Why we are not living in the computer simulation[1]

Abraham Lim

PhD student of the department of philosophy, University of Cologne, Germany

kenzhi12@gmail.com

Abstract

Nick Bostrom considered a number of simulations and contended that the probability that we are living in one of them is high or at least nonzero. I present arguments to refute the claim that we are or might be in any one of them.

Keywords:

Simulation, Brain-in-vat scenario, BIV argument, Nick Bostrom, Tim Button, Hilary Putnam

1. The Easy Simulation

Nick Bostrom put forward the famous claim that we are most likely living in the computer simulation. Bostrom presented this claim in a form of trilemma consisting of three disjuncts:

> (I) the human species is very likely to go extinct before reaching a 'posthuman' stage;
>
> (II) any posthuman civilization is extremely unlikely to run a significant number of simulations of their evolutionary history (or variations thereof);
>
> (III) we are almost certainly living in a computer simulation.

By 'posthuman', Bostrom meant the future generation whose technology is advanced enough to allow them to produce a lot of simulations of their ancestors. And he argued that at least one of the disjuncts is true, and we must at least believe one of them. More crucially, he aimed to argue: "if we don't think that we are currently living in a computer simulation, we

[1] This is a draft for a forthcoming paper which will be published on the International Journal for the Study of Skepticism.

are not entitled to believe that we will have descendants who will run lots of such simulations of their forebears."[2]

One of my aims in this paper is to show that Bostrom's trilemma is not compelling. Specifically, I aim to show that not only are we entitled to have the belief that our descendants will in the future run a lot of simulations of their ancestors without having to have the belief that we are almost certainly living in one of these simulations, but we would also have contradictory beliefs if we have these two beliefs.

Bostrom's argument for his trilemma was roughly as follows. It is predicted by futurists or scientists that enormous amounts of computing power will be available in the future. Later generations will therefore most likely run a great many detailed simulations of their forebears. He further supposed that these simulated people are conscious, assuming that a certain quite widely accepted position in the philosophy of mind is correct (Substrate-Independence is meant here, it will be elaborated later). Then, with some calculation of probability, Bostrom concluded that we are likely among the simulated minds rather than among the original biological ones.

At least two assumptions were made in Bostrom's argument:

> **Substrate-independence**: Provided a system implements the right sort of computational structures and processes, it can be associated with conscious experiences.
>
> **The prospect of computing power evidenced by physical laws we discover**: Running many simulations of human minds would be within the capability of a future civilization whose technologies would be compatible with known physical laws and engineering constraints.

With at least these two assumptions, Bostrom contended that unless we will not reach posthuman stage and will not run a lot of simulations of our forebears, we are actually most likely in the following scenario:

[2] See Bostrom (2003:1).

The Easy Simulation. Simulated beings are eternally simulated by the future generation using advanced technology, which can already be envisioned with the aid of the currently known physical laws.

I call this simulation the *Easy simulation* because the idea that we might be in this simulation is easy to refute, as will be shown in a moment. Many versions of simulation will be discussed later. And when it is clear from the context which simulation I will be discussing, I will sometimes simply call the one in question 'the simulation'.

2. The Refutation of the Easy Simulation

I will refute the idea that we might be in the Easy simulation by refuting the idea that we are in it. And the realization that the idea that we are in it is incoherent would compel us to refute the idea that we are in it.

Without directly questioning the substrate-independence principle, I am going to show in two ways that we cannot make sense of the idea that we are in the Easy simulation.

To make sense of the idea that we are in the Easy simulation, the stricture of the simulation requires us to invoke the physical law to support our thought that we ourselves are in the simulation. But we must first note that physical law is the outcome of inference or induction from a group of observations of reality. Yet no one in the simulation ever observes the reality or comes to make any contact with it, let alone makes any inference or induction from those observations. Thus, if we want to invoke physical laws to back up the idea that we are in the simulation, we must concede that we have been in the reality to have made contact with it.

But, to concede ourselves as being in the reality in order for us to invoke the physical law as evidence to support the idea that we are in the simulation, is to land in incoherence. Consequently, there is no way for us to make sense of the idea that we are in the Easy simulation.

The second way to show that the idea is incoherent concerns the notion of future generation or descendants. And this is much more straightforward: since all the conscious beings can have descendants only if they are not simulated, one cannot be simulated and at the same time can have descendants. Hence, it is incoherent for anyone to think of herself as being in the Easy simulation run by the descendants of her race. This again shows:

Incoherence Claim: The idea of oneself being in the Easy simulation is incoherent.

