Is Consciousness Nothing More Than Input/Output?

The Physicalist claims that consciousness is nothing more than some physical events. For instance, in humans, the experience of pain is nothing over and above some neurons firing. If that’s right, then it should in theory be possible for us to construct a purely physical thing which is conscious! Call it artificial intelligence (A.I.) or a robot. The question is, then, COULD such a thing ever be truly conscious?

1. The Turing Test: In 1950, Alan Turing proposed that an A.I. would be “just like us” mentally if it could pass a certain test. (A nice video about it can be found here.) The test was to have a human being ask it a series of questions via something like text message. If the human couldn’t tell whether she was texting a human or a robot, then the A.I. passed the test—that is, we would judge that the robot was CONSCIOUS.

Thus, the proposal here is that consciousness is present wherever there is a certain kind of behavior. (Call this proposal ‘behaviorism’).

The scientists tell us that such A.I.’s will someday be a reality (and maybe they already ARE! See here). But, some philosophers have questioned Turing’s reasoning.

2. The Chinese Room: Imagine the following scenario, described by John Searle (1971):

The Chinese Room  You know absolutely nothing about the Chinese language. You do not speak a word of it. Some scientists stick you in a room all by yourself and give you a giant 3-ring binder. In the binder are thousands of pages with 2 columns. In both columns are weird strings of symbols. The scientists tell you that they will be slipping pieces of paper under the door. These papers will have weird strings of symbols on them too. When you receive these slips of paper, you are to consult your 3-ring binder and find that string of symbols somewhere in the left-hand column. You are told that you must then copy down onto the piece of paper whatever corresponding string of symbols is written in the right-hand column, and slip the paper back under the door. (Videos here and here.)

In the story above, it turns out that the “weird symbols” are Chinese characters. The strings of symbols in the left-hand column of the binder are simply sentences in the Chinese language, and the symbols in the right-hand column are merely appropriate responses that one might give (in Chinese) if someone communicated that sentence in the left-hand column to them. Thus, in this way, the person in “The Chinese Room” is able to “communicate” with someone on the other side of the door in Chinese.
But, here, Searle notices a problem with Turing’s test for the presence of a mind: The person in The Chinese Room WOULD be able to pass a version of Turing’s test. That is, a Chinese speaker outside of The Chinese Room would be convinced that the person inside of the room was fluent in Chinese. And yet, it is clear in the example that the person inside of The Chinese Room doesn’t understand Chinese. They don’t speak a single word of it! And they don’t understand one bit of what they are doing.

Against A.I.: Many Physicalists claim that the human brain is no different than a really advanced computer. For, our cognitive processes are nothing more than a series of inputs and outputs. One day, it is said, computers will be able to mimic these inputs and outputs and become conscious themselves. Once they begin to process information in a certain way (i.e., the way WE do), computers will begin to have conscious experiences.

Searle’s story challenges this conclusion. Searle’s position is that, even if a computer could IMITATE a mind (by processing information just as we do; i.e., by mastering our language, claiming to have certain thoughts and beliefs, etc.), it would be merely that: An imitation. An artificial intelligence would have no more conscious understanding than the person in The Chinese Room. In order to have CONSCIOUSNESS, something more than a mere input-output pattern is needed. (To illustrate, it may help to simply interact with Mitsuku or Cleverbot for a while, and consider whether all the bot needs in order to become conscious is to become more convincing. What do you think?)

An argument for this conclusion would look something like the following:

1. Physicalist Hypothesis: Whenever an individual behaves as if it is conscious, or aware—that is, whenever the system has the appropriate inputs and outputs—then that thing is conscious/aware.
2. The man in The Chinese Room has the appropriate inputs and outputs.
3. The man in The Chinese Room is not conscious/aware (of what he is doing).
4. Therefore, the physicalist hypothesis from premise 1 is false.

Searle has not proved that consciousness/understanding is NOT reducible to physical processes—but his example IS somewhat persuasive in convincing the reader that consciousness cannot be reduced to THAT sort of physical process; i.e., consciousness is not MERELY a kind of behavior (i.e., it is not reducible to appropriate inputs & outputs).

