Part 1

Suppose that I can upload my brain into a computer? Will the result be me?¹

On the optimistic view of uploading, the upload will be the same person as the original. On the pessimistic view of uploading, the upload will not be the same person as the original. Of course if one thinks that uploads are not conscious, one may well hold the pessimistic view on the grounds that the upload is not a person at all. But even if one thinks that uploads are conscious and are persons, one might still question whether the upload is the same person as the original.

Faced with the prospect of destructive uploading (in which the original brain is destroyed), the issue between the optimistic and pessimistic view is literally a life-or-death question. On the optimistic view, destructive uploading is a form of survival. On the pessimistic view, destructive uploading is a form of death. It is as if one has destroyed the original person, and created a simulacram in their place.

An appeal to organizational invariance does not help here. We can suppose that I have a perfect identical twin whose brain and body are molecule-for-molecule duplicates of mine. The twin will then be a functional isomorph of me and will have the same conscious states as me. This twin is qualitatively identical to me: it has exactly the same qualities as me. But it is not numerically identical to me: it is not me. If you kill the twin, I will survive. If you kill me (that is, if you destroy this system) and preserve the twin, I will die. The survival of the twin might be some consolation to me, but from a self-interested point of view this outcome seems much worse than the alternative.

Once we grant that my twin and I have the same organization but are not the same person, it follows that personal identity is not an organizational invariant. So we cannot count on the fact that uploading preserves organization to guarantee that uploading preserves identity. On the pessimistic view, destructive uploading is at best akin to creating a sort of digital twin while destroying me.

¹ It will be obvious to anyone who has read Derek Parfit’s Reasons and Persons that the current discussion is strongly influenced by Parfit’s discussion there. Parfit does not discuss uploading, but his discussion of related phenomena such as teletransportation can naturally be seen to generalize. In much of what follows I am simply carrying out aspects of the generalization.
These questions about uploading are closely related to parallel questions about physical duplication. Let us suppose that a teletransporter creates a molecule-for-molecule duplicate of a person out of new matter while destroying or dissipating the matter in the original system. Then on the optimistic view of teletransportation, it is a form of survival, while on the pessimistic view, it is a form of death. Teletransportation is not the same as uploading: it preserves physical organization where uploading preserves only functional organization in a different physical substrate. But at least once one grants that uploads are conscious, the issues raised by the two cases are closely related.

In both cases, the choice between optimistic and pessimistic views is a question about personal identity: under what circumstances does a person persist over time? Here there is a range of possible views. An extreme view on one end (perhaps held by no-one) is that exactly the same matter is required for survival (so that when a single molecule in the brain is replaced, the original person ceases to exist). An extreme view on the other end is that merely having the same sort of conscious states suffices for survival (so that from my perspective there is no important difference between killing this body and killing my twin’s body). In practice, most theorists hold that a certain sort of continuity or connectedness over time is required for survival. But they differ on what sort of continuity or connectedness is required.

There are a few natural hypotheses about what sort of connection is required. Biological theories of identity hold that survival of a person requires the intact survival of a brain or a biological organism. Psychological theories of identity hold that survival of a person requires the right sort of psychological continuity over time (preservation of memories, causally related mental states, and so on). Closest-continuer theories hold that a person survives as the most closely related subsequent entity, subject to various constraints.²

Biological theorists are likely to hold the pessimistic view of teletransportation, and are even more likely to hold the pessimistic view of uploading. Psychological theorists are more likely to hold the optimistic view of both, at least if they accept that an upload can be conscious. Closest-continuer theorists are likely to hold that the answer depends on whether the uploading is destructive, in which case the upload will be the closest continuer, or nondestructive (in which case the biological system will be the closest continuer.³

² There are also primitivist theories, holding that survival requires persistence of a primitive nonphysical self. (These theories are closely related to the ontological further-fact theories discussed later.) Primitivist theories still need to answer questions about under which circumstances the self actually persists, though, and they are compatible with psychological, biological, and closest-continuer theories construed as answers to this question. So I will not include them as a separate option here.