Thus, we must concede that we are not in the Easy simulation.[3] The idea that we might be in it is to be refuted as well. And if the belief that our descendants will run a lot of simulations of their forebears is true, then the belief that we are one of those simulated forebears is false. And we are not compelled to believe any one of the three disjuncts Bostrom proposed even if we believe our descendants will run a lot of simulations of their forebears.

Notice that what this refutation shows is that we cannot be inside the Easy simulation. This does not entail that the Easy simulation cannot at all exist. It may someday—in the past, the present, or the future—exist somewhere in the universe. And as Bostrom argues, the emergence of the Easy simulation might be inferred from our currently known physical law, and conscious beings might be simulated by our future generation. All of these might be granted without any reservation. But, emphatically, we ourselves cannot be inside the Easy simulation.

3. The Hard Simulation

The refutation that concerns the notion 'future generations', though an obvious and straightforward one, was curiously, largely or completely missing in the literature. There are papers that discuss the notion 'currently known physical law' and utilize it to criticise the idea that we are in the Easy simulation. Some of these lines of thought come quite near what I have just presented.[4]

[3] Although David Chalmers' reasoning that concludes that we might be in a simulation in Chalmers (2022) is not patently subject to the refutations I have given here, it still comes very close to being subject to them since he at many points does not hold back to explicitly suggest that the upshot of the future development of our simulation technology is that we might be already living in a simulation, which is very similar to Bostrom's reasoning, see for example Chalmers (2022: Introduction, Ch 2, Ch 5). And even though Chalmers' reasoning with his conclusion might eventually be able to avoid the refutations I give in this section, it can certainly be refuted by the *SIM argument* I am going to develop in the rest of this paper.

[4] See Besnard (2004) and Birch (2012).

In view of these criticisms, Bostrom retreated from the Easy simulation to the following:[5]

Simulated beings are eternally simulated by some civilization using advanced technology that is based on some physical law unknown to the simulated beings.

By avoiding the concepts of ‘currently known physical law’ and of ‘future generations’, it might be thought that the claim that we might be in this scenario is a coherent one and cannot be refuted.

A common, but implicit, presupposition in presenting this scenario is that if we are simulated, we have neither had any conscious communication with the simulator, or in general, with any normal individuals in the reality, nor been (consciously) aware of any influence exerted by the latter. So, the more specific formulation of the simulation Bostrom would have in mind should be the following:

The Hard Simulation. Simulated beings are eternally simulated by some civilization using advanced technology that is based on some physical law unknown to the simulated beings. *And no simulated being has ever had any conscious communication with the unsimulated beings or has ever noticed any influence exerted by the latter.*

The idea that we might be in the *Hard simulation* is hard to refute, hence its name. The Hard simulation covers the situation in which the simulated beings (from here on, I may sometimes call them the SIMs) are unconsciously manipulated by the simulators and some information coming from ‘the world outside the simulation’ might be transmitted to the SIMs without them noticing the transmission. But this scenario does not cover the situation in which some conscious communication takes place between the SIM and the normal individual(s).

What could this kind of communication be like? When addressing the question whether the claim that we are in a simulation is testable, Bostrom suggested an example of some conscious communication:

[5] See Bostrom (2008).

> There are clearly possible observations that would show that we are in a simulation. For example, the simulators could make a "window" pop up in front of you with the text "YOU ARE LIVING IN A COMPUTER SIMULATION. CLICK HERE FOR MORE INFORMATION." Or they could uplift you into their level of reality.[6]

This interaction Bostrom envisaged between the normal individuals and the SIMs are apparently some kind of communication, both parties of which are aware or conscious of the content that is being communicated.

However, this example of communication between the normal individuals and the simulated beings can be shown to be utterly incoherent. To show this, I must and will show that we are not in the Hard simulation. By the stricture of the Hard simulation, the conscious communication between the simulated beings and the normal individuals is not allowed. Thus, the simulated beings in the Hard simulation cannot notice the influence on them exerted by the simulators.

As we will later see, determining what interaction could take place between the SIMs and the normal individuals in a given simulation is essential for determining how this simulation can be handled, especially, for determining whether we are in it. This is the reason why the kind of interaction Bostrom would presumably, implicitly, assume between them must be specified at the first place before we can proceed to determine whether the idea of us being in the simulation or some other scenario can be refuted.

4. The Brain-in-Vat Scenario and the BIV argument

My refutation of the claim that we might be in the Hard simulation builds on the refutation of the claim that we might be in the *brain-in-vat scenario (BIV scenario),* given by Hilary Putnam. It would be easier to understand the former refutation if we first understand the latter refutation. The BIV scenario is as follows: [7]

[6] Ibid.

