On the Possibility of Designing a Machine with Mental Capacities

Chris Wrede

December 1998

Class: Minds and Machines I

A strong argument for the possibility of designing machines with mental capacities is based on the identity theory of the mind. The argument holds that the universe is only made up of physical stuff. Every thought, belief and feeling is a result of a corresponding physical brain activity. Since the brain is a physical device and mental capacities come from the brain, and we can design any machine made of physical stuff, it is possible to design a machine with mental capacities. On the other hand the human brain is very complex and we do not understand all of physical law. An understanding of the complexity of the brain and all of physical law is necessary for designing a machine with mental capacities. We will likely never understand the complexity of the brain and all of physical law so we will likely never be able to design a machine with mental capacities. Even if we did understand all of the complexity of the brain and all of physical law, the only machine which we could design with mental capacities would be a biological brain. Nevertheless, a perfect design for a machine with mental capacities lies within human DNA, so if we can understand DNA then we can design strands of DNA to produce human machines with mental capacities.

The identity theory of mind holds that mental states are merely physical brain states. It arose as a response to other theories of the mind which hold that we have separate physical and mental states. The first of these is Dualism which was put forth by Rene Descartes. It holds that there are mental states which are independent of anything physical. However, the concept of nonphysical mental phenomena interacting and changing the physical world is implausible since it violates physical conservation laws of energy and momentum. To account for this, the epiphenomenalists proposed that mental phenomena are caused by physical phenomena, but that physical phenomena are not caused by mental phenomena. However, we know that moods and other mental phenomena affect our physical behavior. Behaviorism, then, suggested that language complicates the mind/body argument; all talk of mental phenomena can be simplified as behaviors or inclinations to behave in a certain way. The flaw with behaviorism is that it denies the existence of our subjective mental states.

Identity theory, by equating each mental state with a physical state avoids this problem. J.J.C. Smart proposed identity theory in 1959 with his landmark paper, Sensations and Brain Processes. In this paper he suggested that sensations are more than simply correlated with or caused by brain phenomena, they *are* brain phenomena. He later extended this argument to include with sensation all mental states including emotions and beliefs. His identity theory contends that when one reports a sensation, he/she is merely reporting on the state which his/her brain is or was in. The person may not realize that he/she is reporting on the state of his/her brain; the person may not even realize that he/she even has a brain. This is the precise reason why people tend to claim that there is a separate mental world: they do not fully realize that they have a brain. However, as science has advanced we have realized that mental phenomena are correlated with the brain. The ultimate goal of neuropsychology is to prove that every mental phenomenon is precisely correlated with a brain state; that identity theory is correct.

Naturally, until neuropsychology reaches this point of advancement, identity theory will not be proven correct. This is one of the main objections to identity theory. Another objection was that the ancients knew about mental states while knowing little or nothing about the brain. The rebuttal to this argument is that mental states are phenomena like the physical phenomenon of lightning. We can know of lightning before we know about electricity. However, when we discover electricity, we realize that lightning is electricity. In the same way, mental states are physical states. A third objection to identity theory was from Leibniz's law that if x=y then x and y must have all the same properties. This objection suggests that mental states do not share the same properties as brain states. For example, when we have an after image (mental state) which is yellow, this does not correspond to a yellow brain state. Conversely, a brain state may be at a specific temperature but a corresponding mental state, say a belief, could not be said to have a temperature. The rebuttal from the identity theorist here is that mental states are not objects so physical properties (like yellowness) cannot be ascribed to them. Mental states do not have physical properties so the fact that a brain state can not be yellow although we can perceive yellow is no objection to the claim that mental states and brain states are the same. A final objection to identity theory is that we can know the brain state of an individual and still not know what it is like to feel what the individual is feeling. These entities have been called "qualia" because of their qualitative aspects. Identity theory does not account for this debatable epiphenomenalistic objection. For the purposes of this paper, epiphenomenology serves the same purpose as identity theory since they are both physically deterministic theories of the mind.

The modern view of physical determinism developed in the seventeenth and eighteenth centuries as modern science began to predict the motions of physical objects using mathematics. One of the best examples of modern physical determinism is the system of Thomas Hobbes. In his philosophy he attempted to describe all of human nature through the motion of physical bodies, physics. Like Smart, Hobbes believed that all psychological processes are physical states of the brain. So, if we know the position and motion of every particle in a brain at a given time, then using the laws of physics we can predict where they will be at any time in the future. From this, and a knowledge of which brain states correspond to particular mental states, we can predict the mental states of the brain. Put simply, we will be able to predict what people are thinking, and how they behave as a result.