³ In the 2009 PhilPapers survey of 931 professional philosophers [philpapers.org/surveys], 34% accepted or leaned toward a psychological view, 17% a biological view, and 12% a further-fact view (others were unsure, unfamiliar with the issue, held that there is no fact of the matter,
I do not have a settled view about these questions of personal identity and find them very puzzling. I am more sympathetic with a psychological view of the conditions under which survival obtains than with a biological view, but I am unsure of this, for reasons I will elaborate later. Correspondingly, I am genuinely unsure whether to take an optimistic or a pessimistic view of destructive uploading. I am most inclined to be optimistic, but I am certainly unsure enough that I would hesitate before undergoing destructive uploading.

To help clarify the issue, I will present an argument for the pessimistic view and an argument for the optimistic view, both of which run parallel to related arguments that can be given concerning teletransportation. The first argument is based on nondestructive uploading, while the second argument is based on gradual uploading.

The argument from nondestructive uploading.

Suppose that yesterday Dave was uploaded into a computer. The original brain and body was not destroyed, so there are now two conscious beings: BioDave and DigiDave. BioDave’s natural attitude will be that he is the original system and that DigiDave is at best some sort of branchline copy. DigiDave presumably has some rights, but it is natural to hold that he does not have BioDave’s rights. For example, it is natural to hold that BioDave has certain rights to Dave’s possession, his friends, and so on, where DigiDave does not. And it is natural to hold that this is because BioDave is Dave: that is, Dave has survived as BioDave and not as DigiDave.

If we grant that in a case of nondestructive uploading, DigiDave is not identical to Dave, then it is natural to question whether destructive uploading is any different. If Dave did not survive as DigiDave when the biological system was preserved, why should he survive as DigiDave when the biological system is destroyed?

We might put this in the form of an argument for the pessimistic view, as follows:

1. In nondestructive uploading, DigiDave is not identical to Dave.
2. If in nondestructive uploading, DigiDave is not identical to Dave, then in destructive uploading, DigiDave is not identical to Dave.

3. In destructive uploading, DigiDave is not identical to Dave.

and so on). Respondents were not asked about uploading, but on the closely related question of whether teletransportation (with new matter) is survival or death, 38% accepted or leaned toward survival and 31% death. Advocates of a psychological view broke down 67/22% for survival/death, while advocates of biological and further-fact views broke down 12/70% and 33/47% respectively.
Various reactions to the argument are possible. A pessimist about uploading will accept the conclusion. An optimist about uploading will presumably deny one of the premises. One option is to deny premise 2, perhaps because one accepts a closest-continuer theory: when BioDave exists, he is the closest continuier, but when he does not, DigiDave is the closest continuier. Some will find that this makes one's survival and status an unacceptably extrinsic matter, though.

Another option is to deny premise 1, holding that even in nondestructive uploading DigiDave is identical to Dave. Now, in this case it is hard to deny that BioDave is at least as good a candidate as DigiDave, so this option threatens to have the consequence that DigiDave is also identical to BioDave. This consequence is hard to swallow as BioDave and DigiDave may be qualitatively distinct conscious beings, with quite different physical and mental states by this point.

A third and related option holds that nondestructive uploading should be regarded as a case of fission. A paradigmatic fission case is one in which the left and right hemispheres of a brain are separated into different bodies, continuing to function well on their own with many properties of the original. In this case it is uncomfortable to say that both resulting systems are identical to the original, for the same reason as above. But one might hold that they are nevertheless on a par. For example, Parfit (1984) suggests although the original system is not identical to the left-hemisphere system or to the right-hemisphere system, it stands in a special relation R (which we might call survival) to both of them, and he claims that this relation rather than numerical identity is what matters. One could likewise hold that in a case of nondestructive uploading, Dave survives as both BioDave and DigiDave (even if he is not identical to them), and hold that survival is what matters.

Still, if survival is what matters, this option does raise uncomfortable questions about whether DigiDave has the same rights as BioDave when both survive.

(Stop here and read the rest later)
Part 2

The argument from gradual uploading.

Suppose that 1% of Dave’s brain is replaced by a functionally isomorphic silicon circuit. Next suppose that another 1% is replaced, and another 1%. We can continue the process for 100 months, after which a wholly uploaded system will result. We can suppose that functional isomorphism preserves consciousness, so that the system has the same sort of conscious states throughout.