[7] Button (2013: 117), Putnam (1981: Ch 1).

The Brain in Vat Scenario. All sentient creatures are eternally envatted brains. That is, for the entire duration of their lives, they were, are, and will always be brains in vats. However, everyone is wired into an infernal machine, which subjects them all to electronic neural stimulations, so that everything appears normal.

Again, if we can refute the idea that we are in the BIV scenario, then the idea that we might be in the BIV scenario is also refuted. There are many versions of the refutation of the idea that we are in this scenario. One version is particularly elegant and compelling, presented by Tim Button. Call a sentient creature in this scenario a BIV. The refutation is as follows:[8]

(1) A BIV's word 'brain' does not refer to brains.
(2) My word 'brain' refers to brains.
(3) So: I am not a BIV.

This refutation is called by Button the *BIV argument*. It is obviously valid. It remains to justify the premises. Let us start with (2). (2) is undeniable. I am going to show why it is so by trying to deny it. That is, let us assert: "my word 'brain' does not refer to brains." To meaningfully assert this denial of (2), I must presuppose that the last word in the assertion does refer to brains. This shows that when asserting the denial of (2), I must presuppose (2) itself. Thus, denying (2) is self-refuting and therefore, (2) is not to be denied.

Note that no empirical fact, including any fact about what a brain is, is invoked in showing that the denial of (2) is self-refuting. And we can generalize (2) by introducing a principle that governs our reference practice:

Disquotation (DQ): My term 'X' refers to X.

As we have seen, from the first-person point of view, any instance of DQ cannot be coherently denied, regardless of what the nature of X is. Hence, DQ cannot be denied from the first-person point of view.[9]

[8] Button (2013: Ch12).

[9] See van Fraassen (1997, 2008: 229-235), Button (2013: 123-127, 2016: 136-138).

We next justify (1). Let Brian be a brain in a vat (or a BIV) in the BIV scenario. I am going to show that Brian's word 'brain' does not refer to brains.

The straightforward way for Brian's word 'brain' to refer to brains would be for Brian to be a brain scientist. But this is prohibited by his being a BIV. He cannot conduct any empirical research when he is a BIV.

An alternative way for his term 'brain' to refer to brains would be by interacting with scientists or science teachers who have some experience of studying brains and thereby have some understanding about brains. But since everyone in the BIV scenario is a BIV, none of them can be a scientist or a science teacher, and therefore none of them can gain the experience of interacting with brains, let alone the experience of studying brains.[10]

By the stricture of BIV scenario, none of the BIVs can ever interact with the reality, part of which are brains. And exactly for this reason, any BIV's word 'brain' cannot refer to brains. And this implies (1).

5. The Limitations of Responding to Scepticisms with a BIV-style Argument

The scenarios studied throughout this paper are usually regarded as sceptical scenarios. Presumably, there are two features that jointly qualify a scenario to be a sceptical one. First, these scenarios present worldviews which are radically different from what we usually believe the real world is like, yet second, there seems no way for us to detect from our experience whether we are actually living in the scenario accompanied by a radically different worldview or in the world we normally believe we live in. The scenarios we have discussed so far——the Easy and Hard simulations, the BIV scenario——all have these two features, and therefore are the paradigmatic cases of sceptical scenario.

Having seen the BIV argument refuting the claim that we are in the BIV scenario, we may wonder if the BIV argument, with some modifications, can refute the claim that we are in any other given sceptical scenario. In fact, many philosophers agree that it is not an easy task to

[10] Button considered several other possible ways for Brian's word 'brain' to refer to brains, and showed why they failed as well. The reasons for their failures effectively boil down to what I have presented here, namely, the total insulation of BIVs from the reality. For more details, see Button (2013: 118-120).

determine whether the claim that we are in the following sceptical scenario can be refuted by deploying a BIV-style argument:[11]

Recent Envatment Scenario. Some of the sentient beings were recently captured and envatted by some scientist(s).

Call any recently envatted brain *REB*. The BIV-style argument against the idea that I might be one of the REBs would be:

(1) An REB's word 'brain' does not refer to brains.
(2) My word 'brain' refers to brains.
(3) So: I am not an REB.

I have shown that (2) is warranted regardless of any situation. The problem of this argument is obviously with (1). Unlike a BIV who has never had the experience of learning any word or of gaining the experience of using any word, an REB, before having become an REB, might have acquired the ability to use language competently. And that ability might allow the REB's word 'brain' to refer to brains even after having become envatted. Moreover, another complication is that although an REB might lose language ability if being envatted for too long, it is also an indeterminate issue as to how long one can be envatted before she starts to lose her language ability.

