Res Cogitans

Volume 4 | Issue 1 Article 5

6-19-2013

Machine Functionalism: Brains as Computing Machines

Yolanda Walker
Humboldt State University

Machine Functionalism: Brains as Computing Machines

Yolanda Walker
Humboldt State University

Published online: 19 June 2013

Abstract

Machine functionalism, or, the computational theory of mind, states that the inner workings of the brain are akin to the information processing of a computer. There are numerous faults with this view. Not only are computers inaccurate models for brain states, but also consciousness--as understood as generating appropriate (behavior) outputs to corresponding inputs--can't be generated through mechanical means.

In everyday language, the functions of the brain are commonly described using mecanical terminology. For instance a friend might say that "the wheels are turning slowly" in the morning before her first cup of coffee, or after hours of mental labor, she's starting to "run out of steam." How we apply mechanical based metaphors to the brain demonstrate how we think about it--as a machine. To be more specific, we tend to think of the brain as type of organic computer.

Likewise, one of the most common themes in science fiction is that of the self-aware computer with individual personality traits. For instance, C-3po and R2-D2 in Star Wars, Marvin the 'paranoid android' from *Hitchhiker's Guide to the Galaxy*, and Data from Star Trek: The Next Generation. These androids/robots have human characteristics and we might say they have consciousness. But these are fictional machines with emotions and thoughts modeled after humans rather than based in our current technological capabilities. It raises the question: At what point do we say that a mess of wiring and electricity--that collectively performs human-like actions-- acquires consciousness? What constitutes consciousness, and can a computer program produce it? Machine functionalism, (a branch of functionalism) theorizes that computers are not just a useful metaphor is describing brain functions and mental states; brains *are* computational machines. Machine functionalism/ computational theory of mind states that the brain is a computer, but how viable is this theory?

Machine functionalism attempts to solve problems that plagued preceding theories of mind (such as physicalism and behaviorism), but it runs into its own issues concerning: dispositional and occurent states, determining if two individuals are in the same brain

state, the problem of inverted qualia, and the finiteness of machine tables. Finally, the understanding gap between syntax and semantics demonstrates that symbol manipulation is not an adequate requirement for consciousness.

Intuitions Guiding Functionalism

What is a brain state? What constitutes consciousness? How are the body and mind related? Starting with Descartes, modern philosophers have tried to tackle these questions. Most philosophers no longer take Cartesian dualism seriously as an adequate theory of mind. While it supports religious beliefs and the idea of a 'soul' (that is related to but not physically bound to the body) most modern philosophers think this is improbable. There's no adequate explanation for how mind can affect matter. Physicalist theories claim that pains and other states are simply their physical manifestations (such as pain=c-fiber firing) fall short as well. By classifying pain exclusively as C-fiber firings, it excludes other organisms that have a different biological makeup than humans but in all other respects appear to feel pain. A good theory of mind ought to have *multiple relizability*, or, recognition that similar mental states are realizable in different beings. Biological similarity to *homo sapiens* is not a requirement in determining if a being has mental states. Logical behaviorism falls short as well. A behavioral disposition to act a certain way is in no way synonymous with the actual brain state. For instance, an individual in pain who lives in a society that has conditioned people to control their feelings, it likely to suppress the desire to say "ouch." Someone in a more expressive society who experiences pain will definitely have some verbal outburst. Functionalism attempts to overcome these problems with physicalist and behaviorist theories.

Functionalism

Functionalists believe that, while we might not have distinct knowledge about what a mental state *is*, we can at least describe it according to what function it has. We don't define 'eye' by listing how many cells it is composed of, where it is located on the body, or saying that it has a pupil, an iris, etc; we consider something an 'eye' if it is an organ that allows a being to see. Eyes developed concurrently through evolution in different organisms. The eyes of a blue whale are different from the eyes of a pelican, but they are both eyes because they serve the same function. A brain state can be described according to what function it has. If I am in pain, I am experiencing a pain brain-state.

