

A: I am here supporting the systems reply to John Searle's Chinese Room thought experiment. In the case of the Chinese Room, it is clear that no single element of the system understands Chinese: the English-speaking operator does not, by definition; the slips of paper cannot, because a slip of paper is incapable of understanding; the book of rules does not because it is entirely inflexible. However, the system of all these elements together does understand Chinese. Searle's supporters immediately jump to calling this conclusion absurd: however you choose to realize the program, they say, whether it be paper, pipes, logic gates, or English speakers, the individual elements of the system cannot understand Chinese, and therefore the system cannot.

I agree with the first part of this argument; I said as much just a moment ago. The conclusion, though, "therefore the system cannot understand," is erroneous. Understanding is an emergent property, and this argument essentially claims that properties cannot be emergent. This is intellectual arrogance. There is no question that a native Chinese speaker understands Chinese, and we hopefully all also agree that his ability arises from the function of his brain. Brains are made up of neurons, and neurons certainly cannot understand. They are passive; they simply *do*, not *act*. They have no intentionality by themselves. The gaps in our understanding of the human brain include how the mind arises from the activities of a system of passive structures. We know, though, that a brain made of neurons can understand. Who's to say the same couldn't happen with pipes or programs? Just because we do not know how an emergent property such as understanding arises does not mean that it cannot; in fact, we observe that it does every day.

Furthermore, the "absurdity" of the system of the Chinese Room understanding Chinese shows a bias against the possibility of non-biological understanding. It does not belong in a serious, mature discussion. The fact that the only understanding we have observed so far has arisen from biological systems does not intrinsically imply that understanding is necessarily unique

to biological systems, that it is impossible in non-biological ones. To show that that is true, it is necessary to prove that there is a unique property of biological systems that allows them to have understanding and that all other systems lack. Searle's argument does not do this.

B: If the system as a whole can understand Chinese, then, have the human operator of the room, Searle, internalize it. He memorizes the book of rules, does all calculations on his head, stores the intermediate information entirely within his mind, and simply stands in the room by himself, looking at slips of paper that come through one slot, performing formal symbol manipulations on what he sees, and spitting back out more slips of paper. Every step in this process is meaningless to him, and he has no further understanding of Chinese.

The proponents of the systems argument then maintain that there now exists a subsystem within Searle's mind that does understand Chinese. However, it is again simple to see that this is not true. Searle, a being who does have understanding of English, follows formal rules in his head. He still does not understand Chinese. Not only are the English-understanding and Chinese-“understanding” subsystems separate, they are entirely dissimilar. One system (the Chinese one) simply manipulates, while the other understands.

I agree that neurons are relatively simple structures. In theory, a computer could be programmed to replicate all the activities of a brain-full of neurons. Say the brain that this computer replicates is that of a native Chinese speaker. This computer would not understand Chinese any more than the Chinese Room. The entire program could be internalized within Searle's mind, but so long as it is only replicating the firings of the neurons, it is ignoring what is truly important about the brain, its ability to produce intentional states.

A: I'll only hold up a mirror to your argument. As Searle himself originally asked in response to the brain-simulation argument, "where is the understanding in this system?" Consider not a simulated brain, but an actual one. We attribute the ability to understand Chinese to the real brain of a Chinese speaker. However, this shows the fruitlessness of asking, "where is the understanding?" Our knowledge of the brain is not thorough enough now to pinpoint where understanding comes from, so there are no grounds to insist that a system such as a communicating computer, in which we cannot pinpoint any particular element that would give rise to understanding, cannot understand.

Let me clarify the issue of the English and Chinese subsystems. A major error that shows up in your understanding of the situation is the idea that the two subsystems operate on the same level. This is not true, though: Searle's consciousness only has access to his understanding of English.

English understanding exists as a part of Searle's own mind. He doesn't have to consider how his understanding arises from the structure of his brain, because his consciousness, that which would consider this, is operating at the same level as his English understanding; both are results of the activity of his brain. The Chinese understanding, however, arises from the structure Searle has memorized. In this case, his consciousness can understand each element of the system, the papers and rules, but the behavior and reality of Chinese understanding are occurring beyond his understanding. Just as we do not yet know how thoughts arise from brains, Searle cannot understand how his apparent understanding of Chinese would arise.

Saying that Searle does not himself understand Chinese is as meaningless as saying that the body of physical laws does not understand English. Physical laws dictate every activity in Searle's brain that lead to his understanding of English, intentionality and causality included. There is nothing else present in the brain that could give rise to these properties, so even though we do not

know how, we must concede that the formal physical activities of the neurons themselves, in a system, somehow do create understanding. Even if understanding could be attributed to physics, its level of interaction with Searle's brain would be the very lowest, and so it would, if it were a disciple of Searle's Chinese Room argument, claim that Searle could not truly understand English.