In any case, the truth of (1) is indeterminate. And it follows that it is an indeterminate matter whether I can successfully conclude that I am not an REB by deploying a BIV-style argument.

Now let us compare Bostrom's Hard simulation with another sceptical scenario Button considered:[12]

[11] Wright (1992: 89-90), Pritchard (2016: 190n4). Cf. Button (2013: 158-159).

[12] Button (2013: 156)

The Vat Earth scenario. Earth has a distant neighbour, Vat Earth. This is a planet whose only inhabitants are eternally envatted brains. There is no relevant causal link between Earth and Vat Earth, but it so happens that, for every brain on Earth, there is a brain on Vat Earth in exactly the same state, and vice versa.

The first thing to notice about the Vat Earth scenario is the existence of the normal individuals in it, who never become envatted. This is one significant aspect that distinguishes this scenario from the original BIV scenario. The existence of normal individuals is what both the Vat Earth scenario and Bostrom's Hard simulation have in common. This existence is certainly a complication we need to consider if we set out to refute the ideas that we are in any of these scenarios. Concerning the Vat Earth scenario, Button reckoned:

> Our sceptical worry ensues: am I an embodied creature on Earth, or an envatted brain on Vat Earth? The crucial premise of our inevitable BIV-style response to the sceptic will be:
>
> > (VE) A vat-earthling's word 'brain' does not refer to brains.
>
> If there is never any connection between Earth and Vat Earth, we might be inclined to say that there is no reason to interpret vat-earthlings as speaking a language that is any different from Brian's language. In which case, we will embrace (VE) for the same reason that we embrace the original premise (1) of the BIV argument.
>
> Granting this, if only temporarily, we can simply change the scenario slightly. Let us move Earth closer to Vat Earth (metre by metre, if it matters). Let us imagine that earthlings visit Vat Earth (in ones, or twos, or threes, if the number of visitors will make a difference). Let us imagine that the earthlings tap on the glass walls of the vats. They pity the state of the 'deluded' brains. And let us suppose now that some native vat-earthlings, and an equal number of visitors from Earth, simultaneously think the following words:
>
> > If I were envatted and someone were looking at me right now through the glass walls of my vat, I would want to be judged according to the language of the embodied, in thinking 'I am envatted'.

> Does this really make no difference? Again, I boggle at the attempt to *start* answering that question.[13]

In the Vat Earth scenario, the normal individuals are not equipped with the power to manipulate the deluded, which is clearly different from the normal individuals in Bostrom's simulation. And given this manipulation exerted on the simulated beings by the simulator(s), the idea that we might be in the Hard simulation is apparently harder to refute than the idea that we might be in the Vat Earth scenario. And if Button boggles at the attempt to start handling the Vat Earth scenario, he most likely would have the same response to the Hard simulation.

David Chalmers explicitly contends that the BIV argument cannot refute the idea that we might be in a simulation. However, the reason Chalmers gives is problematic. He thinks that to refute the idea that we might be in a simulation, the BIV argument must require the premise that the simulated being's word 'simulation' does not refer to simulations. Yet Chalmers thinks this premise cannot be established, for he observes that one's referent of the word 'simulation', unlike one's referent of the word 'brain', does not depend on one's environment.[14]

I am not sure if Chalmers's observation is correct. But even if it is, the reason he gave misses one of the important points of BIV argument, which is to show there essentially exists some aspect of a given sceptical scenario our situation lacks, or vice versa. If the BIV argument can succeed to show that some word, which could be the word 'brain' or any other word, instead of 'simulation', used by any sentient being living inside a given sceptical scenario does not refer to a certain item, and that same word used by us does refer to that item, then this would suffice to show that our situation is not identical to that sceptical scenario.

In any case, it is clear from the discussions above that the BIV argument by itself cannot refute the idea that we might be in the Hard simulation. Nevertheless, in what follows, I am

[13] Button (2013: 156-157).

[14] See Chalmers (2022: 410). Chalmers did not specify what kind of simulation he was discussing. If the simulation scenario he had in mind does not contain normal individuals, then, by the BIV argument, his contention is utterly wrong. If, instead, the simulation scenario contains normal individuals, then as stated in the main text, he gave a problematic reason for his contention.

going to refute the idea that we might be in the Hard simulation, both by invoking much of the line of thoughts already established by Button, and by adding a completely new component that is designed exclusively for addressing the existence of normal individuals in the Hard Simulation. Once the refutation is complete, we will see that the refutation of the idea of us being in the Vat Earth scenario is just an immediate corollary.

