Why Functionalism Works by Forrest Cameranesi

The central thesis of the theory called Functionalism is that what it is to have a mind is simply to perform the proper function of a mind, regardless of how that function is instantiated physically (or perhaps even nonphysically, if such a notion is coherent). That is to say that if some system responds to the same kinds of inputs as something which we acknowledge as having a mind – a person – and in turn gives the same kinds of outputs, then it too has a mind. All a mind is is the right kind of throughput. Something need only play the correct causal role, giving the right kinds of responses to the right kinds of stimuli, in order to instantiate a mind.

This is not to say that everything which plays any functional role is a mind. Rather, it is saying that to have a mind is to play the right kind of functional role. Research would still need to be done to determine what sorts of inputs need to be handled and what sorts of outputs need to be given for something to qualify as a mind. For example, to be a soda machine is simply to play a functional role such that when a sufficient amount of money is placed into the coin slot of the machine, and the appropriate button pushed, that the right soda and any excess money (change) be dispensed out of the machine. It does not matter what goes on inside the machine, whether there is a mechanical system of gears and levels, a digital computer measuring the inputs and activating the right motors to give outputs, or a person sitting in the box counting money, watching lights light up, and pushing change and sodas out the right holes. So long as the function described above is the same, the system is a soda machine. But just because they play some functional role does not mean that all soda machines have minds. To be a mind is to play a different functional role that it is to be a soda machine.

A description of a system of inputs and outputs is called a “machine table”. Above I have described the machine table which must be instantiated to be a soda machine. But to describe the machine table necessary to be a mind may be significantly more difficult, and is fraught with

problems regarding how fine-grained we want to be. For instance, even if it were feasible for me to give a complete fine-grained description of the machine table for my own mind – detailing all of the complex responses I have to every conceivable input – that would not be the machine table for a mind in general. Other people instantiate slightly or even significantly different machine tables and yet we still say that they have minds. A hypothetical alien could give significantly different responses to the same inputs – say having an aversive pain response to the sight of the color green – and yet be qualitatively similar enough that we would still like to say that it has a mind, although a very different mind from that of a human.

Having given this topic considerable thought already, having always seen functionalism as an obvious self-evident truth, I might suggest something like what follows as a high-level general description of the kinds of functions something must play in order to be a mind. A mind in the broadest sense must simply be cognitive – able to take data, either in the forms of sensory impressions or linguistic symbols, and process it and give a response of some kind. There are two distinct kinds of cognition. One is sentience, that is, it must be able to make observations and receive sensory data from its environment and then react to those sensations in some way. All animals have at least this much of a mind. The other piece is intelligence, by which I mean the ability to perform logical operations and symbol manipulation, including the parsing of formal propositions and the formation of theories and explanations based on those propositions, and logical extrapolation of new propositions from them. Some computer programs today can do this, and may be another kind of "lesser" mind. But to have a mind in the greater, more narrow sense, as a human has a mind, requires both sentience and intelligence, and a third component – self-awareness or consciousness, a reflexivity of cognition (both sentient and intelligent), or in other words, the ability to think about what the thinker is thinking.

This still is not a full description of what is necessary to be a person, for people have volitional functions to them as well, which also have sentient components (emotions and the

ability to physically act), intelligent components (practical reasoning in the formation of intentions or 'maxims', and the ability to come up with declarative 'ought' statements), and reflexivity (willpower, the desire to change what one desires). These volitional functions I would consider to make up a person's "soul", in the Aristotelian sense of an "animating principle", rather than being a part of the "mind"; but for those who consider volitional functions necessary for something to be a mind, I acknowledge at least that they are necessary for something to be a person (though they may be found partly in animals, which have sentient volition, or robots, which have intelligent cognition.), and it seems we simply differ on our choice of terminology.

So as you can see, to describe the machine table necessary for something to be a mind may be a daunting task, and my attempt above is certainly fraught with some flaws of which I am at present unaware, but it seems that it is theoretically possible. If this is so, and we assume a strictly materialist universe (for which an argument can be made, but not here) such that anything instantiating this machine table is a physical thing, then any mental state simply is some physical state. It is important to understand that states may also be the outputs of functions. To understand states in terms of physical, multiple-realizable systems, it may be useful to turn back to the more simple soda machine example. Let us assume our soda machine dispenses only one brand of soda, and that it does so as soon as it has received at least 95¢. Let us also assume, for simplicity of this example, that this machine accepts only quarters. When we first approach the machine, it is in State 1, where it's function is to move to State 2 upon the receipt of one quarter. In State 2, it's function is to move to State 3 upon the receipt of one quarter. In State 3, it's function is to move to State 4 upon the receipt of a quarter. In State 4, it's function, upon the receipt of a fourth quarter, is to dispense one soda and one nickel, and move to State 1 again. Not until the machine is in State 4 is it in a position to output any actual behavior; prior to that all outputs were simply state changes.

By this understanding it is easy to see human mental states as similar functional states.

