

Virtual Reality and the Self

Week 8, Unit 3 Knowledge, Reality and VR

PHI3VRP Virtual Reality as Philosophy

Introduction

- What I'm covering in the video lectures this week:
- 1. Uploading and the Philosophy of Mind
- 2. Philosophy of Personal Identity
- 4. Uploading Again

Uploading and the Philosophy of Mind

Week 8: Video Lecture 1

PHI3VRP Virtual Reality as Philosophy

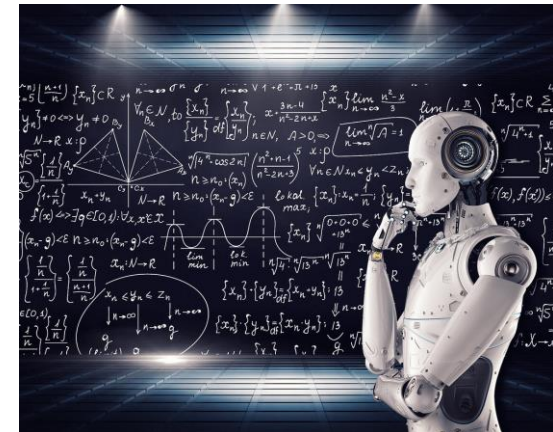
Uploading

- Could you be uploaded?
 - “An *upload* is a creature that has its thoughts and sensations transferred from a physiological basis in the brain to a computational basis in computer hardware. An upload could have a virtual or simulated body or even be downloaded into an android body — indeed, an upload could even primarily exist in the digital world, and merely occupy an android body when needed.” (Corabi and Schneider 2012: 26)



Uploading

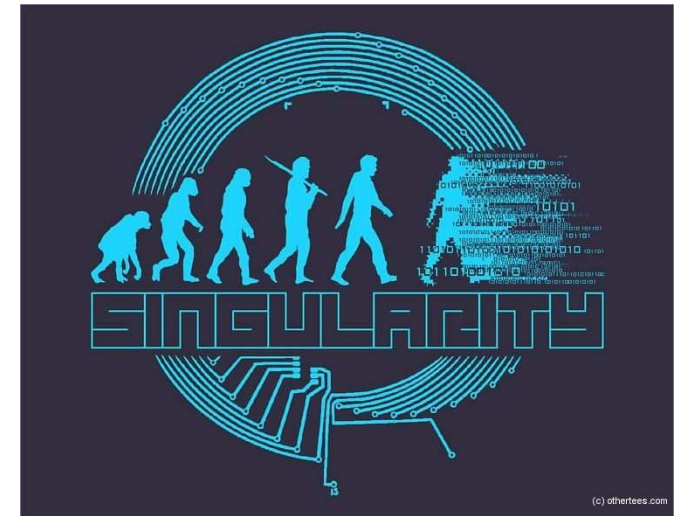
- Why be interested in this question? One reason some thinkers are interested in this issue is that see uploading as potentially the best response we would have to a **singularity** event. E.g. Chalmers (2010)
- A singularity is rapid expansion in artificial intelligence (AI) from intelligence to a super intelligence that far outstrips human intelligence.
 - “Let an ultraintelligent machine be defined as a machine that can far surpass all the intellectual activities of any man however clever. Since the design of machines is one of these intellectual activities, an ultraintelligent machine could design even better machines; there would then unquestionably be an ‘intelligence explosion’, and the intelligence of man would be left far behind. Thus the first ultraintelligent machine is the last invention that man need ever make.” ‘I.J. (1965; see Chalmers 2010 in the further readings for discussion and the reference)



[This Photo](#) by
Unknown Author
is licensed under
[CC BY](#)