Let $\text{Dave}_n$ be the system after $n$ months. Will $\text{Dave}_1$, the system after one month, be Dave? It is natural to suppose so. The same goes for $\text{Dave}_2$ and $\text{Dave}_3$. Now consider $\text{Dave}_{100}$, the wholly uploaded system after 100 months. Will $\text{Dave}_{100}$ be Dave? It is at least very natural to hold that it will be. We could turn this into an argument as follows.

1. For all $n < 100$, $\text{Dave}_{n+1}$ is identical to $\text{Dave}_n$.
2. If for all $n < 100$, $\text{Dave}_{n+1}$ is identical to $\text{Dave}_n$, then $\text{Dave}_{100}$ is identical to Dave.

3. $\text{Dave}_{100}$ is identical to Dave.

On the face of it, premise 2 is hard to deny: it follows from repeated application of the claim that when $a = b$ and $b = c$, then $a = c$. On the face of it, premise 1 is hard to deny too: it is hard to see how changing 1% of a system will change its identity. Furthermore, if someone denies premise 1, we can repeat the thought-experiment with ever smaller amounts of the brain being replaced, down to single neurons and even smaller. Maintaining the same strategy will require holding that replacing a single neuron can in effect kill a person. That is a hard conclusion to accept. Accepting it would raise the possibility that everyday neural death may be killing us without our knowing it.

One could resist the argument by noting that it is a sorites or slippery-slope argument, and by holding that personal identity can come in degrees or can have indeterminate cases. One could also drop talk of identity and instead hold that survival can come in degrees. For example, one might hold that each $\text{Dave}_n$ survives to a large degree as $\text{Dave}_{n+1}$ but to a smaller degree as later systems. On this view, the original person will gradually be killed by the replacement process. This view requires accepting the counterintuitive view that survival can come in degrees or be indeterminate in these cases, though. Perhaps more importantly, it is not clear why one should accept that Dave is gradually killed rather than existing throughout. If one were to accept this, it would again raise the question of whether the everyday replacement of matter in our brains over a period of years is gradually killing us also.
My own view is that in this case, it is very plausible that the original system survives. Or at least, it is plausible that insofar as we ordinarily survive over a period of many years, we could survive gradual uploading too. At the very least, as in the case of consciousness, it seems that if gradual uploading happens, most people will become convinced that it is a form of survival. Assuming the systems are isomorphic, they will say that everything seems the same and that they are still present. It will be very unnatural for most people to believe that their friends and families are being killed by the process. Perhaps there will be groups of people who believe that the process either suddenly or gradually kills people without them or others noticing, but it is likely that this belief will come to seem faintly ridiculous.

Once we accept that gradual uploading over a period of years might preserve identity, the obvious next step is to speed up the process. Suppose that Dave’s brain is gradually uploaded over a period of hours, with neurons replaced one at a time by functionally isomorphic silicon circuits. Will Dave survive this process? It is hard to see why a period of hours should be different in principle from a period of years, so it is natural to hold that Dave will survive.

To make the best case for gradual uploading, we can suppose that the system is active throughout, so that there is consciousness through the entire process. Then we can argue: (i) consciousness will be continuous from moment to moment (replacing a single neuron or a small group will not disrupt continuity of consciousness), (ii) if consciousness is continuous from moment to moment it will be continuous throughout the process, (iii) if consciousness is continuous throughout the process, there will be a single stream of consciousness throughout, (iv) if there is a single stream of consciousness throughout, then the original person survives throughout. One could perhaps deny one of the premises, but denying any of them is uncomfortable. My own view is that continuity of consciousness (especially when accompanied by other forms of psychological continuity) is an extremely strong basis for asserting continuation of a person.

We can then imagine speeding up the process from hours to minutes. The issues here do not seem different in principle. One might then speed up to seconds. At a certain point, one will arguably start replacing large enough chunks of the brain from moment to moment that the case for continuity of consciousness between moments is not as secure as it is above. Still, once we grant that uploading over a period of minutes preserves identity, it is at least hard to see why uploading over a period of seconds should not.