The functional theory of mind states that mental states are characterized by their causal role in relation to inputs (stimuli) and outputs (behavioral dispositions). Functionalism faintly resembles behaviorism since it is characterized by inputs and outputs; however, there is an important aspect that separates functionalism from behaviorism. "According

to functionalism, a mental kind is a *functional kind*, or a *causal-functional kind*, since the "function" involved is to fill a certain causal role" (Kim). There are inputs, outputs, *and* a correlated mental state. That is, there is an internal state that causally relates to the inputs and outputs. These states are beliefs, desires, etc. Functionalism is a *holistic* theory of mind because all of these states are connected. A person acts on his/her desires because of correlating beliefs. On a behaviorist model, if person A is told that there is a burrito in the fridge, and person A gets the burrito and eats it, it's because eating the burrito is the correlating output to being told that there is a burrito. On the functionalist model, the action of eating the burrito (output) is correlates to the other brain states of 'being hungry,' 'wanting food,' 'believing that there is truly a burrito in the fridge' (trusting the person who's telling person A the burrito is there) and 'believing that the burrito will satisfy one's hunger.' Beliefs and desires fuel actions rather than actions being the result of predictable cause/effect.

Since states (such as hunger or pain) can be defined by their functional role, brain states have multiple relizability under a functionalist model. Kim uses the example of a mousetrap to illustrate the multiple-relizability of a function. A mousetrap might have a simple trigger mechanism, or might have a wire box that snaps shut when sensors are activated. It might even have lasers or be hydraulically powered. No matter the size, shape, or method, anything that captures a mouse is a mousetrap. If Martians landed on Earth and had brains composed of some inorganic matter, but could carry on intelligent conversations and tell us what they think of Earth, it could be said that they have mental states according to the functional theory of mind.

Machine Functionalism

Machine functionalism takes the causal role of mental states in functionalism one step further. It shares the characterizations that consciousness can be determined by inputs, outputs, and causal mental states, but states that mental state's causal roles can be laid out by a machine table. Machine functionalists do not think computers are just an analogy for how the brain works; instead, machine functionalism holds that the brain *is* a computer. The brain is a computing machine that processes symbols into consciousness. Turing machines provide a simplistic model and explanation for how this works.

A Turing machine (named after its inventor, Alan Turning) is a basic computing machine that acts according to a list of rules. The Turing machine's scanner scans symbols on a tape and performs actions instructed by the symbols. Reading the instructions is an 'input.' If the Turing machine is counting or doing an arithmetic problem, the Turing machine will either erase or write another symbol as it computes. These are the 'outputs.' The state in which the scanner is in when it is following

directions/ making the corresponding output to the input is a "searching" state. All of the possible actions a Turing machine can perform are listed in its 'machine table.'

An example of a simple turing machine would be that of a water vending machine. They're located outside of grocery stores and typically cost twenty-five cents for a gallon of water. They accept nickels, dimes, and quarters. When the dispenser isn't being used, the machine is in (state) S1, which is a 'waiting' state. If someone wants water and puts in a dime, the machine doesn't emit any water, but goes into S2; it is now waiting for more change. One more dime and a nickel later, the machine dispenses water and returns to S1. S1 also serves as a 'stop' state. If someone were to put in a quarter to begin with, the machine would dispense water, and then immediately return to S1. If someone put in a nickel, the machine would go to S2. Pretend that this person then finds a quarter in her pocket that she didn't know about, and puts that in the machine. She has put in a total of thirty cents, and the machine goes into S2 and performs three actions: it dispenses the water, it returns a nickel in change, and then it returns to S1.

S_1
S_2
Nickel(s) or Dime(s)
Don't dispense water, go to S2
Dispense water, return to S1
Quarter
Dispense water, return to S1
Dispense water, return change, return to S1

This is an example of a simple machine table. According to machine functionalism, the psychology of humans can be mapped out in a similar way (except it's much more detailed and elaborate). The mind computes information in a very similar way to that of a Turing machine with inputs, outputs, and the causal state between them. However, Turing machines are very deterministic. Since it's assumed (by many) that humans are not beings with a pre-determined destiny, deterministic machine tables seem to disagree with our idea that humans have free will. To accommodate this, people's machine tables are seen as probabilistic. It's probable that I will say 'ouch' when I experience a pain, but I might not. This adaptation makes it so that we're probabilistic automatons, instead of deterministic automatons.

Problems

One problem with characterizing human psychological makeup as mappable by a machine table is the issue of dispositional and occurent states. Dispositional states are mental states one has but are not currently experiencing. Occurent states deals with the

now and what one is currently experiencing/thinking. For example, say that you are currently sitting on a bench on a grassy green hill during spring, and are overlooking a small town. You notice the size and colors of the buildings, the skyline, and are generally focusing on everything around you. However, at the same time that you are observing the town, counting the birds in the sky, and thinking about what a nice day it is--without currently thinking about it--you believe that March follows February, if you hold up a pen and release it, the pen will drop, it's a bad idea to put your hand in boiling water, and many other beliefs. Any other belief that, when questioned about it, you'll say you have. There are so many beliefs that you definitely hold, but are unable to actively entertain at the same time.