Uploading

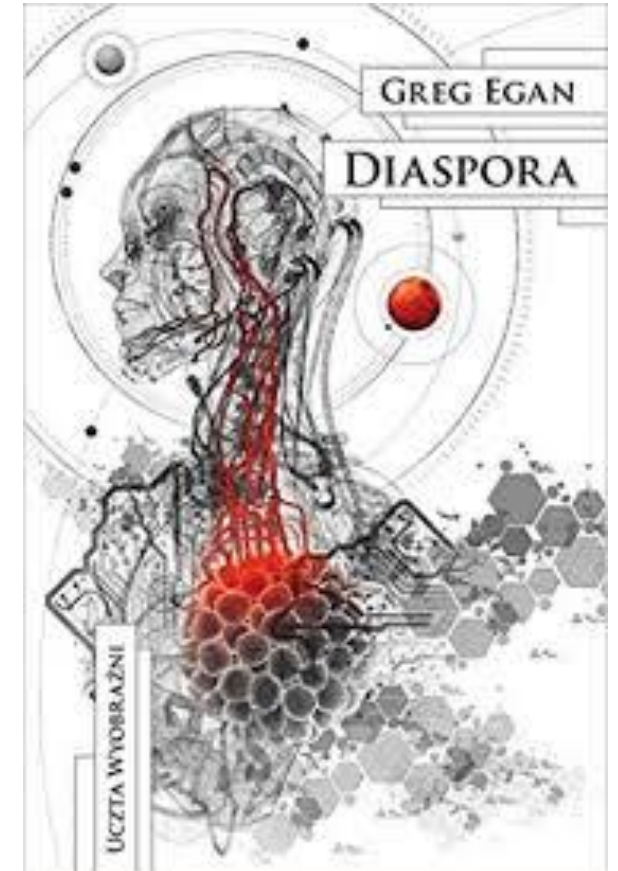
- Chalmers thinks that faced with a super-intelligent AI we would face four choices: extinction, isolation, inferiority, or integration. Of these choices, Chalmers thinks that integration is the only one that would be both workable and attractive. And mind uploading would be the way to achieve that **integration**.
- “This process of migration from brain to computer is often called *uploading*. Uploading can make many different forms. It can involve gradual replacement of brain parts (gradual uploading), instant scanning and activation (instant uploading), or scanning followed by later activation (delayed uploading). It can involve destruction of the original brain parts (destructive uploading), preservation of the original brain (nondestructive uploading), or reconstruction of cognitive structure from records (reconstructive uploading).”
Chalmers (2010: 41-42)



[This Photo](#) by Unknown Author is licensed under [CC BY](#)

Uploading

- A fictional depiction of **destructive uploading** by the Australian sci-fi writer Greg Egan in his book *Diaspora*:
- “Waves of nanoware were sweeping through Orlando’s body, shutting down nerves and sealing off blood vessels to minimize the shock of invasion, leaving a moist pink residue on the rubble as flesh was read and then cannibalized for energy. Within seconds, all the waves converged to form a grey mask over his face, which bored down to the skull and ate through it. The shrinking nanoware spat fluid and steam, reading and encoding crucial synaptic properties, compressing the brain into an ever-tighter description of itself, discarding redundancies as waste.
- Inoshiro stooped down and picked up the end product: a crystalline sphere, a molecular memory containing a snapshot of everything Orlando had been. (Egan, 1997, p. 110)



Uploading

- Other reasons to be interested in the uploading question: concerns with survival after (physical) death, utopian visions (e.g. Danaher's virtual utopia).
- And, we can also be interested in the uploading question as a tool for thinking about what we are *now*—that is, as a tool for thinking about the nature of the mind and the self.



Uploading

- Minimal conception of mind uploading: the functional structure of a biological human brain is replicated in a digital computer.
- Not clear that this is possible but let us grant, for the sake of argument, that it is.
- Two importantly different sets of questions:
- Would the resulting digital system be conscious? Would it have all the same mental states as the original?
- Assume that the resulting system would be conscious and would contain mental states of all the same types etc. Would the uploaded mind be *your* mind? Would this uploaded mind be you or a mere duplicate of you?



[This Photo](#) by Unknown Author is licensed under [CC BY-SA-NC](#)

Are the mental and the physical the same kind of thing?

Yes

Monism

No

Dualism

Idealism

The mental is fundamental and the physical can be analysed in terms of the mental

Physicalism (Non-eliminative)

The physical is fundamental and the mental can be analysed in terms of the physical

The Identity Theory

For any given mental state M: M is identical to some brain/neural state B

Functionalism

For any given mental state M: M is a functional state that can, in principle, be *multiply realized* by different types of physical states

Philosophy of Personal Identity

Week 8: Video Lecture 2

PHI3VRP Virtual Reality as Philosophy



What makes it the case that a person at one time is the same person as a person at another time? What kind of thing is a self/person?