Objection: Some suggest that premise 2 is false, because Searle’s example is too simple. Imagine that the person in the Chinese room gains access to the outside world (e.g., via cameras and microphones), and then begins to connect the Chinese symbols with the things she sees on her monitors. She even learns to form NEW, increasingly complex strings of symbols that were not in the 3-ring binder.
In short, imagine that:

1. she connects the abstract symbols with things in the tangible world,
2. she begins to learn on her own, and
3. the original output that she creates on her own becomes very complex.

Would you say that the person in the Chinese room was still merely IMITATING the Chinese language? Or would we conclude that she now has TRUE UNDERSTANDING?

Alternatively, perhaps premise 3 is false. Perhaps the SYSTEM as a whole DOES have understanding? We would not expect the PERSON in the room to have understanding, because she is only a COMPONENT of a system that understands, just as we would not expect a set of neurons to have understanding, because they are only a COMPONENT of a system that understands. *(What do you think?)*

### 3. The China Brain

Perhaps certain kinds of behavior are a pretty good INDICATION that consciousness is present, but that’s not what consciousness IS. Instead, perhaps consciousness occurs whenever a physical system FUNCTIONS in a certain kind of way (namely, in whatever way our brains function). But, imagine a case from Ned Block, 1978:

**China Brain** The entire population of China (1.4 billion) are asked to engage in an experiment. The motions of a giant robot are dictated by a network of radio signals. Every single person is given a walkie-talkie and a series of instructions. The range of the walkie-talkies only extends to those people nearest to them. The instructions are things like, “When the person in front of you signals you with a beep, signal the person behind you with a beep,” etc. (Or something.) The point of the experiment is to perfectly simulate the sorts of signals that NEURONS give to each other in your brain. So, the population of China perfectly copies the functionality of a human brain, and the robot body that they are controlling performs the actions that its “brain” tells it to.

*[Note: Scientists have since discovered that there are more than ONE billion neurons in the brain—there are more like 86 billion. But, in any case, rats are clearly conscious, and they only have 200 million neurons—the population of China has 7 times that many people!]*

Anyway, if the hypothesis is correct that consciousness occurs whenever a system FUNCTIONS in a certain way, then the country of China would BE CONSCIOUS in the scenario above. Some sort of conscious mind would arise out of the Chinese people!
Block expects you to have the intuition that a mind would NOT arise in his example. And, since it is absurd to think that the China Brain would give rise to a conscious mind, it must not be the case that consciousness is reducible to a system of signals and impulses following a certain pattern. His argument runs as follows:

1. Physicalist Hypothesis: Whenever a physical system functions in a certain way (namely, like the human brain), then that system will be conscious.
2. The population in the China Brain case does function in the appropriate way.
3. But, the population in the China Brain case (or, alternatively, the robot that they are controlling) is clearly not conscious.
4. Therefore, the physicalist hypothesis from premise 1 is false.

Objection: But, is Block right? Some suggest that premise 3 is false, and that a mind WOULD arise. This is not that strange. Consider the ways in which collections of bees or slime mold (also here) make seemingly intelligent decisions, for example. Is it so terribly counter-intuitive to think that a sort of “hive mind” arises in those cases? Some suggest that the individual bees act as individual neurons in such a way that the entire hive becomes one conscious being. Perhaps this same thing would occur in China.

Block’s intuition likely results from an issue regarding scope. To the Chinese people it would not SEEM like there was a China Mind. But, that would be because they would be acting at the level of the neuron. Similarly, at the level of the neuron, it would not SEEM like there was a conscious mind. If we looked at one little ant, it would not SEEM like there was a hive mind. But, that is just because the hive mind is not at the level of the ant. The conscious human mind is not at the level of the neuron. If we ARE the neurons in one giant hive mind, we will never be able to get a glimpse of the mind itself. The conscious experiences would be occurring at an entirely different level.

Reply: Sure, hives might ACT like a single being, just as an artificial body connected to the China Brain might ACT like a human. But, would it BE CONSCIOUS? Is there some feeling or sensation that it is like TO BE a hive? Is there some qualitative feel that is what it is like TO BE the China Brain?