Research Essay

The Problem of the Self

Cosmin Vișan*

Abstract

Consciousness presents us with many aspects. In trying to explain consciousness, one may be tempted to address only the problem of qualia, as for example explaining color red. But can this attempt be done on its own without somehow taking into account also the subject of experience? In this paper, we will concentrate in addressing the problem of the Self without any reference to any particular quale. The best place where the Self can be analysed is at a point where it no longer exists, that being in principal the moment of death. By analysing what life after death might mean, we will shed light on some characteristics of the Self. It will turn out that the problem of the Self is the problem of the continuity or discontinuity of the Self.

Key Words: Consciousness, Self, thought experiment, continuity, discontinuity, sleep, death.

Introduction

The way in which we choose to analyze the Self in this paper is through a series of thought experiments about how the Self might behave when subjected to extreme conditions. This will eventually start to show a specific pattern. Namely, the problem of the Self will turn out to be the problem of the continuity or discontinuity of the Self. The continuity is what we experience every day when we are awake. The discontinuity is what happens when we die, but also when we go to sleep. It will turn out that in order to understand the Self, we will have to understand what exactly is its continuity or discontinuity.

Since the biggest discontinuity of the Self is at the moment of death, let us start by seeing how is death currently seen by the materialistic science. If, as science might suggest today, we are only our atoms, then it is straightforward to see what happens when we die. Since what we see is that our body is destroyed, and we were only our body, then it is clear that we will also be destroyed. But is this all there is to be said? The first thing that in actuality is different from what science assumes, is that besides the atoms that are fairly well described by today's physical theories, we also have consciousness. Actually, we might say that we are that consciousness more than we are our body. And since science has no idea what consciousness is and how it comes about, the problem of the death stops looking that straightforward.

It is commonly assumed that whatever consciousness is, it surely must arise from the activity of the brain. But this is merely an unproved assumption. As long as we don't know how consciousness arises, we also cannot say how it seems to disappear. Since we have no conclusive proof that it is generated by the activity of the atoms in the brain, in the same manner we have no conclusive proof that death is the end of consciousness. And more than this, even if eventually

* Correspondence: Cosmin Visan, Independent Researcher. Email: visancosmin17@yahoo.com

will turn out that indeed consciousness is produced by the brain, this will still not mean that death is the end of the Self. It might turn out that the relations in the brain that give rise to the Self are the same in all persons and this will mean that we are all one and the same person, death merely being the end of this life, but not of the Self, the Self going on to live another life somewhere else.

So what we begin to see is that there are a number of possibilities that might happen at the moment of death. In this paper, by looking at a series of thought experiments, we will try to see what is the best way to regard the discontinuity of the Self at the moment of death, and so we will try to understand what the Self might be.

Me, Future Me, the Same Self

If we were the same entity all along our existence, then we will simply be that entity and there would be no mystery to worry about. However, it is a known fact that all atoms in our brain and body are changing multiple times during our lifetime. But this change has no impact whatsoever on the Self. Even more so, not only that the atoms are changing, but also our personality, our tastes, our way of perceiving the world. Yet, the Self is the same. I am the same Self that I was 50 years ago. How can this be so? How can the Self continue even when the body and its qualia are changing? Let's call this: the problem of the continuity of the Self. At this point, we have to be careful about the concept of continuity. Even though it might appear to us that what we are dealing with here is a continuity, it might turn out not to be so. It might turn out to be just many numbers of discontinuity moments one after the other. But even this view will poses serious difficulties. So we need to be very careful about what continuity or discontinuity might mean. They might simply be two human words that have nothing to do with what is really happening out there. For the simplicity of arguments though, we will loosely employ these two notions, but we will always keep at the back of our head the warning that these might not be the proper notions to use in the problem of the Self.

Before talking more about continuity and discontinuity, let's have a look at some thought experiments without even involving these concepts. We will be back to them after all the experiments have been presented.