6. The First Stage of the Refutation of the Hard Simulation

My refutation of the idea that we might be in the Hard simulation contains three parts. The first part of this refutation is to show that the semantic condition of the conscious simulated beings is virtually the same as that of the BIVs. That is, there is virtually no way for the conscious simulated being's word 'brain' to refer to brains.

To begin, let Simon be a conscious simulated being in the Hard simulation. The straightforward way for Simon's word 'brain' to refer to brains would be for Simon to be a scientist. But this is excluded by his being in the simulation. He cannot conduct any empirical research when he is totally insulated from the reality.

An alternative way for his word 'brain' to refer to brains would be by interacting with scientists or science teachers who have some experience of studying brains and thereby have some understanding about brains. But anyone around him cannot be a scientist or a science teacher for the same reason that they are all totally insulated from reality, and therefore none of them can gain the experience of studying brains.

Simon simply lacks adequate resources that enable his word 'brain' to refer to brains.

7. The Second Stage: A Potential Influence Exerted by the Simulator(s)

Since, by the condition of the Hard simulation, there exist normal individuals who run the simulations, we apparently manage to conceive of another way that enables the simulated beings' word 'brain' to refer to brains. It is as follows:

> **Manipulation (MA)**: The kind of manipulation, unnoticeable by the simulated beings, produced by the simulators utilizing the physical law, that enables the simulated beings to

use the language the normal individuals use, and that enables the simulated beings' word 'brain' to refer to brains.

The idea of MA is specifically designed to be the potentially maximal interaction that could take place between the simulated beings and the reality without them being aware of the influence from the world outside the simulation. MA therefore conforms to the stricture of the Hard simulation. If MA is feasible, then Simon's word 'brain' can eventually refer to brains.[15]

I am going to show, even if MA can occur in the Hard simulation, Simon cannot succeed to coherently articulate or coherently conceive of the idea that he might be in the simulation.

Suppose MA occurs in the Hard simulation. And suppose Simon, like us, is investigating if he might be in the Hard simulation. We can unfold the following thought process, or the ratiocination, that arises in Simon's mind concerning the idea that he himself might be in the simulation:

[15] Cf. Hammarstrom (2008: 44).

The Thought Process (TP)

Simon has this thought:

(i) “I might be in the Hard simulation.”

Since Simon, like us, has been contemplating the idea that he might be in the Hard simulation, and has been thinking through this matter up to this point, it is natural for him to wonder: “what does my word ‘brain’ refer to?”

Simon reasons: “independently of whether I am in the Hard simulation, given DQ, I must concede:

(ii) My word ‘brain’ refers to brains.”

With this in his mind, Simon entertains the following thought:

(iii) “The mass of a brain is smaller than that of the moon.”

Clearly, (iii) is an empirical observation. But since Simon does not reject the claim that he might be in the Hard simulation, he must concede:

(iv) “If I am in the Hard simulation, then (iii) is not an observation about the reality, but an observation about the simulation, and I must further concede that the so called ‘brain’ in (iii), if uttered or conceived of by the individuals with whom I have lived in the same simulated world, does not actually refer to a brain but to some simulated occurrence.”

Simon thus concludes:

(v) “If I am simulated, my word ‘brain’ still refers to brains, but my fellow inhabitants’ word ‘brain’ does not refer to brains.”

Let us call this thought process TP. In the end of TP, Simon, by thinking (v), has placed himself into an extraordinary circumstance, under which his thinking has either come close to being incoherent or has already become incoherent. The explanation is as follows.

Simon, in thinking (v), must concede that if he is in the simulation, the referent of the word ‘brain’ used by Simon himself is radically different from the referent of the word ‘brain’ used by the rest of the simulated beings.[16] Simon must therefore concede that he has never learned the referent of the word ‘brain’ by interacting with his surrounding environment including individuals around him. Instead, he must concede that he has acquired the referent of the word ‘brain’ entirely from the manipulation by the simulators, namely, MA.

Normally, one’s application of language often, if not always, relies on the employment of reason, or the exercise of reason. The problem Simon is faced with, when conceding that if he is in the simulation, he can then only acquire the referents of the word ‘brain’ entirely from the influence by the simulator, is that he is implicitly committed to a view that every instance of his application of the word ‘brain’, has never relied on, or has never resulted from his exercise of reason.

