Think • Vol 22 • No 65 • Autumn 2023, 33–37

DOI: https://doi.org/10.1017/S1477175623000209

Should You Upload Your Mind?

*Sebastian Gäb**

Department of Philosophy, University of Munich
*Corresponding author. Email: gaeb@lrz.uni-muenchen.de

Keywords: upload; mind; brain; survival

Abstract

Could you survive your bodily death by uploading your mind?

Imagine this: you've lived a long and prosperous life, but now you're lying in a hospital bed, terminally ill and dying. Your body feels weak and frail and in a brief moment of clarity among all the pain, anxiety and weariness you realize that this is it. This is the end. This day is your last. Then a well-dressed and professionally concerned looking man approaches you and says: 'We're sorry to hear about your condition, but don't worry. The death of your body is truly unpleasant, but this need not mean that *you* have to die, too. You're not your body – you are your mind. And for a small fee, we can scan your brain, download all the information in it – *you* – and upload it to one of our cloud servers. You'll be able to continue your life indefinitely as a being of pure information, unencumbered by your vulnerable biological body. Just fill in your bank details and sign here.' Would you take that offer?

Well, why should you? Presumably, if you're anything like me, you don't want to die. I like being alive, and I don't want to perish. Taking that offer promises me an opportunity to do just that: stay alive, survive death. And what exactly is it that I want when I say that I want to survive death? Simply this: I want to still be there. Selfish as I am, what matters to me is only whether *I will still be there* after my biological death – or not. So, the crucial question is: will I still be there after the upload?

Proponents of mind-uploading say: yes, and therefore you should take the offer. As you may know, mind-uploading is a procedure in which the complete information stored in the brain is extracted and then transferred into a computer. At the moment, this procedure is purely hypothetical. The whole idea of mind-uploading rests on two fundamental assumptions: (1) a psychological account of persons and personal identity; (2) a computational or patternist theory of mind. According to (1), a person is a being capable of complex mental states like beliefs, desires or imaginations. Also, persons are capable of linguistic communication, rational thought and action, and self-awareness. As you can see, these are all *mental* states and capabilities – the physical shape in which they are instantiated doesn't matter. Any being which shows them will count as a person, and so having a biological body is optional for being a person. So – who am I? I am the person who has these particular thoughts, memories and character traits. A continuous flow of consciousness connects my past, present and

future states of mind and thereby constitutes my identity. I am my mind. According to (2), mental states are essentially computational states. The brain is a computer and the mind is the software it's running. In an ordinary computer, sounds, pictures or texts are coded in a certain series of physical events (the pattern of blocking or allowing the flow of electric current through the transistors). In the same way, my mental states are coded in the pattern of information processing in my brain cells. And since I am my mind, I am this pattern. Now, computational processes are independent of the medium in which they are represented. The sentence 'Roses are red' is informationally the same whether it is coded as a set of ink markings on paper, or as a series of sounds generated by airwaves, or as the flickering of light in Morse code. Information is completely preserved when being transferred from one medium to another. So, if the pattern of information processing is transferred from the brain to a computer, all mental states are preserved and the resulting continuity of these states guarantees that the persons before and after the upload are identical. Even if my body dies, I will survive. I am my mind, my mind is a pattern of information processing (a program, if you will), and this program will continue to run on a different type of computer. Potentially, I am immortal: as long as there is any medium in which these computational processes can continue, I will exist. Or so is the idea. Optimists believe that mind-uploading is a way of survival – pessimists argue that even if mind-uploading works, all it will accomplish is to create a perfect simulacrum of me. The uploaded mind will not be me, but only a self-deluded impostor who thinks he is me, but actually isn't. He (it?) would be *absolutely like me in any way, but still not me*. But if something can be completely *like* me, and still not *be* me, that's not enough, or rather: it's not what matters to me. What matters to me is that I will still be there, and I don't think we have any reason to be sure about that. Here's why.

https://doi.org/10.1017/S1477175623000209 Published online by Cambridge University Press

‘Potentially, I am immortal: as long as there is any medium in which these computational processes can continue, I will exist. Or so is the idea.’