Personal Identity

- What is a person? One way to approach this question: try to informatively fill in the dots below:
- A person x at time t is identical to person y at time t' if and only if ...
- Identity: 'identity' here means **numerical identity** (x and y are *one and the same* thing) **not** mere **qualitative identity** (x and y are distinct things which share a significant number of properties).
- Examples: Mark Twain and Samuel Clemens are numerically identical. When we say that two twins are identical we mean qualitative identity.

Identity

- Some notable facts about (numerical) identity:
- Identity is a **transitive relation**. A relation R is a transitive relation just in case the following is true (for any a , b , and c): if aRb and bRc then aRc .
 - Transitive relations: “is greater than” and “equal to” in math, “is an ancestor of”
 - Non-transitive relations: “not equal to” in math, “loves”, “is the daughter of”
- So, to say that identity is a transitive relation is to say that if x is identical to y and y is identical to z then x is identical to z
- Identity is the relation **that everything bears to itself and nothing else**.
- The **indiscernibility of identicals**: if x is identical with y then x and y share all and only the same properties.

Personal Identity

- **Bodily View:** a person x at time t is identical to y at time t' iff x at t has the same body as y at t' .
- **Brain View:** a person x at time t is identical to y at time t' iff x at t has the same brain as y at t' .
- **Psychological View:** a person x at time t is identical to y at time t' iff x at t is psychologically continuous with y at t' .
- **Soul View:** a person x at time t is identical to y at time t' iff x at t has the same soul as y at t' .
- **No self view:** there are no selves and, hence, there is no identity of a self over time. There is only the *illusion* of these things.

Stop and think

- **Bodily View:** a person x at time t is identical to y at time t' iff x at t has the same body as y at t' .
- **Brain View:** a person x at time t is identical to y at time t' iff x at t has the same brain as y at t' .
- **Psychological View:** a person x at time t is identical to y at time t' iff x at t is psychologically continuous with y at t' .
- **Soul View:** a person x at time t is identical to y at time t' iff x at t has the same soul as y at t' .
- **No self view:** there are no selves and, hence, there is no identity of a self over time. There is only the *illusion* of these things.

- **Questions:**
- Which of these views seems most plausible to you and why?
- Can you think of any objections to one or more of these views?
- What could psychological continuity involve? Suppose we said that it is just matter of the later person stage *remembering* being the earlier person stage? What problems might that memory account of psychological continuity face?

Reid's criticism of Locke's version of the psychological view

- “Suppose a brave officer to have been flogged when a boy at school, for robbing an orchard, to have taken a standard from the enemy in his first campaign, and to have been made a general in advanced life: Suppose also, which must be admitted to be possible, that when he took the standard, he was conscious of his having been flogged at school, and that when made a general he was conscious of his taking the standard, but had absolutely lost the consciousness of his flogging.
- These things being supposed, it follows, from Mr LOCKE's doctrine, that he who was flogged at school is the same person who took the standard, and that he who took the standard is the same person who was made a general. When it follows, if there be any truth in logic, that the general is the same person with him who was flogged at school. But the general's consciousness does not reach so far back as his flogging, therefore, according to Mr LOCKE's doctrine, he is not the person who was flogged. Therefore the general is, and at the same time is not the same person as him who was flogged at school.”
(*Essays*, 276)

The Psychological View (PV)

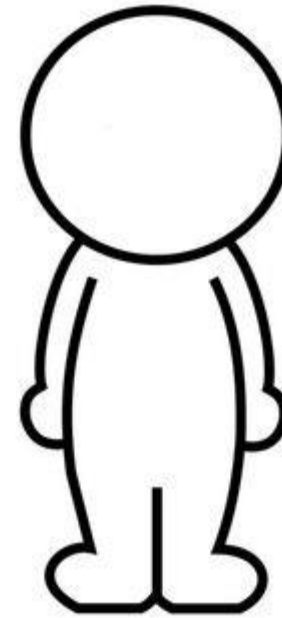
- Psychological *connectedness*:
- x at t and y at t' are **psychologically connected** iff x at t and y at t' have psychological states (beliefs, desires, intentions, apparent memories, character traits, etc.) which are (a) similar in content, and (b) causally connected in the right way, i.e. the psychological states of the later person stage causally depend (in the right way) for their character on the states of the earlier one.

