

The Simulation Hypothesis

Week 7, Unit 3 Knowledge, Reality and VR

PHI3VRP Virtual Reality as Philosophy

Introduction

- What I'm covering in the video lectures this week:
- 1. Simulations and scepticism in the history of philosophy
- 2. The Simulation Argument
- 3. Does it matter if we're living in a simulation?

Simulations and scepticism in the history of philosophy

Week 7: Video Lecture 1

PHI3VRP Virtual Reality as Philosophy

Zhuangzi's Butterfly

“Once the Chinese philosopher Zhuangzi dreamed he was a butterfly, a butterfly flitting and fluttering about, happy with himself and doing as he pleased. He didn't know that he was Zhuangzi. Suddenly he woke up and there he was, solid and unmistakably Zhuangzi. But he didn't know if he was Zhuangzi who had dreamt he was a butterfly, or a butterfly dreaming that he was Zhuangzi.”
(Watson translation 1968, p. 46)



Philosophy thought experiments in drawings, by
@HelenDeCruz

Zhuangzi (about 400 BCE) was dreaming he was a butterfly. When he awoke, he did not know whether he was a man dreaming to be a butterfly, or a butterfly who is now dreaming it is a man.

Plato's Cave



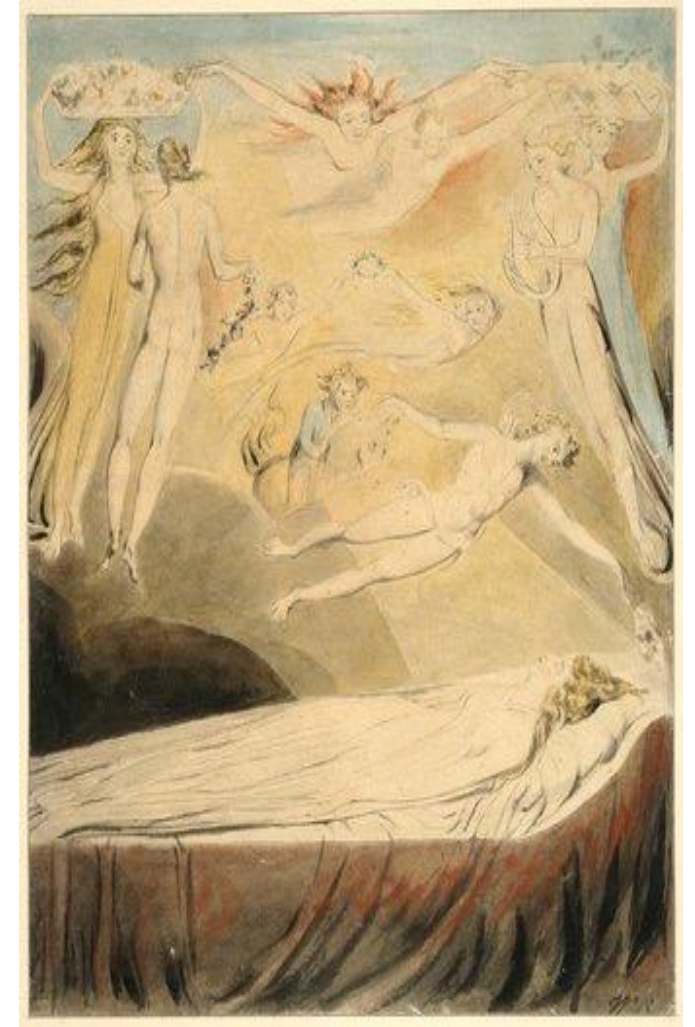
Narada and Vishnu



Descartes' Dreaming Hypothesis

“How often have I dreamt that I was in these familiar circumstances, that I was dressed, and occupied this place by the fire, when I was lying undressed in bed? At the present moment, however, I certainly look upon this paper with eyes wide awake; the head which I now move is not asleep; I extend this hand consciously and with express purpose, and I perceive it; the occurrences in sleep are not so distinct as all this. But I cannot forget that, at other times I have been deceived in sleep by similar illusions; and, attentively considering those cases, I perceive so clearly that there exist no certain marks by which the state of waking can ever be distinguished from sleep, that I feel greatly astonished; and in amazement I almost persuade myself that I am now dreaming.”