It remains unclear how dispositional and occurent states can be mapped out on the same machine table. What casual state would a person be in if she were thinking about math, but also held beliefs that exist but weren't currently being entertained? It can't be said that just because she's not currently thinking that the Earth revolves around the sun doesn't mean that she doesn't believe it. To address this problem, the machine functionalist might say that there are levels to the types of machine states a person can have, and these levels are somehow connected. This is weak and "is incompatible with the view that psychological states are in one-to-one correspondence with machine table states" (Block, Fodor). In order to include dispositional and occurent states, the functionalist would have to alter or dispose of the notion that machine tables accurately map out mental states.

Another problem concerns the functionalist's holistic view of mental states. How can we determine if two people are in the same mental state, such as 'being in pain'? In order to recognize the same mental state in multiple individuals, they'd need to have the same complete machine table. This is impossible, since not everyone has the same experiences and beliefs, and so also have different machine tables. One simple difference, such as, when I stub my toe, I say "Damn!" and when you stub your toe you say "Darn!" would make our machine tables different. Since the output is different, can it be said that we share the same mental state?

A functionalist response to the Damn/Darn distinction, could be that these behavioral dispositions can be lumped together as an "angry outcry." So, we'd be able to share the same mental state because when we stub our toes, we experience pain, and have a pained outcry. Although it's unclear how this would work for other causal mental states. Everybody still has a vastly different machine table from everyone else because their ranges of experiences and beliefs are different. For instance, a masochist who is in 'pain.' Any other non-masochist who experienced pain sensations would say they're in 'pain,' but the masochist would experience a type of pleasurable 'pain.' The masochist and the non-masochist experience the same sensation, but due to their preferences, have very different mental states.

Another issue with the functionalist holistic view of human psychology is the problem of inverted qualia. How can the different variations in human experience be mapped out by a machine table? How would the machine table of someone who is color blind correlate or resemble that of someone who has in depth knowledge and perception for the many different shades of green? Take for example person B. When person B sees a banana, B experiences a yellow sensation. When asked what color the banana is, he says, "It's yellow." Person C, when she looks at a banana, experiences a reddish sensation. However, C has been raised calling (what I would consider to be) 'red,' 'yellow.' So, when person C is asked what color the banana is, she says, "It's yellow." Again, how would person B and person C be able to have the same functional state when the qualitative elements of the object they are perceiving is different? They are not both perceiving 'yellow' as we know it, but they are perceiving the same object. The inputs of person B and C are different, even though they produce the same output.

Another issue with comparing the mentality of a human to that of a machine is the finiteness of a computer's machine table. Computers have fully mappable states. After all, they're programmed that way. (Unless you buy into the idea that humans are 'designed' by some invisible hand), humans are not actually 'programmed.' Instead, we have some base abilities/thoughts that can be added to (learning) and combined in infinite ways. It would be impossible to completely draw out the machine table for a person, since it has the ability to change in any possible way. 'Pain' is a functional state, but there are many different types of pains; when a person steps on a Lego block barefoot, that pain is different than the pain experienced when that same person slams her finger in a doorway. (And what about emotional pain? Does that count?) The infinite varieties of intensity of physical sensations pose a problem for machine functionalism. This dilemma might be answered by saying that these variations within a type of state (pain, etc) can be lumped together. However, by lumping all pains together, it saps the qualitative value out of pain. So, say that my mental states can be completely represented in a machine table, using the lumped version of 'pain.' I step on a Lego and say "Ouch, that hurt!" and then I go and slam my finger into a door and yelp, "Ouch, that hurt!" it seems inadequate to classify different pain experiences under one title and expect the richness and variation of each experience to be maintained when someone says, "She was in pain at time1 and at time 2." I was in pain, but different types of pain, and lumping them together doesn't account for this. I'm not sure how someone who is experiencing multiple types of pain at one time (a toothache, a broken toe, and an intense sunburn) would be accounted for in a machine table either.