Furthermore, the personality and attitude of the Chinese-speaking subsystem would be different from those of Searle himself. As a symbol manipulator, Searle's emotions never come into play; there is no room for interpretation in the rule book. In the same way that Searle's own emotions arise, somehow, from the structure and functioning of his brain, the virtual Chinese speaker's emotions arise from the structure and functioning of the hardware and software that Searle has internalized. In this way, the two are essentially separate minds within the same body, two distinct people. It seems strange, certainly, that this would be the case, but, as mentioned above, to engage in a mature discussion we must accept the possibility of strange occurrences and rely only on logic and reasoning to disprove them, rather than gut reactions. Also, it is only strange because of the absurd premise on which it relies: the internalization by a single person of the entire structure of a complex computer and its program. This is beyond the capacity of any human's memory.

B: Your language itself indicates that yours is a behaviorist argument. This behaviorist basis for your argument relies on the assumption that the Turing test is sufficient proof for intelligence, but that is at issue. You are simply pointing and waving; the Chinese Room argument is itself a contestation of the validity of the Turing test on this exact point. Behavior itself is not sufficient proof for understanding or intelligence. We may be fooled by any sufficiently complex program, whether we interact with it in the limited context of written communication or in the real world, but once we see that its behaviors arise from an instantiation of a formal program, we know

it is not truly intelligent. We can reliably grant intentionality to fellow humans or other animals because we see that they are made of the same stuff as we are, and we know that we ourselves are intentional.

To get more basic, the problem is that any program in a Chinese Room-like computer is formal, but intentional states are not. For the Chinese Room, the symbols transferred as input and output only have meaning to the Chinese speakers providing and interpreting them. A stomach's input and output could be represented as information in the same way, and so by your argument, we can grant cognition to it. To the stomach, though, the input and output are completely meaningless just as the Chinese characters are meaningless to Searle; the people outside the system are the ones assigning the meaning. To the stomach itself, or the machine itself, the symbols are not truly symbols, for they have no meaning. This is the true reason why assigning intentionality and understanding to a stomach, or any other symbol manipulator, such as the Chinese Room, is erroneous.

A: A quick clarification is necessary here. In Searle's original Chinese Room argument, he claims for his opponents the view that "mind is to brain as program is to hardware." He insists this is not true, and indeed, it is not. Instead, mind is to brain as virtual mind is to the conjunction of programs and hardware. The mind is a nebulous, non-physical entity arising from the brain, characterized by understanding and consciousness, and identified by the behaviors that follow from its activities. Likewise, programs and hardware, together, give rise to a similar entity that does not yet have a simple name. The right programs for the right hardware could give rise to a virtual mind, an entity also characterized by understanding and consciousness and identified by behavior.

The argument, then, that programs are formal while intentional states are not misses the point. No one is claiming that the mind and its states are analogous to programs. This situation returns to the same simple issue: we observe that mental states can arise from a formal system of neurons and chemicals. Why not from a different formal system? It is the hallmark of computer science that different complex behaviors, activities, or functions can be realized in infinitely many different systems, and we have no indication that the mind is any different.

What is wrong with behaviorism? Behavior is all we ever truly have to judge intentionality in each other. The important factor is consistency. Chinese observers believe that the Chinese Room understands their language for the same reason that English observers believe that Searle himself understands theirs. This reason is that the meanings, which are set, of the symbols, be they Chinese or English, of all the inputs and outputs in these cases are consistent. A stomach would not be consistent, so we would not conclude it was cognizant. If we assign inputs and outputs such that food and food products represent words and ideas, the outputs will not be consistent in anything nearly as complex as language. We define "correct" functioning in terms of its inputs and outputs; if the concept that an output represents is consistent with the consequences of the action of a conceptual function on the concept represented by the input, the particular instance with this set of input and output is functioning correctly. If the functioning of an entity is always correct, that is, its behavior is always consistent with the conceptual function to which it is supposed to correspond, then we say that it truly has this function. A stomach, of course, cannot then have intentionality, but if a computer is always consistent with intentionality, then it is intentional.

Your argument that my stance is behavioristic, a negatively-connoted buzz word, is perfectly matched with my own belief that your stance is animistic. You repeatedly claim the same thing in different ways; the great hole in your reasoning is most apparent in Searle's own *Minds, Brains, and Programs*: "I am a certain sort of organism with a certain biological structure, and this

structure under certain conditions is causally capable of producing... intentional phenomena. And part of the present argument is that only something that had those causal powers could have that intentionality.” This argument is so circular it is difficult to create a sensible response to it. Essentially, Searle claims that something has intentionality because we see it has intentionality, but this somehow relies on a magical aspect of biology that has the exclusive ability to create intentionality. Nowhere is this aspect identified. I again maintain, then, that we must base our assignment of intentionality only on the reliable grounds we use to identify it in each other: behavior. Since the only thing we know of that gives rise to intentional states is the human brain, and its entire structure, its entire being, is formal, we must accept the possibility of other formal systems, such as computers and their programs creating true intentionality.