

Introduction to Human Sciences (HS8.102)

Spring 2022, IIIT Hyderabad
18 Jan, Tuesday (Lecture 4)

Taught by Prof. Don D'Cruz

Thinking & Consciousness

Searle's Chinese Room

The Chinese Room thought experiment was a response to the Turing Test, with regard to the nature of thinking.

There are two different types of AI – narrow (which performs tasks in a specific domain that would typically require human intelligence) and general (which has the capacity to learn new intelligence-requiring tasks).

Searle introduced a distinction between strong and weak AI as well. Strong AI posits that a computer programmed in the right way is a real mind – it can “understand” and explain human cognition. Conversely, weak AI only considers a computer a useful tool for the study of the human mind; it helps us to formulate and test our hypotheses in a rigorous way.

He attacked strong AI, pointing out that it was ignoring the fundamental biological quality of the human mind, and that it considered the mind simply one of the infinite kinds of hardware that could simulate human intelligence. His Chinese Room thought experiment was a rebuttal to this.

Searle's argument against strong AI can be outlined as follows:

- If strong AI is true, then any system that runs the Chinese-understanding program understands Chinese.
- Searle can run the so-called Chinese-understanding program without understanding Chinese.
- Strong AI is false.

The implications of the Chinese Room experiment are that there is no formal algorithm for the mind; computer programs have only syntax while thought processes have a semantics, which syntax is not sufficient for.

Responses to Searle's Argument

There are some rebuttals to Searle's argument that can be made:

→ Given the differences in complexity between the brain (containing billions of neurons) and the room, a comparison cannot be made.

→ The *system* (consisting of the person, the room and the manual) understands Chinese, even if the part of the system that the person is does not.

→→ This is subject to the same objections; there is no way for the system to get to semantics from syntax. In principle, the manual could be internalised by the person and they could wander around and possibly even converse in Chinese without understanding it.

→ What if the system simulates the actual sequence of neuron firing, in a Chinese speaker's brain?

→→ Same issue; the system can still be contained in the operator's imagination, without them understanding Chinese.