One way of salvaging machine functionalism is to consider it as only one possible way mental processes can be understood. Rather than a full machine table that lists all of the causal states, "we may therefore refine questions about the truth of [Computational Theory of Mind] to questions about its truth as a theory *of particular kinds* of mental processes" (Horst [Fodor quote]). In order to escape problems it has with its holistic

view of mental states and the finiteness of machine table states, the machine functionalist could claim that the computational theory of mind might only apply to certain computing-like processes. Perhaps it's inappropriate to describe qualia experiences using machine table states, but how the human brain processes math problems, or any other process that depends exclusively on syntax.

Computers and Consciousness

Not only is a machine table an inaccurate demonstration on how mental processes work, but also the mechanical modes of computation a computer goes through cannot be considered 'consciousness.' Alan Turing devised the 'Turing test' to demonstrate when a computer is 'conscious.' There are two people and a computer. One person is the interrogator, and the computer and the other person are behind a curtain/screen so that the interrogator can't see them. The interrogator asks a series of questions, and tries to determine which is the person and which is the machine. If the computer can trick the interrogator into thinking that the computer is a person, this is considered proof that the computer has intelligence.

John Searle's 'Chinese Room' example is a thought experiment designed to demonstrate how computers which go through the motions of 'thought' and 'understanding' can't be considered to have consciousness. The Chinese Room thought experiment is supposed to represent a computer's 'understanding' of Chinese. Say that you are stuck in a room with an enormous look up table and a little window through which you are handed a sheet of paper that has a series of Chinese characters on it. It's a question. If you're like me, and don't know any Chinese; you don't know what the paper says. But you go over to the lookup table in the room, and the instructions read "When you receive ___symbol, return ____ symbol. So you find the match in the table that matches the one in your hand, write that symbol down, and return it through the little window. You've answered the question without any understanding of its content yourself. Throughout these series of inputs and outputs, the system you are participating in, on a computational theory of mind, would be said to 'understand' Chinese. Yet, you, the causal reason for why the inputs and outputs correspond, certainty do not know and are unable to learn Chinese through this symbol manipulation (Searle).

The reason for why the Chinese room can't be said to exhibit understanding, according to Searle, is the difference between syntax and semantics. Syntax is just the arrangement of words (which are symbols). The problem is that these words have content. Computers are unable to understand this content. It might be able to correctly formulate the sentence "Grass is green," but it will never have the association of the smell of fresh cut grass, nor from this sentence alone—if it isn't watered enough--know that sometimes grass is brown. "The reason that no computer program can ever be a

mind is simply that a computer program is only syntactical, and minds are more than syntactical. Minds are semantically, in the sense that they have more than a formal structure, they have content" (Searle). Symbol manipulation is inadequate for consciousness because symbols have meanings. Accurate use of symbols does not automatically entail that these meanings are understood.

Someone might just might just bite the bullet and say that the Chinese room does in fact 'understand' Chinese since it produces all of the appearances of understanding that we might use to judge it. Besides for the fact that this objection is very behavioristic and so subject to relevant criticisms, this is a troublesome objection. According to the computational theory of mind, there needs to be the internal causal state, along with the inputs and outputs. The internal state, you, the person who is matching symbols but not learning any Chinese, does not know Chinese, despite the understanding behaviors. The internal state does not 'understand.' "To posit a mechanism that understood the meanings of mental symbols would in effect be to posit a little interpreter or *homunculus* inside the head, and then the same problems of coordinating reason and causation would recur for the homunculus, resulting in a regress of interpreters. On the other hand, it is hard to see how a process specified in purely causal terms could thereby count as a reasoning process, as calling something "reasoning" locates it with respect to norms and not merely to causes" (Horst) Furthermore, if all that matters for something to have understanding are outputs correctly corresponding to inputs, then this model of understanding could have some absurd implications. Ned Block's Chinese nation example demonstrates just this.

Imagine that the entire nation of China collaborates together and every citizen is fully connected to each other "with a specially designed two-way radio that connects them in the appropriate way to other persons and to the artificial body" (Block, "Troubles with Functionalism"). Satellites are set up in space so that everyone can read the same message (input) and though a chain reaction, act like neurons and produce outputs. "Whatever the initial state, the system will respond in whatever way the machine table directs. This is how *any* computer realizes the machine table it realizes" (Block, "Troubles with Functionalism). Someone who admitted that the Chinese Room exhibits understanding and is 'conscious' might be driven to also say that the unified nation of China, acting as one being is 'conscious' too. Many would find this counterintuitive to grant a synchronized nation the same level of consciousness that we claim human beings have.