Descartes, *Meditations* I.5



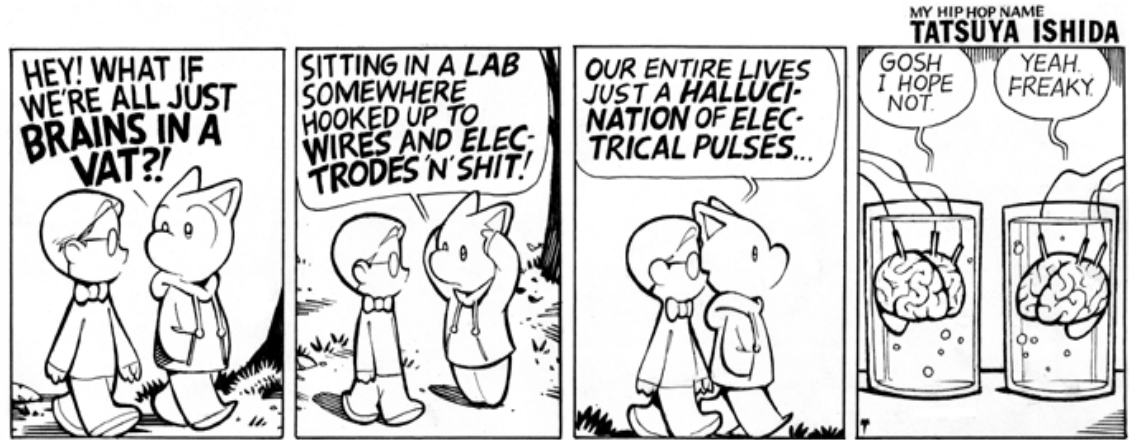
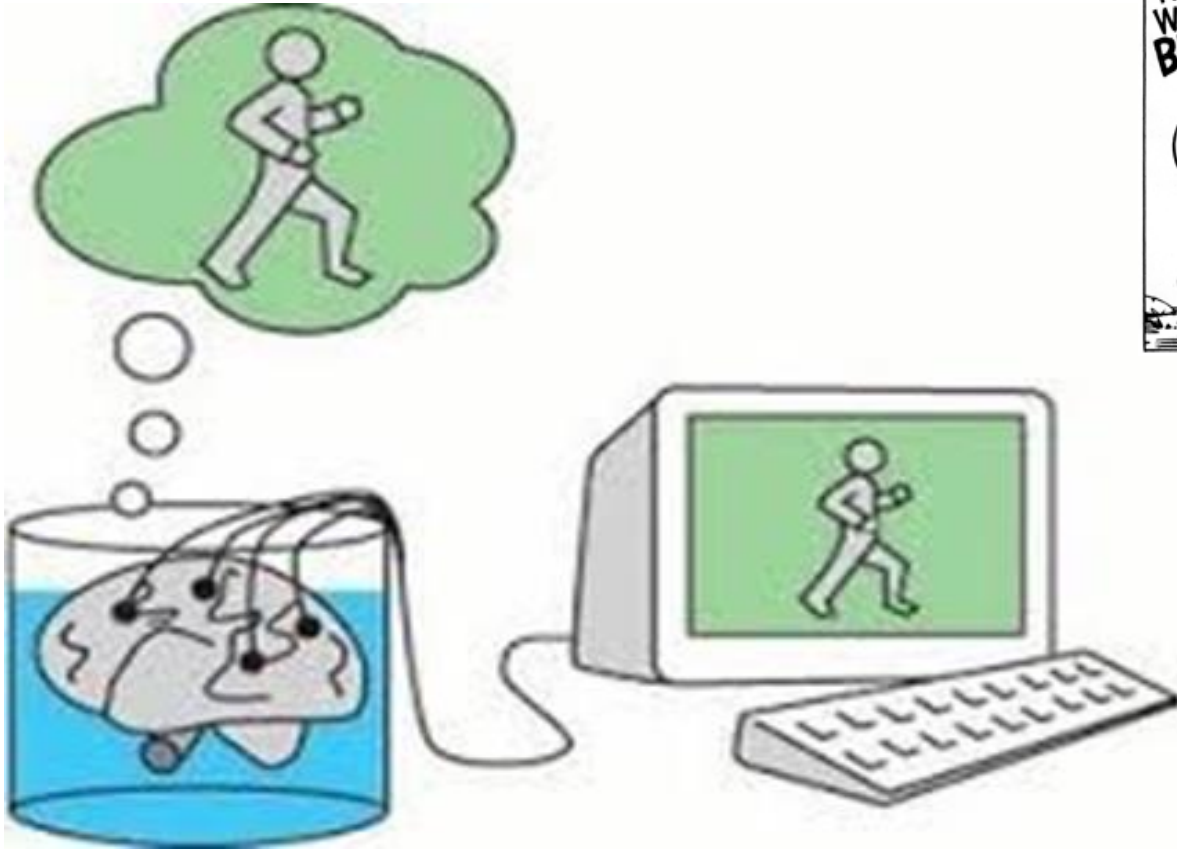
Descartes' Demon