Let us take anger for an example. Sitting here now, I am not at all angry; I am in a rather neutral state of mind, lets call it Mad-0. A number of possible inputs could get me slightly upset, moving me to the state of Mad-1. Repeat of those inputs could move me to Mad-2, as could have some harsher inputs while I was in Mad-0. Further repeat of those annoying, obnoxious, frustrating inputs could move me to Mad-3, and again, different more severe inputs could have moved me to Mad-3 directly from Mad-0 or Mad-1. Finally while in Mad-3, a state where I am already very angry, those same mildly annoying, obnoxious, or frustrating inputs which earlier simply made me slightly upset (moving me to the state of Mad-1) could cause me to lash out and scream and want to hit someone – that state of wanting to hit someone itself being a move to Mad-4. If we are assuming a strictly materialist universe (for which, again, an argument can be made, but there is not room in this essay), and my mind is instantiated in the neurochemical system of my brain, then those mental states of graduated anger simply are physical states of my brain.

But several counterexamples have been raised to this functionalist account of the mind, purporting to show a system which is functionally identical to a person, and yet knowing how that function is attained, we the listeners are supposed to intuitively find the attribution of mental states to that system absurd. Perhaps the most popular of these counterexamples is John Searle's "Chinese Room" thought experiment. In this thought experiment, we are asked to imagine an experiment with a group of philosophers and scientists, some of whom speak Chinese, feeding written Chinese stories and questions about those stories into two closed-off rooms. In one room is a native Chinese speaker, who can read and understand the stories and thus easily answer the questions about them. In the other room is John Searle himself, who does not speak a word of Chinese, but who has a vast set of instruction manuals telling him what the appropriate responses to the questions fed into his room are, even including the retrieval of key terms from the stories fed in earlier so that he can answer questions about them. But in all this, Searle has no idea what the symbols he manipulates mean, either in terms of English words or in

terms of experienced sensations, and so he still does not understand Chinese. Thus this thought experiment is meant to show that two systems may be functionally identical and yet still, one does not have a mind, or at least that their mental states differ.

One response to this is to point out that it is not supposed to be Searle who understands Chinese, but rather him and all his instruction books together which understand Chinese. However, to this objection I must actually side with Searle: even if his thought-experiment self had perfectly memorized the rule books and so could respond to the questions about the stories just as quickly and accurately as the native Chinese speaker, without the aid of any rule books, still Searle does not understand Chinese. Another response is that a native Chinese speaker can actually understand spoken Chinese and could speak with the Chinese-speaking scientists outside the room orally, while Searle could not. To this I must again side with Searle, for even if his thought-experiment self had memorized not only the proper way to manipulate the Chinese symbols but also how to pronounce them, he has gained no greater understanding of Chinese by virtue of being able to manipulate spoken phonemes as well as written ideograms.

But I must raise my own objection to this thought experiment. Searle intended to prove by this that syntax alone is not enough for semantics – that being able to manipulate symbols on a page or even spoken phonemes properly is not enough to constitute understanding – and thus that machines cannot be made to think despite any appearances to the contrary, implying that functionalism is false. I do agree with him that syntax is not sufficient for semantics, but I do not think that this disproves functionalism, for this thought-experiment is by Searle's own admission comparing only a limited range of functionality, namely intelligence.

The imaginary Searle, with his instruction books, is able to receive propositions and parse them syntactically, perhaps even to do logical operations on them, and given two statements in Chinese such as <*all birds have wings*> and <*all ducks are birds*> be able to respond correctly <*yes*> to the question <*do ducks have wings?*>. Even reflexivity may be possible here, for Searle

could respond not only to questions like <*what did you just sa?y*> or <*what did I just ask?*>, but also questions like <*why must it be true that ducks have wings?*> (<*because they are birds and all birds have wings*>), or even <*why must it be false that ducks have wings?*> (<*it is not false that ducks have wings*>). But none of this has anything to do with the understanding of what <*bird*>, <*duck*> or <*wing*> mean.

What is lacking here is the other component of cognition which I earlier mentioned: sentience. The native Chinese speaker in the other room could respond to a photograph of a duck on a lake with a question attached <*what kind of bird is on the lake?*>, with the answer <*a duck*>. Searle could not. That is a very important functional difference. Searle, even with his books memorized, could not leave the room and walk about and identify things in the room in Chinese. The native Chinese speaker could. Since these systems are functionally different, Searle has not refuted functionalism. He has merely stated that intelligence alone does not connote a complete mind, with which I agree. But this does not mean that machines could not be made to think.

If Searle's instruction books had also included an assortment of photos linked to each Chinese term, such that <*lake*> and <*bird*> and <*duck*> are associated with certain kinds of images, then he could respond to the photographic example above. If the book had thorough associations of all Chinese terms to the right kind of sensory impressions, as well as how those terms relate to each other, then the book would simply be an instruction book on how to learn Chinese. If Searle had memorized that, he would have simply learned Chinese. Consequently, if any system, even just a robot with sensors attached and a good pattern-recognition algorithm in it's program, was capable of executing the cognitive functions detailed in this picture-book, and describing photographs and identifying objects in the world, I don't see how anyone could justify saying that that robot did not understand Chinese as well as any native speaker did. The robot and the Chinese speaker are functionally identical, at least in the domain of cognitive (non-volitional) functions, and so they are mentally identical. Functionalism still stands.