For example, Simon may recall that his teacher showed the students a brain from a donor in a high school biology class. But Simon must concede that if he is in the simulation, he then did not learn anything about brains from that occasion because his teacher was not interacting with a real brain but with some simulated occurrence, and he himself was of course also not interacting with a real brain in that occasion, and the effort he put into exercising reason to understand what his teacher taught about brains, which would supposedly add to his understanding of the referent of the word ‘brain’, was also in vain.

The reasoning illustrated above does not just apply to the occasion in Simon’s biology class, but also applies to all occasions when Simon uses the word ‘brain’. He must concede that if he is in the hard simulation, the effort he put into exercising reason to apply and to understand the word ‘brain’ is in vain in each of these occasions. And he must therefore further concede that if he is in the simulation, every instance of his application of the word ‘brain’ has never relied on his exercise of reason.

Most crucially, in thinking (v), Simon must concede that his thinking (v), which itself contains the word ‘brain’, is the very instance where the effort he put into exercising reason to apply as well as to understand the word ‘brain’ is in vain, and he thereby must further

[16] In fact, Simon must concede that the referents of most of the words used by himself are radically different from that of those words used by the rest of the simulated beings. And this is reminiscent of Wittgenstein’s idea of private language.

concede that his thinking (v) does not rely on his exercise of reason. Since Simon, by thinking (v), denies the efficacy of reason for thinking (v), he has become deeply incoherent.

****

Some may want to question both (iv) and (v) in TP and contend that, given the supposition of MA, instead of (iv) and (v), the latter half of TP could be unfolded as follows:

An Alternative Ending of TP

(after thinking of (iii)) … since Simon does not reject the claim that he might be in the Hard simulation, instead of (iv), he may concede:

*(iv) "If I am in the Hard simulation, then, because of MA, the word 'brain', uttered or conceived of by any SIM, would still refer to the same group of objects, namely, the brains."

In conceding *(iv), Simon must concede that none of the SIMs has acquired the referent of the word 'brain' by interacting with brains. And he must further concede: "if I am simulated, then neither myself nor any of my fellow inhabitants can gain any bit of understanding about the word 'brain' by interacting either with brains or with someone having the experience of interacting with brains." Thus, Simon must concede:

*(v) "If I am simulated, I cannot gain any bit of understanding about the word 'brain' by interacting either with brains or with my fellow inhabitants (because none of them has ever interacted with brains), and it is solely due to the result of MA that I understand the word 'brain'."

In conceding *(v), Simon is once again committed to the view that his exercise of reason plays absolutely no role in every instance of his using the word 'brain', and this includes the very instance of conceding *(v), which contains the word 'brain'. Again, this commitment of Simon's is incoherent.

TP begins with Simon's holding the idea that he might be in the simulation. Each of two alternative endings of TP inevitably gives rise to a circumstance under which Simon is being incoherent. What TP has shown is, Simon cannot coherently articulate or conceive of the idea that he might be in the simulation.

8. The Final Stage

Since Simon is any simulated being, the key observation we are to draw from the endings of TP is that in order for a SIM to avoid being incoherent, she must reject the idea that she might be in the simulation and must concede that she is not in it. That is,

> **Rejection Move (RM)**: If one is in the Hard simulation, she must reject the idea that she might be in the Hard simulation and must concede that she is not in it.

Given RM, we must eventually concede that we are not in the Hard simulation. I present three ways to argue for this.

****

The first way. Simon's situation is, by definition, the Hard Simulation. But, by RM, Simon must concede that he is not in the Hard Simulation. This implies that he can never correctly comprehend his own situation, namely, the Hard Simulation. We thus obtain:

(1) Simulated beings cannot comprehend the Hard Simulation.

Self-evidently, I can comprehend the Hard simulation. This is especially true of those who want to impress upon us with the idea or worry that we might be in the Hard Simulation. Hence, the following premise obtains:

(2) I can comprehend the Hard Simulation.

In conclusion,

(3) I am not a simulated being.

****

Arguably, the fact that two situations are identical might not imply that not being able to comprehend one of them entails not being able to comprehend another. That is to say, the implication might not hold, and it could still be the case that Simon can comprehend the Hard simulation without being able to comprehend his own situation.

Nevertheless, the second way to show that we are not in the simulation does not hinge on this implication. Suppose I hold the idea that I might be in the simulation. That is,

(a) I might be in the simulation.

Holding this idea simply means that I am unsure of whether or not I am in the simulation. Now assume that,

(i) I am in the simulation.

By (i) and RM, I obtain

(ii) I must reject the idea that I might be in the simulation and must concede that I am not in it.

However, I cannot coherently maintain both (i) and (ii).[17] I must therefore refute the assumption (i). That is, I must concede,

(1) It is not the case that I am in the simulation.