The Psychological View (PV)

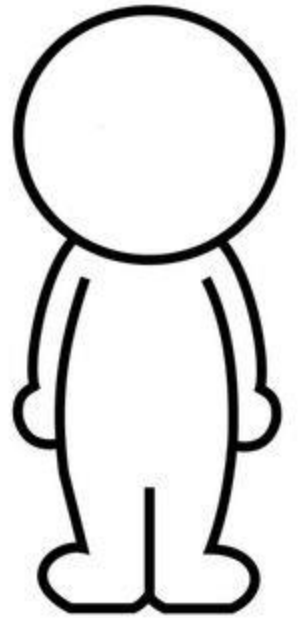
- *Psychological continuity* is the ancestral of psychological connectedness. That is, person-stage x is **psychologically continuous** with y if there is a step-by-step path from x to y which consists of person-stages each of which is psychologically connected to its immediate predecessor and immediate successor.
- I am psychologically connected to my 5-second-ago self.
- I am psychologically continuous with (but not psychologically connected to) my 5-year-old self.

The Psychological View (PV)

- Common motivation for the Psychological View: 'transplant' thought experiments, e.g.:
- Suppose your brain is successfully transplanted from your body A into another body B. Body A is then destroyed. The resulting person with body B will have all of your beliefs, desires, apparent memories, character traits, etc.. Note that this person herself will think she is you (she has your apparent memories and beliefs), and she will act and behave just like you (she has your belief, desire, and intentions). But is she you? Or did you die when body A was destroyed?
- http://www.bbc.co.uk/radio4/today/reports/archive/science_nature/brain.shtml