“ will suppose...that some malignant demon, who is at once exceedingly potent and deceitful, has employed all his artifice to deceive me; I will suppose that the sky, the air, the earth, colors, figures, sounds, and all external things, are nothing better than the illusions of dreams, by means of which this being has laid snares for my credulity; I will consider myself as without hands, eyes, flesh, blood, or any of the senses, and as falsely believing that I am possessed of these; I will continue resolutely fixed in this belief, and if indeed by this means it be not in my power to arrive at the knowledge of truth, I shall at least do what is in my power, viz, [suspend my judgment], and guard with settled purpose against giving my assent to what is false, and being imposed upon by this deceiver, whatever be his power and artifice.”

Descartes, *Meditations* 1. 12



The Brain in a Vat Hypothesis



The Sceptic's Argument

- (1) You don't know that you're not a Brain in a Vat (BIV)
- (2) If you don't know that you're not a BIV, then you don't know that you have hands (for example).
- (3) Therefore, You don't know that you have hands.



The Sceptic's Argument

- Some important things to note:
- The sceptic is **not** claiming that you are a BIV, and they are **not** even claiming that it is *likely* that you are a BIV.
- The sceptic is claiming that you cannot know that you are not a BIV and then using that premise as part of their argument for the conclusion that most (if not all) of your supposed knowledge about the external world is not actually knowledge at all.



Questions raised by simulation thought experiments

- **Knowledge:** What can we know? Can we know that we're not living in a simulation?
- **Reality:** Would a simulated reality be less real than a non-simulated reality?
- **Value:** What is the value of being in contact with reality? Is a simulated life always of less value than one lived in reality?



The Simulation Argument

Week 7: Video Lecture 2

PHI3VRP Virtual Reality as Philosophy

Bostrom's simulation argument

- Nick Bostrom (2003) gave a very influential argument for the conclusion that it is **there is a significant probability that the simulation hypothesis** (i.e. the hypothesis that we are living in a computer generated simulation) **is true**.
- Note the differences here from the sceptic's BIV and evil demon arguments:
- The sceptic's argument only relies on the claim that we cannot know that we are not living in a certain kind of simulation scenario. Bostrom is arguing that it is actually likely that we are living in a computer generate simulation.
- Also, unlike the BIV scenario, if Bostrom's simulation hypothesis is true our brain is also a part of the computer generated simulation.
- *Mixed simulation and reality* scenarios: the BIV scenario, the Matrix scenario
- *Full simulation* scenarios: Bostrom's simulation hypothesis