In classical logic, (1) is equivalent to

(2) I am not in the simulation.[18]

And either (1) or (2) compels me to drop (a) as well.

****

The third, and a slightly more straightforward way to establish the claim that I am not in the Hard simulation is as follows: either I am in the simulation or I am not in it. On the one hand, if I am in it, then by RM, I must reject the idea that I might be in it and must concede that I am not in it; on the other hand, if I am not in it, then the idea that I am in it is false. Hence,

[17] This is reminiscent of Moorean paradox or Moorean contradiction. It is imaginable that some might concede that (i) and (ii) cannot be maintained together coherently while still insisting that he himself might be in the simulation. Yet I think this insistence is again incoherent. Hence, in what follows, I regard putting (i) and (ii) together as the usual contradiction. For a closely related sentiment about the brain-in-vat scenario, see Nagel (1986:73), Folina (2016: 172), Pritchard and Ranalli (2016: 88-89). For the response to this sentiment, see Button (2013: 13.3, 13.4, 13.5, 14.4).

[18] Even if, in view of some other kind of logic, it is not legitimate to conclude from (1) to (2), what is at least established here is that to avoid being incoherent, one must restrain from thinking the idea that she herself is in the Hard simulation.

the idea that I am in the simulation is either incoherent or is false. Either way, I am compelled to concede that I am not in the simulation.

****

In the end, each of these three arguments leads to the same conclusion that I am not in the Hard simulation.

9. The SIM Argument and its relevance to the Vat Earth Scenario

The reasoning presented in the previous sections refutes the idea that we might be or are in the Hard simulation. Let us call this reasoning as a whole the *SIM Argument*. It contains three parts. The first part shows that the semantic situation of the SIMs is virtually the same as that of the BIVs.

The second part is the most essential one. It addresses the existence of normal individuals in the Hard simulation. This existence may enable the SIMs and the normal individuals to use the same language, which ensures that we are able to unfold the thought process, namely TP, of a SIM who, like us, is investigating the idea that she might be in the Hard simulation.[19] TP eventually brings us with the discovery that no simulated being can coherently maintain the idea that she might be in the simulation.

The third part begins with the observation that to avoid being incoherent, any SIM must abandon the idea that she might be in the simulation, and must concede that she is not in it. I have called this move the Rejection Move, RM. By invoking RM, we have constructed three parallel sub-arguments for establishing the conclusion that we are not in the Hard simulation.

It is not difficult for us to see from the SIM argument that the absence of the conscious communication, between the simulated beings and the normal individuals in the Hard simulation, constitutes a fundamental reason why the simulated beings cannot maintain the idea of them being in the Hard simulation without becoming incoherent, as well as why we cannot be in the Hard simulation. Moreover, since the argument does not rely on any specification of physical law, it indicates that whatever physical laws governs the reality, and

[19] It is important to note that whether we are in the Hard simulation is not settled at this stage yet.

however the simulators can utilize those physical laws, none of those can override that reason and prevent the simulated beings from being incoherent in maintaining that idea.

And for the same reason that there is no conscious communication between the simulated beings and the normal individuals, we cannot be in the Vat Earth scenario as well, regardless of how close the BIVs and the normal individuals can be from each other. Thus, contrary to Button's view that whether the claim that we might be in the Vat Earth scenario can be refuted is an indeterminate matter, this claim can be refuted conclusively.

10. The Encounter Simulation

Although the occurrence of some conscious communication between the normal individuals and the deluded beings is a necessary condition for the deluded beings to be able to conceive of the idea that they might be deluded, such communication is not sufficient for the deluded beings to make sense of the idea that they themselves might be deluded, as what the following discussion shows.

Recall that in section 3 we briefly discussed some conscious communication Bostrom delineated, when he addressed the question whether the claim that we are in a simulation is testable:

> There are clearly possible observations that would show that we are in a simulation. For example, the simulators could make a 'window' pop up in front of you with the text "YOU ARE LIVING IN A COMPUTER SIMULATION. CLICK HERE FOR MORE INFORMATION." Or they could uplift you into their level of reality.

As usual, here is a succinct formulation of what Bostrom proposed:

The Encounter Simulation. Simulated beings were, are, and will be simulated by some civilization, who will make contact with the former, and will explicitly inform them that they are simulated. Prior to that moment, all the simulated beings do not know that they have been simulated.