If identity theory or epiphenomenology are accepted, then one can argue from physical determinism that it is possible to design a machine with mental capacities. Julien de la Mettrie wrote in 1748 that "man is a machine". Wooldridge argues simply that "if there is a purely physical explanation of brain performance, then computer-like structures are in principle capable of precisely duplicating such performance". He goes on to conclude that since all intelligence is based on complex switching networks, "the ordinary laws of the physical scientist are adequate to account for all aspects of what we consider to be intelligent behavior". If we can account for all aspects of intelligent behavior by physical means, we can design a machine which has mental capacities.

There are several objections to this argument. Firstly, some would argue that the unpredictability that results from quantum mechanics means that we cannot completely predict the behavior of organisms or machines with complete accuracy. Wooldridge addresses this point by arguing that there, nevertheless, is no way for a nonphysical agent to participate in the behavior of an organism or machine. So the organism or machine is still determined by physical law. Another objection arises from biological chauvinists who take the stand that only things made of biological material can have mental capacities. An example of this is John Searle's Chinese room argument which says that if we build a machine which looks, acts and behaves exactly like a human being, it can only follow the rules it has been programmed to follow for reading input, processing it, and behaving accordingly. It does not have mental capacities, however, because it does not understand what its input means. Hilary Putnam's rebuttal to the Chinese argument is the question: What if we were created by a race of superior creatures (with a small 'c' he stresses)? Would this mean that we do not have mental capacities? An interesting point is raised here that perhaps we consider "mental capacities" to be too unique and special to our own organic material. We find it hard to imagine that our mental capacities are not unique, in the same way that we found it hard to imagine that the sun does not revolve around us before the Copernican revolution, that the galaxy does not revolve around us before Shapley, and that we are not at the center of the expanding universe though we appear to be from our subjective perspective. Arguments like Searle's are, then, biologically chauvinistic and, perhaps, biased.

On the other hand, these objections to the potential mental capacities of machines can be combined to form one of the strongest arguments against the possibility of designing a machine with mental capacities. Biological chauvinism, the quantum nature of consciousness, and the complexity of the brain suggest that we will never design a machine with mental capacities. Although it is easy to make the argument that there is some nonphysical entity controlling brain function that we will never understand or reproduce, this argument acknowledges that the universe is purely physical as a starting point.

Clearly the human brain is very complex. Much of the input and output abilities of the brain are performed by electrical pulses flowing through a net of interconnected neurons. As a result, people have attempted to reduce brain function to the function of a complex computer containing only a mechanical analogue of these neural nets. However, the brain does not contain only these neural nets for transporting information. There are neurons in the brain which do not take part in these neural nets. These dendritic neurons have no axon (the long transmitter of electric current from neuron to neuron). A recent study (Pribram,1991) seems to show that the workings of the cerebral cortex are better understood through a network of dendritic neurons with no axons. Furthermore, the neural net model does not account for the glial cells surrounding the neural network which clearly play a part in the transmission of electric pulses down the axon. The Japanese researchers Mari Jibu and KunioYasue conclude that "neurons without axons and glia play the principal roles in the fundamental processes of the brain". Furthermore, the brain is dynamic, not static. For example, the connections between neurons (synapses) are not static, they move around to find an appropriate place to transmit current. This makes the amount of current which they transmit variable. Furthermore, it was once thought that the human brain does not grow new neurons once it has matured. However, it has been shown recently (Gage, 1998) that the human brain is capable of growing new neurons throughout life. Simplifying the brain to a net of neurons with axons is, then, far too simple a model to use to design a machine with mental capacities.