Another problem with attributing consciousness to computers, are their limits. This was previously discussed in relation to why computers are an inaccurate model of the mind; this time it's why computers can't (or are very unlikely able to) reproduce typical indicators of 'intelligence'/'consciousness.' The human mind, is not infinite in its capacity to store information, but is infinite in its potential to produce completely new

thoughts. Computers on the other hand are limited by the hand that programmed it. Computers are capable of mixing the symbols it is provided with in any variety of ways, but the machine table it has is finite. The Isaac Asimov short story "Risk" demonstrates (in a thought experiment way) how ingenuity and interpretation are necessary for true 'intelligence.' Engineer Gerald Black is sent aboard a spaceship--that was supposed to enter hyperspace with its robot pilot (since previous tests showed that hyperspace [as of yet] was too dangerous for humans), but something malfunctioned. The robot was conditioned to, when the signal came, to pull the bar back *firmly*, and that was exactly what it did. Nothing happened, and since the ship could potentially (dangerously) leap into hyperspace at any moment, Gerald Black was sent to find out what went wrong. He discovered that the robot did pull the bar back firmly, and since it had superior strength compared to a human, and unable to interpret what 'firmly' really meant, it completely bent the control bar and dislocated the lever. When Black successfully returns after smashing the robot, he asks robopyschologist Susan Calvin why he was sent, since it was so dangerous, instead of an expendable robot. She replies:

> "Robots have no ingenuity. Their minds are finite and can be calculated to the last decimal. That, in fact, is my job. Now if a robots is given an order, a *precise* order, he can follow it. If the order is not precise, he cannot correct his own mistake without further orders. Isn't that what you reported concerning the robots on the ship? How then can we send a robot to find a flaw in a mechanism when we cannot possibly give precise orders, since we know nothing of the flaw ourselves? "Find out what's wrong" is not an order you can give to a robot; only to a man. The human brain, so far at least, is beyond calculation" (Asimov)

The scope of a computer's abilities is shaped by the hand that designs it. Whereas, in some respects, computers can be superior to humans (once designed correctly) and perform any number of algorithms without flaw, they are limited in ways that humans are not. This might be a temporary, arbitrary problem; whereas it can't be empirically proven that the brain is actually a machine, consciousness might someday arise from a mechanical source. The sentient mechanical beings of science fiction might become a reality. Even today we're capable of producing androids with human-like appearance and some behavioral capabilities. However, as long as an understanding of semantics and the capacity for ingenuity are requirements for consciousness--instead of simply generating appropriate outputs to inputs-- artificial consciousness hasn't been achieved.

Works Consulted/Cited

Asimov, Isaac. "Risk." *The Rest of the Robots*. Hammersmith, London: HarperCollins, 1997. 123-55.

Block, Ned. "Introduction: What Is Functionalism?" *Readings in Philosophy of Psychology*. Vol. 1. Cambridge, MA: Harvard University Press, 1980. Pages 171-84.

Block, Ned, and Jerry A. Fodor. "What Psychological States Are Not." *Readings in Philosophy of Psychology*. Vol. 1. Cambridge, MA: Harvard University Press, 1980. Pages 237-50.

Block, Ned. "Troubles with Functionalism." *Philosophy of Mind: Classical and Contemporary Readings*. New York: Oxford UP, 2002. 94-98.

Horst, Steven, "The Computational Theory of Mind", *The Stanford Encyclopedia of Philosophy (Spring 2011 Edition)*, Edward N. Zalta (ed.), http://plato.stanford.edu/archives/spr2011/entries/computational-mind/.

Kim, Jaegwon. "Mind as a Computing Machine." *Philosophy of Mind*. Boulder, CO: Westview, 2006. 115-45. Print.

Levin, Janet, "Functionalism", *The Stanford Encyclopedia of Philosophy (Summer 2010 Edition)*, Edward N. Zalta (ed.), http://plato.stanford.edu/archives/sum2010/entries/functionalism/.

Putnam, Hilary. "The Nature of Mental States." *Philosophy of Mind: Classical and Contemporary Readings*. New York: Oxford UP, 2002. 73-79. Print.

Searle, John R. "Can Computers Think?" *Philosophy of Mind: Classical and Contemporary Readings*. New York: Oxford UP, 2002. 669-75. Print.