We cannot be in the Encounter simulation. To show this, let Simone be a SIM in this scenario. By the stricture of this scenario, Simone, for the time being, has not communicated with, and therefore has not been informed by the simulator that she is simulated. Thus, her semantic resources at the moment could not be any better or richer than, and could at most be identical to, Simon’s. And this cannot allow her to coherently conceive of the idea that she might be in the Encounter simulation.

More specifically, we can similarly run a SIM-style argument. The first part of the argument would show that her semantic situation is the same as that of Simon, and virtually the same as that of BIVs.

The second part would show that the maximal influence on Simone’s semantic ability exerted by the simulator could, at the moment, only still be the one that is implicit and unnoticeable by the simulated beings, namely, some influence like MA. And her thought process regarding the idea that she might be in the Encounter simulation could not be any relevantly different from Simon’s thought process regarding the idea that he might be in the Hard simulation.

A new portion can be added into the thought process in which Simone imagines herself having first conscious encounter with the simulators. But this new portion has little relevance in helping her avoid being incoherent, as long as she realizes and concedes that if she is in the Encounter simulation, then since she has not had this encounter yet, her own semantic situation could currently be no better than Simon’s.

Inevitably, Simone’s thought process can in the same way end with either the circumstance under which Simone would think: “if I am simulated, although my word ‘brain’ still refers to brain, the word ‘brain’ used by those individuals with whom I have been living in the same simulation does not refer to brains”; or the circumstance under which Simone would think: “If I am simulated, I cannot gain any bit of understanding about the word ‘brain’ by interacting either with brains or with my fellow inhabitants, and it is solely due to the result of MA that I understand the word ‘brain’.”

The move like RM can be taken in a similar vein from either of these two ending thoughts, and the rest of the SIM argument remains the same except that we need to substitute the Encounter simulation for the Hard simulation every time the latter appears, and by doing so, we can eventually obtain the conclusion that we are not in the Encounter simulation.

11. Conclusion

Nick Bostrom contended that we might be living in one of the three kinds of simulation. I call these three simulations the Easy, the Hard, and the Encounter simulations. In this paper, I have presented the refutations of the ideas that we might be or are living in any of these three simulations. These refutations are greatly inspired by Tim Button's reconstruction of Hilary Putnam's famous refutation of the idea that we might be brains in vats. Incidentally, although these refutations show that we cannot be living in any of these simulations, they do not show that there cannot be any of these simulations in our universe.

References

Besnard, F. 2004. "Refutations of the Simulation Argument," Retrieved from http://fabien.besnard.pagesperso-orange.fr/pdfrefut.pdf.

Birch, J. 2012. "On the 'Simulation Argument' and Selective Scepticism," *Erkenntnis* 78: 95-107.

Bostrom, N. 2003. "Are we living in a computer simulation?" *Philosophical Quarterly* 53: 243–255.

Bostrom, N. 2008. "The simulation argument FAQ". Retrieved on 25 March 2012 from http://www.simulation-argument.com/faq.html.

Bostrom, N., & Kulczycki, M. 2011. "A patch for the simulation argument," *Analysis*, 71, 64–71.

Button, T. 2013. *The Limits of Realism*. Oxford: Oxford University Press.

Button, T. 2016. "Brains in vats and model theory." In S. Goldberg (Ed.), *The Brain in a Vat*, 131-154. Cambridge: Cambridge University Press.

Chalmers, David, J. 2022. *Reality+: Virtual Worlds and the Problems of Philosophy*. W. W. Norton & Company.

Folina, J. (2016). "Realism, skepticism, and the brain in a vat." In S. Goldberg (Ed.), *The Brain in a Vat*, 155-173. Cambridge: Cambridge University Press.

Hammarstrom, A. 2008. *I, Sim - An exploration of the Simulation Argument*. MA thesis.

Nagel, T. 1986. *The View from Nowhere*. New York: Oxford University Press.

Pritchard, D. 2016. *Epistemic Angst*. Princeton, N.J.: Princeton University Press.

Pritchard, D. and Ranalli, C. (2016). “Putnam on BIVs and radical skepticism.” In S. Goldberg (Ed.), *The Brain in a Vat*, 75-89. Cambridge: Cambridge University Press.

Putnam, H. 1981. *Reason, Truth and History*. Cambridge: Cambridge University Press.

Van Fraassen, B. 1997. “Putnam’s paradox: Metaphysical realism revamped and evaded,” *Noûs* 31(s11):17–42.

Van Fraassen, B. 2008. *Scientific Representation: Paradoxes of Perspective*. Oxford University Press.

Wright, C. 1992. “On Putnam’s Proof that We Are Not Brains in a Vat,” *Proceedings of the Aristotelian Society 92*: 67–94.