The Cryogenic Experiment

Let there be 2 cases:

1) Your brain is cryogenically frozen, or some similar procedure that can keep your brain unchanged for an indefinite amount of time. You are then brought back to life after 1 million years. I think this is obvious to anyone that you will be that person that is waking up.

2) You suddenly die. Your brain is destroyed. However, after 1 million years, a brain identical with the one that you had when you died, is built. Who will be in that brain?

It appears that the 2 cases are identical. There was a moment (A) in time when your brain was working and it was you who was in that brain. Then there is a period when the brain is not working (in one case it is cryogenically frozen, in another it doesn't exist). And there is another moment (B) when a brain identical with the brain at moment (A), is working again and someone is in that brain. Since in the case 1), that someone is you, why should the case 2) contain a different person? Let's conclude that in case 2), that someone is also you - a fair conclusion to accept. After all, why should it matter at all what happened to the brain in the interval between (A) and (B), if (A) and (B) seem identical in both cases?

But now let's consider a third case, which will make the situation more suspicious.

3) Let us modify a little case 2) and say that you don't die, but while you're still alive, an identical brain with yours is built. Since we established for case 2) that an identical brain with yours will contain you, then also in this case an identical brain should contain you. But how can this be possible? You're still alive and you definitively know that you are in the brain that you've been since you were born. How can you now be in a second brain at the same time? We clearly consider this case absurd. But if this is absurd, then it will imply that also in case 2), the person that is in the reconstructed brain, is not you either. But if is not you, then there is a difference between case 1) and case 2). And the only difference is during the time period when the brain was unconscious. But why should it matters what the state of the brain was (cryogenically frozen or not existing), if it was unconscious anyway?

Thus, we find ourselves in an intricate situation in which no clear conclusion can be drawn. One way out of this is to assume that there is only one Self in the entire universe, and that Self is in all of us. We are all one and the same Self. So in this case, it makes no sense to worry what will happen if you die or if an identical copy of you is made. They are all you anyway. You are immortal, so no worry. Of course, other way out is to assume that one of the premises is wrong. We will talk more about this solution after we would have seen all the thought experiments.

As weird as this experiment might be, let's make it even weirder. Let's introduce a twist in case 1). After your brain was cryogenically frozen, it is then destroyed. And after it was destroyed, it is put back together in the original state. Then everything is good, and after 1 million years, you are brought back to life. Will it still be you in that case? Let's have a closer look at what we just did. By destroying the brain while frozen, we just replicated case 2). But why would it matter anyway? The brain was frozen anyway. So why should it matters what you do with it, if there is no consciousness present there anyway? Is like disassembling a car and then assembling it back. Nothing changes in that case. Why should it matter in the case of the unconscious brain either? But somehow it feels to us that it is not the same thing. Did you die in the process of destroying the brain in case 1) just in the same way you died in case 2)? To our intuition, it somehow feels that is not the same situation.

After all, in case 1) you were unconscious when you died, while in case 2) you died while conscious. But why should it matter the state you were in at the moment of death? A death is a death. And if the brain is reconstructed, the death is reversed and here you are back to life. This twist of the experiment makes cases 1) and 2) more similar to each other. But then there is also case 3) who turn the cases 1) and 2) upside down. One solution is to eliminate case 3) completely by assuming that you cannot make a copy of an object, without destroying it. This would imply

that our Self is uniquely bound to a specific object. If the object is destroyed and then replicated somewhere else, the Self moves safely to the new object. But as we saw earlier, the atoms in our brain are changing continuously throughout our lifetime, but our Self endures. So it is not restricted to only one object. The atoms that were in you when you were a child, are now floating around in the universe. You can in principle gather them up and rebuild the child you. So you just replicated you as a child. But is it you in the newly recreated child? How can this be when you know for sure that you are in the body of a grown up now?