Let’s begin with a thought from Thomas Nagel’s book *The View from Nowhere*. There, he notes that the self is an odd kind of thing. From the inside, it seems to have no connection with any other facts whatsoever, mental or physical. I can perfectly well imagine having a different body from the one I currently have; I can imagine having different experiences or a different character; I can even imagine having lived a completely different life up until now or losing all my memories from one moment to the next – and yet I could still be me. Any and all experiences could potentially be *my* experiences. Let’s say that the self is *opaque*: we know for certain that statements like ‘I am the one who feels this feeling’ are definitely true or false, but we don’t know what makes them true or false – apart from the very fact that I am the one who feels this feeling. I know that I am myself, but I don’t know what it is that I am and why I am myself (and not someone else). All I know is that of all the experiences happening all over the world in all kinds of brains right now, a few of them have a certain quality of *mineness* – for whatever reason. In a way, our use of the pronoun ‘I’ is similar to the use of general terms like ‘gold’ according to a direct theory of reference: we can talk about gold as soon as we have been in contact with gold. It’s not necessary to know what gold actually is to use the term. Likewise, we can successfully use ‘I’ to say things like ‘I am hungry’ and thereby refer to some kind of fact which makes a certain experience of hunger *my* experience, although we have no idea what this fact is.

But then again, the question whether some experience is mine or not seems to have a definite answer. If an uploaded mind continues to have experiences, these experiences will objectively either be mine or not. Sure, it might not be possible to find a set of external criteria to proof this. But subjectively, I know very well what continuity of the uploaded mind would mean: if I press the button on the scanner and begin the upload, I will either continue to have experiences or not. Either my stream of consciousness will keep on flowing into the next moment or it won’t. It’s either like falling asleep and waking up – or like dying (and never waking up). From my internal perspective, it is absolutely clear in which case I persist and in which case I do not (although if I don’t, I won’t be able to notice). From an external perspective, though, it is absolutely *un*clear what determines whether I have survived or not. If you ask the uploaded mind, he (it?) will certainly say that he is me; he’ll remember things only I can remember; he’ll even feel like he is me and vividly remember my past. But he might still be not me.

Now, you’ll probably ask: ‘What is this *me* you’re talking about after all?’ And rightly so. What does it mean to be *me*? Well, I am Sebastian. I may not know what exactly I refer to when I say ‘I’, but at least I know this much, that I am Sebastian, right? No. As Nagel argues, sentences like ‘I am Sebastian’ are neither identity statements nor necessarily true. Sometimes, identity statements *are* necessarily true, if the terms on both sides of the ‘is’ refer necessarily to the same thing, like in: ‘The Morning Star is the Evening Star’. But ‘I am Sebastian’ is not like this – it’s more like ‘I am the vice president of this club’. I simply happen to be Sebastian, just like I happen to be the vice president. The fact that I am this particular person – Sebastian – is ultimately arbitrary. Why Sebastian? I could very well have been someone else, which means: whatever ‘I’ refers to wouldn’t be located in Sebastian’s body, but in another one. Granted, I cannot really imagine what this other body might feel like (especially if it is

https://doi.org/10.1017/S1477175623000209 Published online by Cambridge University Press

very different from my current body), but the mere fact is conceivable (if you have doubts, just get on Netflix and watch the first season of *Altered Carbon*). It's merely a coincidence that this 'I' is attached to this particular body. Nothing about Sebastian's body necessitates that he is *me*. Moreover, it seems as if I could have had different experiences, feelings, or memories as well. In short: I could have had a different mind. I am experiencing Sebastian's mental states, but merely by accident. If I had been a Polynesian fisherman in the eighteenth century instead of a philosopher in twenty-first-century Europe, my mental life would have been completely different. So then, I am not Sebastian's mind, either. I just happen to feel his feelings and think his thoughts, to view the world through his eyes (and mind) like a window, but again, this is just a coincidence. If I were not Sebastian but someone else, I would be experiencing their thoughts and seeing the world through their eyes; but I would still be *me*. But if *me* is not Sebastian, then what is it?

