Hating John Searle

March 14, 2007

Original link

I think John Searle might be my favorite living philosopher. But when I tell my friends this, they recoil in horror. "That bastard?" one friend cries. "He tripled my rent!" "Oh geebus," another cries. "The Chinese Room argument is *awful*."

I will not profess to have an opinion on whether the fourteenth amendment requires the City of Berkeley to provide a rational basis for only allowing its landlords to raise rents at forty percent of the increase in the consumer price index, but I must admit I fail to see what the question has to do with Searle as a philosopher.

The other complaint seems less *ad hominem*, but I think is fundamentally similar. Reading Searle's published books, it's striking how little space the Chinese Room Argument takes up. Indeed, his book on the subject of consciousness — *The Rediscovery of the Mind* — gives it little more than a paragraph and notes that his more recent argument against functionalism is far more powerful.

Nonetheless, I will defend the Chinese Room Argument. The basic idea, for those who aren't familiar with it, is this: imagine yourself being placed in a room and given instructions on how to convert one set of Chinese symbols to another. To outsiders, if the instructions are good enough, it will seem as if you understand Chinese. But you do not consciously understand Chinese — you are simply following instructions. Thus, no computer can ever consciously understand Chinese, because no computer does more than what you're doing — it's simply following a set of instructions. (Indeed, being unconscious, it's doing far less.)

The Chinese Room Argument works mainly as a forcing maneuver. There are only two ways out of it: you can either claim that no one is conscious or that everything is conscious. If you claim that no one is conscious, then there is no problem. Sure, the man doesn't consciously understand Chinese, but he doesn't consciously understand English either. However, I don't think anyone can take this position with a straight face. (Even Daniel Dennett is embarrassed to admit it in public.)

The alternative is to say that while perhaps the man doesn't consciously understand Chinese, the *room* does. (This is functionalism.) I think it's pretty patently absurd, but Searle provides a convincing refutation. Functionalists argue that information processes lead to consciousness. Running a certain computer program, whether on a PC or by a man with a book or by beer cups and ping pong balls, will cause that program to be conscious. Searle points out that this is impossible; information processes can't cause consciousness because

1

they're not brute facts. We (conscious humans) look at something and decide to *interpret* it as an information process; but such processes don't exist in the world and thus can't have causal powers.

Despite the obvious weakness of the arguments, why do so many of my friends continue to believe in functionalism? The first thing to notice is that most of my friends are computer programmers. There's something about computer programming that gets you thinking that the brain is nothing more than a special kind of program.

But once you do that, you're stuck. Because one property of computer programs is that they can run on any sort of hardware. The same program can run on your Mac or a PC or a series of gears and pulleys. Which means it must be the program that's important; the hardware can't be relevant. Which is patently absurd.

I used to think that part of the reason my friends believed this was because they had no good alternatives. But I've since explained to them Searle's alternative — consciousness is a natural phenomenon which developed through evolutionary processes and is caused by the actions of the brain in the same way solidity is caused by the actions of atoms — and it hasn't caused them to abandon their position one bit.

So I tried a different tack. I asked them what they thought was wrong with Searle's position. And the answer always seems to come down to a confusion between ontology and epistemology. Ontology is a fancy word for the facts of the matter — what actually exists out in the world. And epistemology is the world for the way we know about it. Unless you subscribe to a bizarre philosophical theory, things in the world exist irrespective of whether we know them or not. Behind the TV game show door, there either is a car or there isn't, even if no one can see in to tell which one is the case. Furthermore, things continue to exist even if we can't even know them in principle. There appears to be no way for me to ever tell what it feels like for you to taste an orange; nonetheless, there is indeed something that it feels like for you.

My programmer friends' argument always ends up coming down to this: if a computer program acted conscious, if it plaintively insisted that it was conscious, if it acted in all respects like the conscious people we know in the real world, then it must be conscious. How could we possibly tell if it was not? In short, they believe in the Turing Test as a test for consciousness — anything that acts smart enough to make us think it's conscious must be conscious.

This was the position Ned Block was trying to refute when he postulated a computer program known as Blockhead. Blockhead is a very simple (although very large) computer program. It simply contains a list of all possible thirty minute conversations. When you say something, Blockhead looks it up in the

list, and says whatever the list says it's supposed to say next. (Obviously such a list would be unreasonably long in practice, perhaps even when heavily compressed, but let us play along theoretically.)

Having a conversation with Blockhead would be just like having a conversation with a real person. But nobody could seriously claim the program was conscious, right? Well, in fact they do.

One wonders whether these people think their cell phones are conscious. After all, talking to a properly-enabled cell phone works just like talking to a properly-enabled person! (I asked one friend this and his response was that the whole system containing the cell phone and the wires and the person on the other end was conscious.)

The point is that we don't assign consciousness purely based on behavior. Blockhead acts like it's conscious and a completely paralyzed person acts like they're not, yet we all know that the first isn't conscious and the second is. Instead, we assign consciousness based on causes. We know dogs are conscious because we know they have brains that are very much like ours which cause behavior very much like ours. We don't make that judgment based on behavior alone.

Criticisms aside, what is the positive argument for John Searle? First, he has done important work in a wide variety of fields. As far as I can tell, he began following up the works of his teachers (like J. L. Austin) on the topic of speech acts, which he generalized to the subject of intentionality, which he solved by saying it was a property of conscious beings, which led him to develop a theory of consciousness. Second, all of his points seem quite reasonable to me and (with a few exceptions) I agree with them. Third, he writes extremely clearly and entertainingly and for a popular audience.

These three seem like a fairly low bar — they're about what I would expect from myself were I a philosopher — but its shocking how few prominent philosophers seem to meet them. Daniel Dennett is a dreadfully prolix writer and is insane. Thomas Nagel comes close but is a fairly committed dualist. Hilary Putnam doesn't write for a popular audience. Peter Singer doesn't seem to develop any actual theories. So I can't think of any. Can you? Suggestions appreciated in the comments.