The quantum nature of the brain is another significant factor in attempting to design a machine with mental capacities. The brain, like everything else, is made up of microscopic particles which behave in a way that is not like the way the classical macroscopic world behaves. The quantum behavior of these particles is, at the current state of science, predictable only in a probabilistic sense. This would not pose a large problem if the functions of the brain were only dependent on larger molecular and biological parts. However, it seems quite possible that what we know as consciousness is largely dependent on quantum phenomena. The mathematical physicist Roger Penrose gives the example of a single celled organism, the paramecium. Naturally, a paramecium has no neural net since neurons are cells themselves and it is a single celled organism. Nevertheless, a paramecium is able to "swim towards food, retreat from danger, negotiate obstacles and , apparently, learn by experience". This behavior has been attributed to tiny micro tubules in the little hairs, or cilia, with which the paramecium is covered. Penrose attributes the dynamic behavior of neuron synapses to information being transported through micro tubules. He then advances a theory that the micro tubules serve as an encasement from random outside phenomena. Within this encasement occurs some type of coherent quantum oscillation which extends over very large areas of the brain. Some people, like Wooldridge, would argue that this is still a physical phenomenon and is therefore capable of being replicated by a computer-like device. However, Penrose suggests that this system would significantly involve the quantum phenomenon of non-locality. This is a phenomenon in which two particles a significant distance apart (even kilometers) can exhibit simultaneous related behaviors. This phenomenon is not well understood at this time and it seems highly implausible that is it intentionally reproducible by humans to make a machine think.

Here Searle's biologically chauvinistic standpoint is relevant. The brain is very complex, and since the essential functions of the brain seem to rely on some of the most complex phenomena in the brain, we must be able to produce a very complex model of the brain to produce a machine with mental capacities. In fact, these complex models of the brain would likely converge to an exact model of a biological brain.

DNA is far better than we are at producing mental capacities. We are close to mapping the human genome, and close to understanding which genes on DNA correspond to given characteristics in living organisms. It is far more plausible that we will build living machines based on manufactured DNA before we design a machine from scratch to have mental capacities. Why would we take on the arduous task of putting together a brain piece by piece to perform a certain way when we can produce the same results by simply constructing a strand of appropriate DNA and making a clone from it? Attempting to build a machine with mental capacities out of material which is not biological is completely illogical. We already have the design for machines with a wide variety of mental capacities in the molecule DNA.

Since the human brain is purely physical, it can be argued that its workings and the mental capacities that result from these workings can be duplicated by a machine. However, all of the mechanical models of the brain that have been proposed to mimic the workings of the brain are too reductionist and simple. We do not know and likely will never know all of the laws of physics, and as a result will likely never fully understand how the human brain works. As we discover more about the workings of the brain, these simplified models of the human brain will become more complex and seem as though they will converge with time to an exact model of the human brain. Currently, the best design for a machine with mental capacities known to us lies in human DNA. We can manipulate this design to produce a human machine with a wide range of mental capacities.

Bibliography

Jaroff, Leon. (1997). "Keys to the Kingdom". Time. Special Issue: The New Age of Discovery. (pp. 56-61).

Churchland, Patricia S . Neurophilosophy. (Ch. 2).

Feigl, Herbert. (1981). "Mind-Body, Not a Pseudoproblem". (1960) In Inquiries and Provocations. Holland: Reidel. (pp. 342-351).

Gibbs, W. Wayt. (1998). "Dogma Overturned". Scientific American. Nov, 1998. (pp.19-20).

Graham, Robert B. (1990). Physiological Psychology. United States: Wadsworth. (pp.32-33).

Hobbes, Jesse. (1991, June). "Chaos and Indeteminism". Canadian Journal of Philosophy. (pp. 141-164).

Jackson, Frank. (1998). "Mind, Identity Theory of". In The Enyclopedia of Philosophy, Vol.6. London: Routledge. (pp.394-399).

Jackson, Frank. (1998). "Smart, John Jamieson Carswell". In The Encyclopedia of Philosophy, Vol.8. London: Routledge. (pp.813-815).

Jibu, Mari. Yasue, Kunio. (1995). Quantum Brain Dynamics and Consciousness. Philadelphia: John Benjamins. (pp. 103-116)

La Mettrie, Julien Offray de. (1748). Man a Machine. Illinois: Open Court. (p.148).

Morgan, Charles G. (1998). Phil 342A Class Discussions.

Navarro, Julio. (1998). Helpful Discussions on astronomy (humans at centre of everything etc.)

Searle, John R. (1980). "Minds Brains and Programs". In The Behavioural and Brain Sciences. Vol. 3. (pp. 427-457).

Penrose, Roger. (1997). The Large, the Small, and the Human Mind. Cambridge: Cambridge University Press. (pp. 50-139).

Putnam, Hilary. Robots: Machines or Arificially Created Life?. (p191).

Taylor, Richard. (1967). "Determinism". In The Encyclopedia of Philosophy, Vol.2. Great Britain: Crowell Collier and Macmillan. (pp. 359-373).

Wooldridge, Dean. (1968). Mechanical Man

Edited by Patrick Edward Meyer.
Revised May 1,1999.