it can be represented in many ways using multiple expressions (e.g different linguistic systems) (14.2).

3. Mental states are the **product of the brain** but a program is not a product of the computer (14.3). The mind is not to the brain as the program is to the hardware.

Why do people even ascribe intentionality to AI in the first place?

1. Confusion surrounding **"information processing"** (14.7 - 15.1). The notion of "information processing" must either imply intentionality or not. If it is the former, then computers cannot be said to process information because they do not have intentionality. If it is the latter, the level of information processing in computers is not the same as that in humans.

2. **Residual behaviourism or operationalism** (15.2). Because AI operates similarly to humans, it is tempting to say it has mental states. However, as seen from the Chinese Room example, just because AI behaves in certain ways similar to humans, they do not necessarily mean it has intentionality.

3. **Residual form of dualism** (15.3 - 16.1). Many AI workers still hold a dualistic worldview, which gives rise to the claim that the human mind can be entirely captured by strong AI. Such a view is dualistic as it suggests that the mind can exist without a biological body. However, intentionality is a biological phenomenon that depends on the brain; one cannot exist without the other.

Clarification needed - Gab

"If you type into the computer "2 plus 2 equals?" it will type out "4". But it has no idea that "4" means 4 or that is means anything at all. And the point is not that it lacks some second-order information about the interpretation of its first-order symbols, but rather that its first-order symbols don't have any interpretations as far as the computer is concerned" (12.1).

What does Searle mean by "first-order symbols" and "second-order information"? Isn't the use of a script by the computer to process these first-order symbols in some way a form of interpretation (even though it might be formal)?

Clarification Needed - Wying

"[W]e find it completely natural to ascribe intentionality to members of certain other primate species such as apes and monkeys and to domestic animals such as dogs. The reasons we find it natural are, roughly, two: We can't make sense of the animal's behavior **without** the ascription of intentionality and we can see that the beasts are made of **similar stuff to ourselves**...Given the **coherence of the animal's behavior** and the assumption of the same **causal** stuff underlying it, we assume both that the animal must have mental states underlying its behavior, and that the mental states intent be produced by mechanisms made out of the stuff that is like our stuff" (10.4).

a) **What does Searle mean by "coherence of animal's behaviour", given that primate species and domestic animals would behave very differently due to species differences and environment which they live in?**

Discussion Question - Gab

"The idea is that while a person doesn't understand Chinese, the conjunction of that person and bits of paper might understand Chinese" (5.5).

From this quote, it seems that Searle believes that each element of a whole system must be able to understand before the whole system can understand. However, when we look at neurons in the brain, is it accurate to say that each neuron can understand? Each neuron is also simply reacting to chemical stimuli from its environment, yet the combination of individual non-understanding neurons is able to give rise to a brain/mind that can understand.

Is Searle valid in making such a claim that the conjunction of non-understanding parts is unable to give rise to a whole system that can understand?

Discussion Question - Wying

Searle argues that "in the literal sense the programmed computer understands what the car and the adding machine understand, namely, exactly nothing" (5.1).

Hence how does the processing of a computer program happen, because arguably, the computer will have to be able to read and understand the program, just like how the man in the Chinese room understands the English rules given to him.