Avinash Dindial

PHL340

The hard problem of consciousness can be defined as explaining subjective experience in terms of brain processes. It is already understood how we can explain objective experience through brain processes and perhaps the best example of this is in explaining intelligence. This is because intelligence does not rely on how a system feels, it only focuses on how to arrive at certain goals. Subjective experience has the problem of appearing not to have any neural correlates that ties this experience to the brain. It is the belief of Chalmers that if we can somehow copy each neuron into an artificial being one at a time, at some point the artificial machine would have the same functions (and same consciousness) as the original. This bypasses the issue of finding where subjective experience lies in the brain by simply copying the entire brain. This paper seeks to defend this stance as a plausible way of maintaining consciousness. The following paragraphs further explains, analyzes and criticizes this view.

To begin to understand this question, one must first understand what consciousness is defined as in this paper. Chalmers describes this as subjective experience; what it is like to experience a certain event. To understand subjective experience, consider the Mary the scientist thought

experiment. Suppose Mary is blind but knows everything there is to know about colors, light, wave lengths etc. She would know the wave length of the color red but now suppose she suddenly regains sight and sees the color red for the first time. Despite her knowledge of everything concerning the color red, she would now have the experience of seeing red; something she did not have before. Chalmers proposes the problem of consciousness can be explained by defining easy and hard problems of consciousness. The easy problem refers to understanding how humans can categorize, solve problems etc. By no means is understanding the easy problem of consciousness but we have some sort of idea on how to approach mechanizing this aspect of our consciousness. For instance, modern AI exhibit narrow intelligence. Modern machines have the ability of identifying colors (based on wave length), but they would not know what it is like to experience seeing the color (they do not exhibit subjective experience). If a machine is capable of solving a wide range of problems, we call this general intelligence. There is steady progress in Cognitive science to understand how we can achieve general intelligence by using neural networks and other theoretical breakthroughs such as marks of the cognitive or relevance realization which can tell is whether a machine is a general problem solver. Contrast this with the hard problem of consciousness which is explaining subjective experience.

The goal of Chalmers is to find a way for machines to exhibit subjective experience. One way he thinks about arriving at this is to copy an already existing consciousness from a human into a machine. To do this, he relies on a process known as gradual uploading. The basis of this process is the belief that if a system with a biological brain is conscious, then an exact, artificial replica of this brain would also be conscious. To do this, he considers perfectly simulating every neuron in a particular brain, one at a time. In addition, blood flow and neuronal pathways etc., would also need to be replicated. However, for the simplicity of this argument, I will only consider replicating neurons in this paper. He argues the easiest way to simulate a brain is to do this in stages or small increments, so he suggests building an artificial brain one cell at a time. So, at first, a couple of cells would exist in this brain but slowly all neighbouring cells would be perfectly joined as in the original brain. The argument can be made that this gradual uploading retains consciousness because if it doesn't, then it raises the question "Where did consciousness vanish along the way?". This question poses a challenge in that if consciousness is lost, then there must be a neuron or groups of neurons that is responsible for subjective experience. In either case, it is a desired outcome as it would tell us either consciousness is retained, or we can find a location in the brain that is responsible for subjective experience. There are two possibilities in which consciousness can vanish- either all at once or gradually. Indeed, it may be a challenge to determine whether or not consciousness is gradually vanishing because if the subject of this experiment is asked

questions, they would provide answers consistent with their fully conscious self. In such a case, one can pose the question of whether there exists partial (subjective) consciousness. This should not be confused with partial consciousness as in vegetative states or impaired cognitive functions. On the other hand, the possibility of consciousness vanishing completely at a specific point seems even more implausible than this according the Chalmers. Then he proposes a way around this issue. Perhaps there exists the crucial neuron responsible for consciousness. Then, perhaps replicating this in even smaller increments may allow us to reach what Chalmers calls the crucial quark. Even if this is the case, this would be a huge step forward in the study of consciousness. It may be the case that this crucial quark can be replicated in the distant future if this theory stands. However, this does not seem to be the case given current literature. The last possibility of the gradual uploading experiment is that consciousness is retained. Given the implausibility of the previous two possibilities, we are left to conclude that simulated brains must be conscious themselves given they are the exact copy of an organic brain.

