Kandidat 4 Oppgave 1

Citation 1 by David J. Chalmers

Uploading and the self

The citation is discussing what happens if we upload our mind to a computer. Could we upload something conscious to a computer? And if so, would the upload be the same as what was uploaded? Could you upload a person into a computer, such that that person would “live on” in the computer, and all the person’s mind and personality would be intact? The deeper and underlying question is what is the “self”? Could a self be removed from the body, and what happens if there are several “selves”? Which of the “selves” are the true “self” or can there be several true “selves”?

First of we need to know why something can even be conscious, or why is there any “selves” at all. Biologically it seems that our brain is just a lot of nerves signaling to each other in a complicated pattern and responding to different hormones to produce a certain behavior. For some reason are cells alive, even though they only comprise dead parts. It could that it is some concept or reason, like God, that we haven’t discovered yet, or it could be just that things randomly pan out and make a stable system. Anyway, we don’t know why we are conscious so it is hard to say that a computer can’t be. We are in fact quite similar in that we use a lot of tiny parts that signal each other to process information.

If we for some reason presume that computers could work the same way as humans, and computers work almost deterministically, would that assumption imply that humans work in a deterministic manner. We like to think that humans can make choices. Truly free choices that are not forced and are just based on the persons wants and needs in that exact moment. We tend to believe that if I wanted to do something truly random, I could do that, and that creativity in some sense is not deterministic. In that way it is very tempting to reject the assumption that computers and humans work in the same way, because that would limit all of human existence to a forced path which could have been computed and simulated beforehand. Nevertheless, it could be possible. While we are able to study how computers compute things, we are not yet able to study exactly how our brain would do the same. Anyway, it would not be possible to make a computer take on the full potential and extent of a human mind without humans acting deterministic (unless we make computers that are non-deterministic (1)).

If we assume that humans are in some sense special in their intelligence, and there is something truly nondeterministic about it, it will mean that nondeterminism is possible and that our computers can’t work in the same way. The only problem is that that nondeterminism would come from somewhere. It could be a God, but that is in some way to reject finding an answer to the question. Calling it God would not give us any more insight since it can’t be tested and would always remain speculation. Another way to have nondeterminism is if the laws of physics are nondeterministic, which quantum physics suggests. The problem with that though is that on any other scale it seems that the nondeterminism even out to become sort of deterministic again, almost like if you have a random normal distribution, it is very unlikely that you get a result that has a large deviation from the average. It also seems that most of the

worlds macro systems work in a deterministic way, so it would be strange for the human brain to be the only thing based on the slim chance of repeated nondeterminism. Therefore, it seems likely that you could in some way simulate a brain.

So, if we assume that there is nothing special about our brain and that our mind is mostly governed by our brain, and that we could simulate it on a computer or device, what would that mean? We would have a definition of what we upload, since our brain stores information in another way than a human. Our brain is also based on the hormones and amount of sugar in our blood, so to replicate a human, we need to simulate all the different parts of the body that could influence the brain. If we do that, we could simulate a person, and at least for any practical sense; something that cannot be distinguished from a human mind.

As an example, if we assume that someone is about to die and gets uploaded, such that afterwards only the upload is remaining. Would the same mind still be alive in the upload? One might say no simply because it has taken a different form. If we make a copy of something; let's say a table, we now have two identical tables. They are similar in every sense, but they are separate tables. Same characteristics, different "essence". But likewise, say that you copy a unique table, such that you now have two of them. Then you destroy one of them and ask: does a table of that type or characteristics exist. The answer would be yes, like it would be before the copying. These are two different ways to view an object, either by its characteristics or by its essence. If we value a mind only by its characteristics then we could say that the copy is equal to the original, but if we say that the copy is different since it is a copy, then we have difference. The question boils done to whether we give the copy an aspect that is not based on any physicality, that the original has. In any practical sense the person lives on, but to some degree the "essence" has changed.

Another question is whether two minds can be the same. Am I the same as I was yesterday? Obviously, some processes have happened in my brain that led me to have a slightly different understanding than yesterday, but we still treat us alike. The "self" is still the same "self" even though it has changed. So, in some sense, I am more than the exact state of my mind. And to what degree can we change something, and it would still be the same? And if a computer is in another state that is deemed me, are there then several different versions of me?

You could reject that something like a "self" exists and imply that you are not the same "self" at all times. Simply put, you are not the same as yesterday. A counter argument to that is that then you in some sense must reject fairness in consequences. If you are not the same, why should you be punished for something you haven't done? And why do we deem it fair? It seems that our system and way of thought implies some sense of responsibility, that you need to comply with the consequences of that responsibility. One could argue that this is just in a juridical and social sense, and that in a philosophical sense we are never the same. That then becomes a question of definition, based on how we define a "self".

If we make a *perfect* upload of me, I think it makes the most sense to label both the versions as me. If one version is created and the characteristics are completely similar, you can't distinguish it, and no difference appear. Since there is no difference, it doesn't make sense to differentiate them. However, immediately afterwards, the two alike minds would diverge since they would get different sensory information and they can't occupy the exact same state and position in the world. If the two minds got the exact same external influence, it again becomes

a question about determinism whether they would stay the same. If we don't have determinism, they would gradually by chance diverge, otherwise they remain similar.

One could argue that we distinguish based on non-physical all the time in the day-to-day life. For example, if we have two paintings that look completely similar in every way. One of them made by a famous painter, the other a forgery. Most of us would probably want the non-forged one, but honestly, we would not know the difference. They might have been swapped, and unless you're very skilled at checking forgery or are going to pay for someone else to check it, you would never know. Nevertheless, we would still want the one that is original. However, at least in this case, I think we don't want it because it was a certain person who physically painted it, but we want to give the honor to the actual creator of the painting, and not the person who reproduced it. We value the actual work behind and the corresponding story, but if they are both similar it would be no reason to differentiate. If two painters painted the exact same painting the exact same way with the exact same story, the value of the paintings would be the same. So, if we're able to upload in a way such that all the characteristics would be the same, it would make sense to differentiate. If we copy in such a way that there is something different about the story or any other characteristic we value, then we could treat it differently and it would not be the same.

Whether we could make a copy with all the same characteristics is hard to say. We could maybe make it in a way that we can't differentiate it based on empirical evidence, but its unclear whether we could at any point in the future differentiate it. Maybe it is not possible to make a copy good enough for us to have to refer to it in the same way as the original. But if for all possible practical purposes something is the same, we should be able to have them as equal.

(1) It could be that we make quantum computers that are based on chance. They could be nondeterministic, since quantum states has a necessary nondeterministic aspect.