Let's make the experiment even more extreme. Why should we let such a long time to pass? Let's shorten the time interval. How about 1000 years or 1 year? Or let's go for the extreme: 1 nanosecond. In the first case, the brain is cryogenically frozen for 1 nanosecond. In the second case the brain is destroyed for 1 nanosecond and then reconstructed. For such a short period of time, our intuition is again confronted with a difficult decision to make. Both cases seem even more similar now. What is 1 nanosecond? It is so short that our intuition is telling us that the two cases are clearly identical. After all, our brain reaction time to everyday situations is much longer than this. So this brief moment should pass unnoticed. But there is always the third case that seems to present us with an absurd situation of being in two places at once.

It is clear that we cannot get to a conclusion with this experiment. So let's try another approach.

The Gradual Change Experiment

Instead of destroying the brain and rebuilding it, this time we will try a gentle approach. We will change the brain slowly. Let's see 4 cases.

0) You simply pass from state A to state A in a continuous way. The state can be regarded as a person, for example a gradual change from you to you. This is the control case. It is our everyday experience of the Self. A time Δt is passing, while you remain the same Self. This case can be written as $A \rightarrow A$.

1) Your brain is destroyed in state A, and rebuilt back in state A. This is the problematic destruction that we encountered in the previous experiment. In this experiment we try to solve this problem. This case can be written as $A \rightarrow \ldots \rightarrow A$.

2) Your brain is gradually transformed from state A to state B. This might be for example a transformation from you to another person. For example, your neurons are being replaced one by one with the neurons of another person. At the end of the process, you are person B. Since it is a continuous process, I think everyone can agree that the same Self will remain in that brain. Of course, its personality will change. But this change will not bring about another Self. It will simply make the Self to feel a change in its qualia. But throughout this process, the same Self remains. This case can be written as $A \rightarrow B$.

3) Your brain is destroyed in state A, and another brain is constructed in state B. This is weird from the very start. It is the case of you dying of old age for example, and a child being born in another place in the world. It is beyond any doubt that that child cannot be you. You just died. You are lost forever. How can you say that you are that baby? This would imply reincarnation.

Everyone would agree that you cannot be that baby - right? This case can be written as A→…→B.

We went up to another experiment in the hope that we can find a solution for the cryogenic experiment, but we are actually finding ourselves confronted with even weirder ideas, like reincarnation. We would dismiss case 3) right away, if it weren't for case 2). But if a gradual change can change you from a grown up to a child, then why would an abrupt change not do the same thing? For cases 0) and 1), we feel more at ease to accept that after the abrupt change, the Self is recreated. But for cases 2) and 3), our intuition disagrees with the possibility that the Self can also be preserved. But there is actually no difference between cases 0),1) and 2),3). If we accept 0), 1) we should also accept 2), 3).

The Sleep Prank Experiment

Another angle from which the problem of the Self can be regarded is the distinction between a conscious state and an unconscious state. If the gradual change is made while the Self is awake, the Self will be conscious of the gradual change, so at the end of the transformation, the Self knows for sure that he survived. But what if the transformation is done while the Self is sleeping? Let 10 people enter a room and sleep for the night. In the first night, you spare them from a sleep prank. You let them wake up in the same state they went to sleep. None of them will notice anything unusual. They are still themselves and everything is ok. The next night though, you prepared the experiment to mess up with their brains. While they sleep, you cut their brains in pieces and mix them between the 10 people. You can for example take one half of the brain of a person and put it in the head of the next person, and so circularly, such that each head will contain now two halves from two different persons. You can even be more creative and take 10% of each person's brain and put it in each head. This way, each head will contain a part of each of the 10 persons.

The morning then comes and a big mystery arises. Where are the ten original Selves? Who are the new ten Selves? This is simply wrong. Our intuition is completely helpless in predicting what the outcome might be. But take notice here. If the experiment were to be made while the ten persons were awake, we would have no trouble imagining how the Self of each of the ten persons would gradually change to acquire elements of the other ten persons. Why is this difference in our perception? Why a transformation with consciousness turned on seems no problem imagining, while a transformation during the unconscious state seems so mysterious?