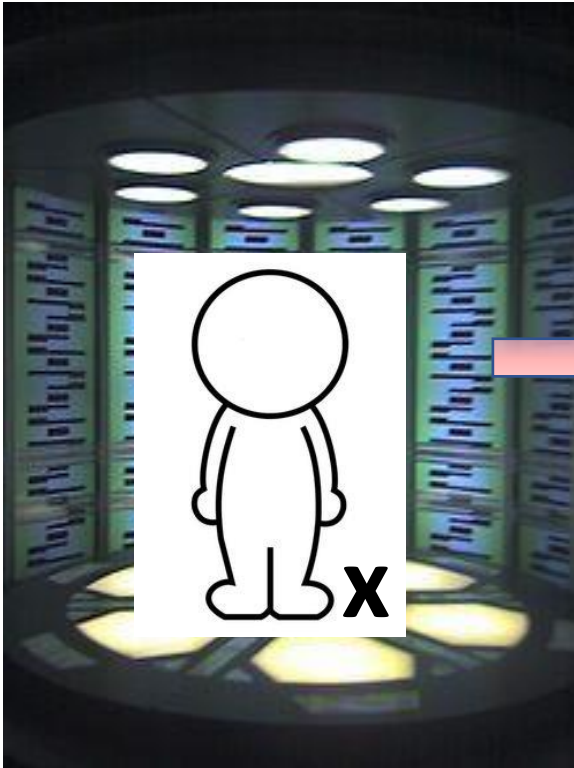
A



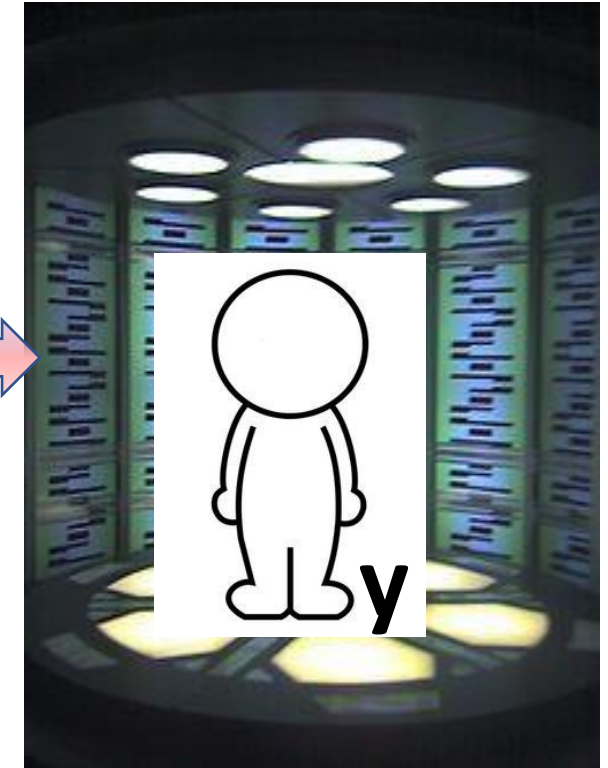
B

- Motivation for just the Psychological View (and *not* also the brain view): teletransportation.
- Is X at t_1 identical to Y at t_2 ?
- The psychological view tells us yes because X is psychologically continuous with Y. And, arguably, this is the intuitively correct thing to say about the case.

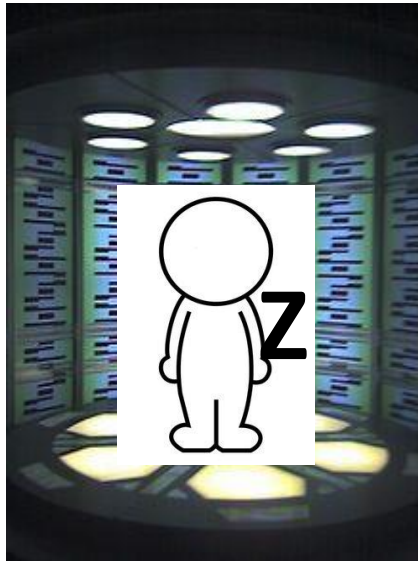
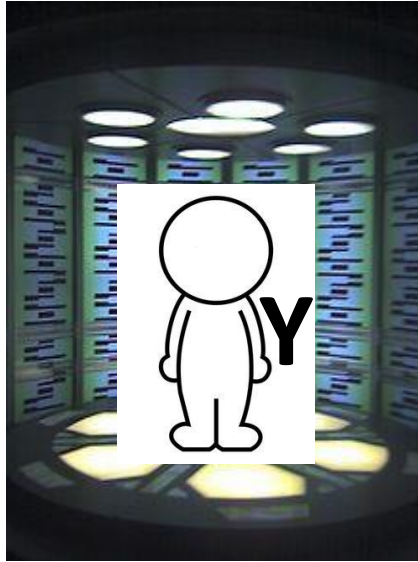
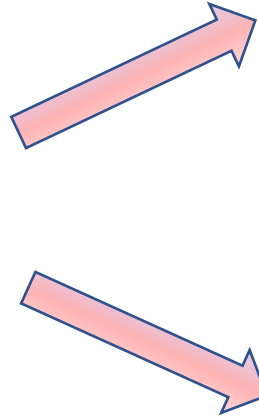
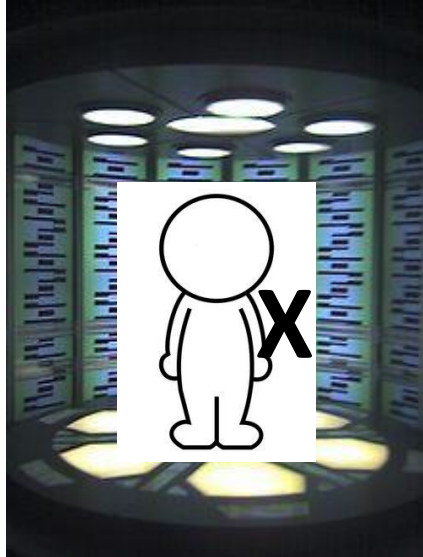
t^1



t^2



t¹



t²

- X is psychologically continuous with Y. So, if the psychological view is true then $X=Y$.
- But X is also psychologically continuous with Z. So, if the psychological view is true then $X=Z$.
- But then, by the transitivity of identity, $Y=Z$! Which is absurd.