The Substrate Independence Assumption

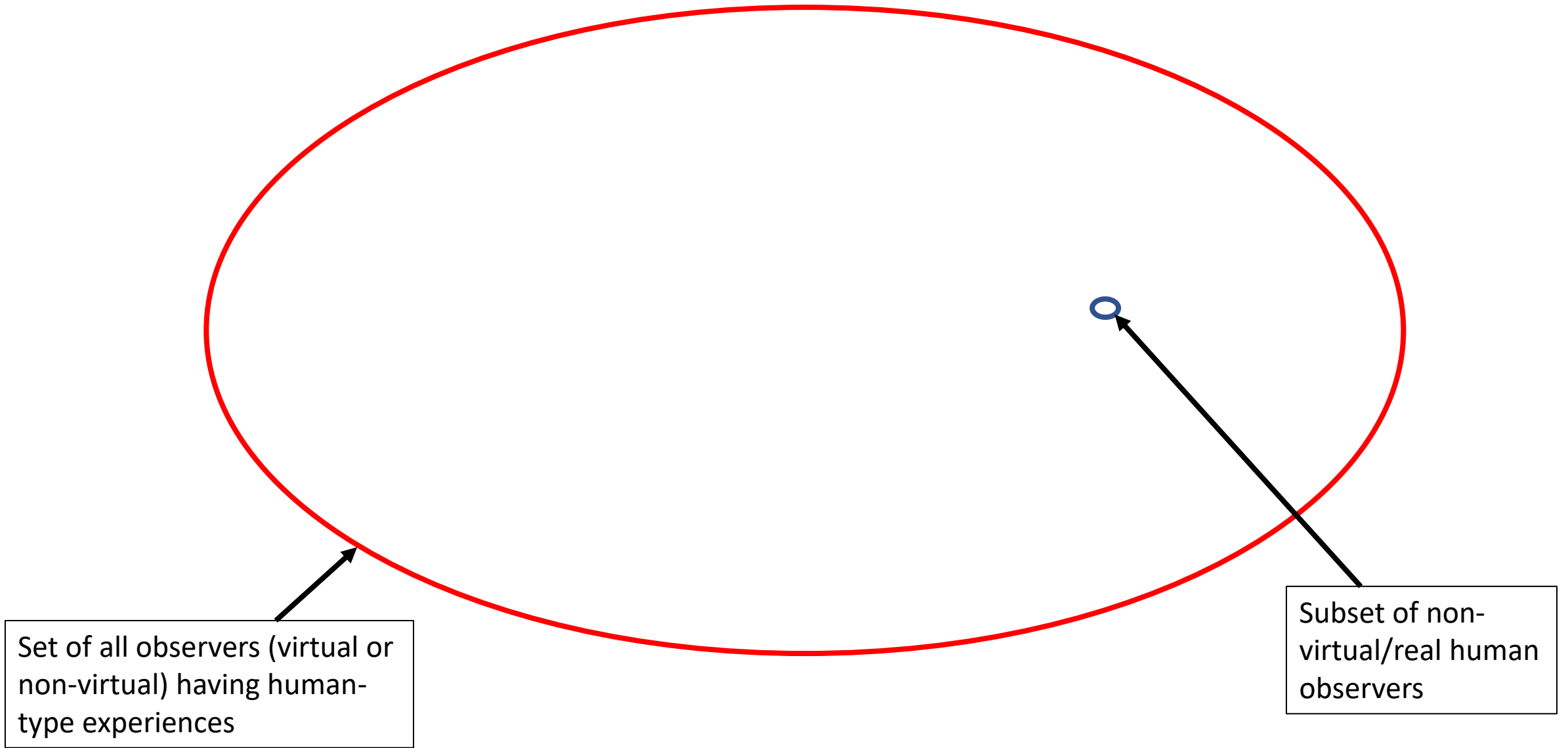
- “A common assumption in the philosophy of mind is that of *substrate independence*. The idea is that mental states can supervene on any of a broad class of physical substrates. Provided a system implements the right sort of computational structures and processes, it can be associated with conscious experiences. It is not an essential property of consciousness that it is implemented on carbon-based biological neural networks inside a cranium: silicon-based processors inside a computer could in principle do the trick as well.” (Bostrom 2003: 244)

- In the philosophy of mind it is widely assumed that mental states and processes are **multiply realizable**, that is, the same mental state type (say a belief or a visual experience, or...) can be realized by very different kinds of underlying substances (a human brain, an animal brain, a silicon brain, an alien brain).
- Comparisons:
- Babbage’s Analytic Engine, versus Colossus, versus modern computers.
- Coke machines, air conditioners, ovens, bank tellers,...
- Neural implants

Bostrom's technological assumptions

- If the substrate independence assumption is true then a human mind could be realized not only by a biological brain but by a sufficiently powerful computer.
- We can't do that with the computers we have now but we have good reasons to believe that a more technologically advanced society in the future could do that. Furthermore:
 - “Post-human civilizations would have enough computing power to run hugely many ancestor-simulations even while using only a tiny fraction of their resources for that purpose.” (Bostrom 2003: 248)
- Comparisons with the rapid development of computers, and computer simulations, to date are often used to support these assumptions.