As we upload faster and faster, the limit point is instant destructive uploading, where the whole brain is replaced at once. Perhaps this limit point is different from everything that came before it, but this is at least unobvious. We might formulate this as an argument for the optimistic view of destructive uploading.
Here it is to be understood that both the gradual uploading and instant uploading are destructive in that they destroy the original brain.

1. Dave survives as $\text{Dave}_{100}$ in gradual uploading.
2. If Dave survives as $\text{Dave}_{100}$ in gradual uploading, Dave survives as $\text{DigiDave}$ in instant uploading.

3. Dave survives as $\text{DigiDave}$ in instant uploading.

I have in effect argued for the first premise above, and there is at least a prima facie case for the second premise, in that it is hard to see why there is a difference in principle between uploading over a period of seconds and doing so instantly. As before, this argument parallels a corresponding argument about teletransportation (gradual matter replacement preserves identity, so instant matter replacement preserves identity too), and the considerations available are similar.

An opponent could resist this argument by denying premise 1 along the lines suggested earlier, or perhaps better, by denying premise 2. A pessimist about instant uploading, like a pessimist about teletransportation, might hold that intermediate systems play a vital role in the transmission of identity from one system to another. This is a common view of the ship of Theseus, in which all the planks of a ship are gradually replaced over years. It is natural to hold that the result is the same ship with new planks. It is plausible that the same holds even if the gradual replacement is done within days or minutes. By contrast, building a duplicate from scratch without any intermediate cases arguably results in a new ship. Still, it is natural to hold that the question about the ship is in some sense a verbal question or a matter for stipulation, while the question about personal survival runs deeper than that. So it is not clear how well one can generalize from the ship case to the shape of persons.

*Where things stand.*

We are in a position where there are at least strongly suggestive arguments for both the optimistic and pessimistic views of destructive uploading. The arguments have diametrically opposed conclusions, so they cannot both be sound. ... My instincts favor optimism here, but as before I cannot be certain which view is correct.

Still, I am confident that the safest form of uploading is gradual uploading, and I am reasonably confident that gradual uploading is a form of survival. So if at some point in the future I am faced with the choice between uploading and continuing in an increasingly slow biological embodiment, then as long as I have the option of gradual uploading, I will be happy to do so.
Unfortunately, I may not have that option. It may be that gradual uploading technology will not be available in my lifetime. It may even be that no adequate uploading technology will be available at all in my lifetime. This raises the question of whether there might still be a place for me, or for any currently existing humans, in a post-singularity world. …

Reconstructive uploading.

The final alternative here is reconstruction of the original system from records, and especially reconstructive uploading, in which an upload of the original system is reconstructed from records. Here, the records might include brain scans and other medical data; any available genetic material; audio and video records of the original person; their writings; and the testimony of others about them. These records may seem limited, but it is not out of the question that a superintelligence could go a long way with them. Given constraints on the structure of a human system, even limited information might make a good amount of reverse engineering possible. And detailed information, as might be available in extensive video recordings and in detailed brain images, might in principle make it possible for a superintelligence to reconstruct something close to a functional isomorph of the original system.

The question then arises: is reconstructive uploading a form of survival? If we reconstruct a functional isomorph of Einstein from records, will it be Einstein? Here, the pessimistic view says that this is at best akin to a copy of Einstein surviving. The optimistic view says that it is akin to having Einstein awake from a long coma.

Reconstructive uploading from brain scans is closely akin to ordinary (nongradual) uploading from brain scans, with the main difference being the time delay, and perhaps the continued existence in the meantime of the original person. One might see it as a form of delayed destructive or nondestructive uploading. If one regards nondestructive uploading as survival (perhaps through fission), one will naturally regard reconstructive uploading the same way. If one regards destructive but not nondestructive uploading as survival because one embraces a closest continuer theory, one might also regard reconstructive uploading as survival (at least if the original biological system is gone). If one regards neither as survival, one will probably take the same attitude to reconstructive uploading. Much the same options plausibly apply to reconstructive uploading from other sources of information. …