

Name:
Group:

Patrick Ehrenbrink
Achim Stefan

29.05.2006

"Minds, Brains and Programs" - John R. Searle

Task: Does Searle's Chinese Room-Gedankenexperiment show that robots - solely by virtue of being a computer with the right sort of program - do not have any understanding or intentionality?

Introduction: In his text "Minds, Brains and Programs", John R. Searle tries to argue that it will never be possible to build the so called **strong AI** (**strong Artificial Intelligence** means an artificial intelligence which really has understanding, intentionality and consciousness, in contrast to **weak AI** which is only a powerful tool and does **not** have these attributes).

In order to do that, he invents the **Chinese Room**. A *Gedankenexperiment* in which he is locked into a room and gets chinese input, but he does not understand chinese. He also has certain rules in english. According to these rules, he gives chinese output (which he also does not understand). Due to the given rules, he will always give the correct output even if he does not know what the input or the output means.

For outsiders, it is not possible to distinguish if the person inside the room does understand chinese or not. These system would even pass the **touring**¹ test.

The Chinese Room allegorizes the function of computers, which also simply do manipulation of formal symbols. Searle uses his Chinese Room to answer six arguments of partisans of the strong AI. In this essay, I will only present three of them. On one hand because I want to keep this essay rather short, on the other hand because the remaining ones seem less relevant to answer the given task.

The Systems Reply: This claim argues, that even if the person who is locked in the room does not understand neither the input nor the output. That does not mean that the system **in the whole** does not understand, because the person is only one part of the system.

Searle replies that the locked person could **include the whole system** by learning the rules by heart and simulating the complete process in his thoughts. Due to the fact that the system would now be a part of the person, the person should be able to understand chinese, if the system would understand chinese. But the person still does not understand chinese, hence the system does not understand chinese.

Humans understand, computers do not: Searle thinks that computers have a lack of **causal powers**, they are not able to establish a semantic connection between the symbol manipulation which they perform and the real outer world. Because of that, they have no intentionality.

Humans have these causal powers because they are produced by unknown features of the brain and the interaction with the

¹ Test to determine whether a machine is intelligent or not. If a test person can not differ between answers of a machine and a real human, the machine passes the test.

Name:
Group:

Patrick Ehrenbrink
Achim Stefan

29.05.2006

environment.

There are two other claims which deal with these problems. I will present them in the following.

The Robot Reply: The Robot Reply assumes a robot with a computer brain and furthermore sensory devices. The robot would be able to perform all kinds of actions to interact with its environment, like eating, walking around and so on.

The claim says that this robot would have *genuine understanding and other mental states*.

Searle does not share this claim. The whole behaviour of the robot could still be reduced to simple symbol manipulation, the only change of the Chinese Room would be that the input would consist of sensory information and the output would be motor actions. There would be no intentionality.

The Brain-Simulator Reply: If there would be a program which does not simply process given input, but simulates the whole brain of a native Chinese speaker with all synapses and neuron firings, this program would get Chinese input and correct Chinese output. The crucial point of this claim is that if one would deny that this program would understand Chinese, one would have to deny also that a native Chinese speaker would understand Chinese.

In order to reply to this claim, Searle modifies his Chinese Room in the way that the person inside the room does not give answers to certain input, but tunes water pipes which simulate the whole brain functions according to given rules. The output would be correct and the person inside the room would again not be able to understand Chinese.

Searle says that intentionality and brain states are products of the brain and a simple simulation of the brain processes could not produce these things. Like no one would expect to get wet if a rainstorm is simulated by a computer, no one should expect that the simulation of brain processes could produce intentionality.

My opinion: I think that the question if robots could have understanding or intentionality can not be answered. Since we simply do not know how understanding works in reality.

Everything in this world follows certain physical rules and hence **could** be reduced to formal processes. Independent of our knowledge about these rules. If we would be able to simulate a whole human and the environment (means that every single quantum jump or even lower level possible physical happenings are simulated), the only thing that can follow is that also the understanding and intention is simulated. Even if this would not exactly be like our real understanding, it would still be understanding. One would be able to interact with the simulated person exactly like if the simulated person would be controlled by a real person.

If one would not regard this simulated person as an intentional being, the obstacle of strong AI would not be the formal system, but the beholder.