The Simulation Argument

- Bostrom presents his argument as an argument that at least one of the following three claims must be true:
 1. The chances that a species at our current level of development can avoid going extinct before becoming technologically mature is negligibly small.
 2. Almost no technologically mature civilisations are interested in running computer simulations of minds like ours.
 3. You are almost certainly in a simulation.
- Bostrom claims if (1) and (2) are both *false* then it follows that (3) is true. In which case, if (3) is false then either (1) or (2) is true [note when I gave the lecture this accidentally said the incorrect 'false']

The Simulation Argument

- “Suppose that proposition (1) is false. Then a significant fraction of all species at our level of development eventually becomes technologically mature. Suppose, further, that (2) is false, too. Then some significant fraction of these species that have become technologically mature will use some portion of their computational resources to run computer simulations of minds like ours. But, as we saw earlier, the number of simulated minds that any such technologically mature civilisation could run is astronomically huge.
- Therefore, if both (1) and (2) are false, there will be an astronomically huge number of simulated minds like ours. If we work out the numbers, we find that there would be vastly many more such simulated minds than there would be non-simulated minds running on organic brains. In other words, almost all minds like yours, having the kinds of experiences that you have, would be simulated rather than biological. Therefore, by a very weak principle of indifference, you would have to think that you are probably one of these simulated minds rather than one of the exceptional ones that are running on biological neurons.
- So if you think that (1) and (2) are both false, you should accept (3). It is not coherent to reject all three propositions. In reality, we do not have much specific information to tell us which of the three propositions might be true. In this situation, it might be reasonable to distribute our credence roughly evenly between the three possibilities, giving each of them a substantial probability.” (Bostrom 2009)

Stop and think:

The Simulation Argument

1. The chances that a species at our current level of development can avoid going extinct before becoming technologically mature is negligibly small.
2. Almost no technologically mature civilisations are interested in running computer simulations of minds like ours.
3. You are almost certainly in a simulation.
 - Bostrom claims if (1) and (2) are both *false* then it follows that (3) is true. In which case, if (3) is false then either (1) or (2) must be true.

- Questions:
- Can you think of any reasons to support (1)?
- Can you think of any reasons to support (2)? Perhaps any technologically mature civilisation would refuse to run ancestor simulations for ethical reasons?
- Bostrom relies on the assumption that sufficiently powerful computer simulation of a human's brain would have the same kind of mental states and consciousness that the original human has. Can you think of any potential reasons for disputing that assumption?

Does it matter if we're living in a simulation?

Week 7: Video Lecture 3

PHI3VRP Virtual Reality as Philosophy

Ethics and the simulation hypothesis

- If we think that there is a strong chance that we are living in a simulation should that change our views about the value of the people and objects in the world around us?
 - Nozick's discussion of the experience machine suggests a view on which we should conclude that if we are living in a simulation than our life has less value.
- Should it change our views about our ethical obligations?
 - "If you might be living in a simulation then all else equal it seems that you should care less about others, live more for today, make your world look likely to become eventually rich, expect to and try to participate in pivotal events, be entertaining and praiseworthy, and keep the famous people around you happy and interested in you." (Hanson 2001)
- But if at least some of the other simulated characters around us are also conscious beings then we would still have ethical obligations to them, even if we also have prudential reasons to act as Hanson suggests.

God and the simulation hypothesis

- “Although all the elements of such a system can be naturalistic, even physical, it is possible to draw some loose analogies with religious conceptions of the world. In some ways, the posthumans running a simulation are like gods in relation to the people inhabiting the simulation: the posthumans created the world we see; they are of superior intelligence; they are ‘omnipotent’ in the sense that they can interfere in the workings of our world even in ways that violate its physical laws; and they are ‘omniscient’ in the sense that they can monitor everything that happens.” (Bostrom 2003: 253-254)
- "The Simulation Argument is perhaps the first interesting argument for the existence of a Creator in 2000 years." David Pearce



Knowledge and the simulation hypothesis

- Suppose, for the sake of argument, that the simulation hypothesis is true.
- It would seem to follow that most of the beliefs we ordinarily assume to be knowledge are in fact false (and hence cannot be knowledge).
- However, Chalmers in his paper “The Matrix as Metaphysics” gives an intriguing argument against this assumption. We’ll look closely at these issues in Week 9 but, roughly, the idea is that if the simulation hypothesis were true then all the objects and people in our environment still exist and are perfectly real, its just that their underlying nature is surprisingly different from what we had assumed it to be.
- My belief that I have hands is still true, its just that the fundamental nature of a hand is computational.