I argue that this gradual uploading process does indeed retain consciousness. I agree with this because I would describe myself as a functionalist. Under functionalism, which I argue is the best theory we have of describing mental processes and relating it in such a way that we can achieve mechanization of these mental processes. Functionalism claims that mental states are

constituted by their functional roles. In the case of gradual uploading, biological neurons are replaced by functionally equivalent artificial neurons. Under this theory, there is no difference between the artificial and the biological brain as the functional roles are one and the same. Under this, it raises the consequence of identity. For if we have two or more functionally equivalent brains, then the consciousness between them would be identical. To discern the consciousnesses, one might suggest that the original consciousness is the one who has a biological body. However, upon death- as is a possibility of destructive uploading, would the instances of the consciousness still be regarded as the person who died? Would we have to treat the machine the same way, as if death is now an obstacle of life that can now be evaded? Perhaps this is the key to eternal life as biological beings tend towards death and decay whereas artificial consciousness does not face this inevitability. Functionalism also has the characteristic of allowing mental states, processes and by extension, consciousness to be multiply realizable. This states consciousness does not require a biological system to harbour it, rather consciousness can exist in artificial systems as well. The best example of multiple realizability lies in nature itself. The mental state of pain has neural correlates in humans that is vastly different from animals such as the octopus. In this case, the same mental state is run on different operating systems, similarly, the same consciousness can exist in both a biological system as well as an artificial one perhaps even simultaneously.

The biggest consequence regarding this functionalist perspective concerns identity and ethics. Suppose we successfully uploaded a person's (Call this person Daniel), mind to an artificial equivalent (Call this AIDaniel). It is natural to say that Daniel would be the 'real' Daniel and therefore be allowed to claim his possessions, relationships etc. However, under Functionalism, AIDaniel would have the same memories, consciousness etc., as its biological counterpart, so how would we deal with situations such as this? Suppose this artificial Daniel has feelings as we would expect, then upon realizing everything he thought was his, including his wife etc., now does not belong to him, in what mental torture would we bring AIDaniel into this world? It seems almost cruel to allow such situations. This consequence of identity cannot be overlooked when attempting to consider the possibility of replicating a mind. This consequence is only unique to gradual uploading and not building a whole new consciousness from scratch. In fact, Chalmers agrees that the only way we can truly have artificial consciousness is through the gradual uploading process. So, we may never get fully unique AI, rather only copies from biological volunteers. Given this account, identity would always pose a serious threat to uploading consciousness. This consequence can also be seen in current media such as the Netflix series Black Mirror. In one particular episode, people are allowed to upload their consciousness that would serve in a similar manner to a Google home device. The biological person would give their uploaded consciousness orders such as turning on lights etc. This uploaded consciousness was depicted in the episode to be in an eternal state of torment,

similar to slavery. The topic of ethics regarding uploaded consciousness must be studied in more detail as we cannot discount the feelings and claims that AIDaniel would have as they would be equivalent to the claims made by its biological counterpart.

We may never know how to get around such circumstances, however, I argue the only way in which gradual uploading should be encouraged is when the biological host is close to death or willing to sacrifice his organic body for a potentially eternal artificial replacement. In either case, it should not be permitted that there exists two identical consciousness. This above consequence does not refute the plausibility of gradual uploading, rather it addresses what may happen if it were to be realized. I do not see any strong arguments against the gradual uploading theory other than its applicability in the real world given the large number of neurons in the brain. However, providing future technology can bypass this, I do not see any obvious or even nuanced objections to this theory. The consequences of this seems to be a much larger issue to study rather than its plausibility. It is perfectly possible that in the distant or near future, there would be ways of replicating the functionality of neurons on a large scale as it is already being done in small scale operations currently. The question is not if this gradual uploading would be attempted, but it is a question of when humanity would reach a point in which this could be done experimentally. Until then, the question of identity should be seriously considered and debated as there does not seem to be any clear-cut rules or procedures on how to handle such situations. In conclusion,

it seems this theory of gradual uploading is possible and that it retains consciousness especially when assuming functionalism. It does not seem to be a question of plausibility, rather the consequences seem to entail situations we may not be yet ready to deal with just yet.

References

Chalmers, D. J. (2022). *Reality+: Virtual worlds and the problems of philosophy*. Penguin UK.