Duplication/Fission worries

- So, duplication cases like this appear to raise serious problems for psychological views of PI. **Possible reply:** Teletransportation cases are just mere sci-fi. There is no way anyone could ever build such a machine so the duplication worry is not a serious worry.
- **Counter:** brain hemisphere transplant cases involving split brains ('fission' cases) can be used to generate the same kind of problem cases.
 - Suppose that we could divide my brain, and transplant each hemisphere into a different body. And suppose that the two resulting persons both had beliefs, desires, intentions, apparent memories, and character traits which were similar to mine. In short, suppose that both resulting persons were psychologically continuous with me.
 - See: <https://www.scientificamerican.com/article/strange-but-true-when-half-brain-better-than-whole/>

Parfit on Identity and Survival

- Derek Parfit adopts a **non-branching version of the psychological view**. That is, he thinks that I would survive teletransportation or a brain hemisphere transplant only if there was no duplication or fission (i.e. one person in, one person out). So fission (/duplication) is *death*.
- But then whether or not I survive a hemisphere transplant depends on whether or not both (as opposed to just one) of the hemispheres is transplanted. How can that be? “How could a double success lead to failure?”. Violates Wiggins's “only x and y” principle:
 - Whether or not x and y are stages of the same person should depend only on facts about how x and y are related to each other, not about how either is related to something else.

Parfit on Identity and Survival

- Parfit accepts these consequences. On his view situation in which one or more future persons is psychologically continuous with me is just as good (other things being equal) as a situation in which I continue to exist. Since in the former type of situation, I wouldn't continue to exist—I wouldn't be identical to any of those future person—Parfit says that **identity is not what matters**. What matters is there being some future person that is psychologically continuous with me, whether or not he is actually me.
- I would survive a single hemisphere transplant but not survive fission; but this doesn't matter much because surviving in this sense—i.e. being identical to something after fission—is not what matters.
- I should care about there being some future person who has my psychological properties, not about there being some future person who is identical to me.

Uploading Again

Week 8: Video Lecture 3

PHI3VRP Virtual Reality as Philosophy

Uploading again

- Assuming that we could create a functional duplicate of you in a virtual world running on a computer then we can ask two questions:
- (1) Would it be a mental duplicate of you? (So, it would be conscious and have all the same mental states)
- (2) Would it be you?
- Let us grant, at least for the sake of argument, that the answer to (1) is 'yes'. What is the answer to (2)?



Uploading and Personal Identity

- If we accept the body or brain view of PI then it is not possible that the person in the virtual world is you.
- If we accept the psychological view then uploading seems, in principle, possible as long as (i) that there would be the right kind of connections between the earlier biologically realized mind and the later computer realized mind, (ii) the original body and brain is destroyed, and we might also need to require (iii) only one uploaded mind is created.



Chalmers argument for pessimism

- “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
- (Chalmers 2010: 51)



Chalmers argument for pessimism

- 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.
- (Chalmers 2010: 51)



Chalmers argument for optimism: step 1

- “Suppose that 1% of Dave’s brain is replaced by a functionally isomorphic silicon circuit. Next suppose that after one month another 1% is replaced, and the following month 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 Dave n be the system after n months. Will Dave₁, the system after one month, be Dave? It is natural to suppose so. The same goes for Dave₂ and Dave₃.” (Chalmers 2010: 52)



Chalmers argument for optimism: step 1

- 1. For all $n < 100$, Dave _{$n+1$} is identical to Dave _{n} .
- 2. If for all $n < 100$, Dave _{$n+1$} is identical to Dave _{n} ,
- then Dave₁₀₀ is identical to Dave.
- —————
- 3. Dave₁₀₀ is identical to Dave.
- (Chalmers 2010: 52)



Chalmers argument for optimism: step 2

- 1. Dave survives as Dave₁₀₀ in gradual uploading.
- 2. If Dave survives as Dave₁₀₀ in gradual uploading, Dave survives as DigiDave in instant uploading.
- —————
- 3. Dave survives as DigiDave in instant uploading.



Chalmers' conclusion

- “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 own view is that the optimist’s best reply to the argument from nondestructive uploading is the fission reply, and the pessimist’s best reply to the argument from gradual uploading is the intermediate-case reply. My instincts favour 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.”
(Chalmers 2010)