Let's say that what I mean here by 'me' is a *minimal self*. There is a single vantage point from which I observe the whole universe. It is the centre of all my subjective experience, and happens to be situated right here, in Sebastian's body, at this arbitrary and rather unremarkable point in time and space. This is the minimal self, and it is not identical to Sebastian. Why? Imagine the following situation: in the not so far future, technology has made it possible to link brains to each other. There are interfaces in my head and in yours into which we plug our brain-connector. Once we are connected, all your thoughts are replaced by mine. Your brain mirrors my consciousness, so to speak (like two monitors connected to the same computer). Then you will see what I see, you will feel what I feel, but you will not be me. If I see a flower in a vase on the table before me, you'll see the exact same thing and feel exactly as I feel when I see it. But this mental event will be your mental event. Your experience has a different *mineness* from my experience. This is what I mean by minimal self: the specific quality of an experience which makes it immediately and non-inferentially clear that this experience is mine. It is the fact that there is a specific first-person perspective on some experiences.

So, in my little thought experiment, you are seeing the world through my eyes but you are not me. One brain (yours) is in a contingent relation to some other brain. Well, one of them you *call* yours and one of them you call mine, but that's just a convention. I just call the particular brain which is connected to those experiences I experience as mine 'my brain'; but I call it mine because of the *mineness* of these experiences, not the other way round. Mineness doesn't depend on any brain. If we link up our brains, my brain becomes our brain, but my experiences will still be mine. So, the relation between me, namely my minimal self, and any brain is arbitrary. But so is the relation between me and any patterns of information that might make up my mind or make me the person I am.

If this is correct, personal identity is truly not what matters (as Derek Parfit famously claimed). All the facts that make up Sebastian's personal identity and which guarantee his continued existence ultimately don't matter to me. The identity of some arbitrary person is not the same as the persistence of *me* (again, my minimal self), namely my specific vantage point from which I experience the universe. The minimal self is not the personal self, and while I might cease to be the person I am, I cannot cease to be my minimal self. And it is this minimal self I care about when I think about the question whether I will still be there. I don't care about personal identity; I care about *me*. If I ask whether any future experiences will be my experiences, I don't want to know if some being will be physically continuous with my current body; or whether some mental states that will exist are continuous to any mental states I have now; and I also don't care if some immaterial entity that is now connected to me in some way will continue to exist. I only care about whether some of these future experiences will be *mine*, that is, whether these future experiences will have the distinct quality of mineness which my experiences have now.

https://doi.org/10.1017/S1477175623000209 Published online by Cambridge University Press

‘So: *should* you upload your mind? Probably yes. You’re gambling, sure – but it’s not like you’ve got something to lose.’

But if the self is opaque, as I just said, then we don’t know what a minimal self is and under what conditions it will survive any physical or mental changes. In fact, I don’t even know how my minimal self can persist from one day to the next; I just know it does, because I keep on experiencing the *mineness* of my experiences. So, would the uploaded mind be me? Who knows! Minimal selves are distinct (you are not me), but not distinguishable (I don’t know why you are not me). What we would need to answer this question is a theory which explains how the self emerges from physical and/or mental and/or further facts. But we don’t have one. Without such a theory, uploading my mind will be a shot in the dark. I simply have no idea if I will get what matters to me: the continuity of my minimal self and my first-person-perspective. So: *should* you upload your mind? Probably yes. You’re gambling, sure – but it’s not like you’ve got something to lose.

Sebastian Gäb

Sebastian Gäb is Professor of Philosophy of Religion at the University of Munich.

Cite this article: Gäb S (2023). Should You Upload Your Mind? *Think* **22**, 33–37. https://doi.org/10.1017/S1477175623000209