The Memories Experiment

In case we are too troubled by this mess of brain mixes, let's only play with one brain. That should be safe. Or should it? This experiment is much less violent than the preceding ones. We don't kill anyone this time. We are just replacing memories. For the first case we will explore what happens to a person who is awake. We will gradually replace that person's memories with some fake memories. In this case, we have no trouble in assuming that the Self endures. Surely,

he will forget his old life and he will start believing he is the new person, but he will still be the same Self.

For the second case though, we will replace the memories of the person while he is sleeping. No point emphasizing that we are again in a troublesome case. Who will be that person that wakes up? In order to make the experiment complete, we will even remove the person from his house and put it in another bed in another city and another country, together with his new memories. When morning will come, he will wake up, and go to his job, greet his colleagues as if they knew each other for a lifetime. It appears that the old Self is dead for good. We are clearly dealing with a fresh new Self. But what about the case when the transformation is done while the Self is awake? Surely the Self survives in that case. So how can an innocent sleep state kill the first Self and give birth to a new Self?

Brain Mergers Experiments

This is a more classical set of experiments that probably all of us wondered about. What happens if you split a brain in two, or if you merge two brains together? Or if you merge all the brains in the world into one brain and then split that brain not in 7 billion pieces, but in 7 million or 70 billion pieces, or if you make all the possible combinations? What these experiments seem to suggest is that the Self is a very malleable entity and that it might not be a clear distinction between 2 Selves.

Continuity/Discontinuity

The problem that seems to arise in all these experiments is the problem of the continuity or discontinuity of the Self. It appears that when a transformation is being done while the Self is awake (so is in a continuous state), the Self survives the transformation. While if the transformation is being done while the Self is unconscious/in a discontinuous state (either asleep or dead), we have trouble imagining who is the person that emerges after the transformation. Is it the same Self? Is it a different Self?

Luckily, we experience both states, so maybe we can find an answer there. While awake, we experience the continuous change and even though our state is changing throughout the day, we are still the same person. We also experience discontinuities each night, but when we wake up, we are still the same Self, even though certain changes are taking place in our brain during sleep. But what about the more serious discontinuity that will await for us at the end of our life?

In what respect does that discontinuity differ from the one that we encounter each night? In the light of the thought experiments presented in this paper, I would suggest that a nice surprise is in store for us at the end of our lives. We may find out that death is only a moment of discontinuity in which the Self encounters a transformation, but is not destroyed. This can be seen from *The memories experiment*. We can ask there how is the case of the sleeping-state Self whose memories are being replaced different from the case when a person dies and another is born in another part of the world? It appears to be no different. Then, if the case of the waken-state Self

whose memories are being replaced is no different from the sleeping-state Self whose memories are being replaced, and the sleeping-state Self whose memories are being replaced is no different from the death-state, then is seems that death is merely a phase where our memories are being replaced, while our Self survives. So death will not be final.

This has even far-fetched implications. This would mean that this is not our first life, but is one in a long series that started with the first conscious being in the universe. And it will not end until the last being of the universe is dead forever. And if there are universes forever, then we are immortal. If we also take into account that when one person dies, a billion other beings are born on Earth and countless more in the universe, then there is no way in which you can choose in which being we will find ourselves after death. It is equally valid if we find ourselves in one being or another. This lead to the conclusive that not only we are immortals, but there is only one Self in the entire Existence. Our present lives are merely small aspects of that enduring unique Self.

Of course, most probably our premises are strongly mistaken. But nevertheless, these ideas of continuity or discontinuity can at least be guidelines in our attempt to understand who we are. As amazing the hypothesis that there is only one Self in the entire Existence, and each of us are that same Self might be, future scientific progress might reveal truths that are more amazing than we can ever imagine. I just hope that by presenting these multiple thought experiments, I succeeded in sparking some new ideas in the mind of the reader.