5-2024

Beyond the Matrix: Navigating the Implications of Nick Bostrom's Simulation Hypothesis on Reality and Existence

Michell-Lee Graham '24
DePauw University

Beyond the Matrix: Navigating the Implications of Nick Bostrom's Simulation Hypothesis on Reality and Existence

Michell-Lee Graham

DePauw Honor Scholar Program

Class of 2024

Sponsor: David Alvarez

Second Chair: Erik Wielenberg

Third Chair: Justin Glessner

Acknowledgements:

I would like to thank my thesis committee, Professors's David Alvarez, Erik Wielenberg, and Justin Glessner, for their assistance in completing this thesis. It was by their guidance that helped to realize this project, and I am grateful for their help. I would also like to thank my friends and family for their encouragement while completing this thesis.

Content:

Abstract:

Suppose for the sake of the argument we are all living inside of a simulation, is that a problem? Nick Bostrom, a Swedish philosopher has proposed the question of whether we are living in a simulation. Philosophers in the past have grappled with this concept, from Plato to George Berkeley. This concept is also readily seen in movies such as “The Matrix" which is a science fiction film that explores the concept of a simulated reality. In the story, humans live in a simulated world created by sentient machines to distract and control them, while their physical bodies are used as an energy source. The idea of a simulated reality raises philosophical questions about the nature of existence, perception, and reality itself. For this paper, I will start with summarizing Nick Bostrom's argument. Then, I’ll explore some of the implications which he mentions, such as questions of reality, the afterlife and God and the problem of evil. Lastly, I will explore the ethical implications of the simulation. Our hypothesis is that we are not living in a simulation, but that we could, in the future, create one. However, I also argue that we should not create one, regardless of the benefits because of the effects it will have on the intrinsic value of humans.

Nick Bostrom, in “Are We Living in a Simulation” argues that at least one of the

following propositions are true, (1) "the human species is very likely to go extinct before reaching a "posthuman" stage;" (2) any posthuman civilization is extremely unlikely to run a significant number of simulations of their evolutionary history (or variations thereof); (3) we are almost certainly living in a computer simulation"[1]. If (1) and (2) are both false, then (3) must be true and we are certainly living within a simulation[2]. At the conclusion of Bostrom's paper, he argues that if we are not currently living within a simulation, then most likely none of our future generations ever will. However, Bostrom also claims that given the information we currently have, we should consider each of (1), (2) and (3) to be equally likely.

Introduction:

By reviewing Bostrom's argument, we gain a more thorough understanding of his reasoning and evidence. It also helps us to assess how likely his argument is and clearly identify questions that remain unanswered and that Bostrom has failed to elaborate on. In the first section of this paper, I present Bostroms's arguments and reasoning for his claims. While I agree overall with Bostrom's argument and find it persuasive, I note that there are some possible objections to his third proposition. Instead of focusing on the argument, I'll focus on a few valuable philosophical questions that are illuminated by the argument. For the latter part of the paper, I present many implications stemming from Bostrom's argument. The first (1) implication I present is the question of who created the simulation. In the second implication, (2) I explore the possibility of an afterlife/resurrection within the simulation. In the third (3), I discuss the question `what is reality" ? In the fourth (4), I discuss the concept of free will and if simulated

[1] Nick Bostrom "Are We Living Inside a Simulation" pg 1
[2] Nick Bostrom "Are We Living Inside a Simulation" pg 1

beings would have the ability to exercise it. In the fifth (5), I discuss the issue of the existence of evil within the simulation. In the sixth (6), I question if there is a reward/punishment system within the simulated world. In the seventh (7), I discuss the ethical concerns of creating the simulation. Lastly, I provide several pieces of evidence supporting my hypothesis that in the future, based on our current advances in technology, we could create our own ancestor simulations.

Nick Bostrom's Argument:

According to Bostrom, proposition (1) which is that "the human species is very likely to go extinct before reaching a "posthuman" stage;" may likely be true based on our current technological development. He supposes this will be the "result of the development of some powerful but dangerous technology"[3]. He states, "One candidate is molecular nanotechnology, which in its mature stage would enable the construction of self‑replicating nanobots capable of feeding on dirt and organic matter – a kind of mechanical bacteria"[4]. These bots, once aged, create a type of mechanical bacteria which someone can use for malicious purposes. Bostrom adds that proposition (1) in itself mostly implies that to build the devices needed to create a simulation will take a long time, and most likely will not be completed before our extinction. Additionally, he argues that humanity might become extinct because of warfare. With increased

[3] Nick Bostrom "Are We Living Inside a Simulation" pg 10
[4] Nick Bostrom "Are We Living Inside a Simulation" pg 10

tensions worldwide, there is a possibility that we may enter into another World War. Granted, these are all hypotheticals and Bostrom encourages us not to dwell on them too much.

Bostrom's second (2) proposition is that "any posthuman civilization is extremely unlikely to run a significant number of simulations of their evolutionary history (or variations thereof) "[5]. He claims that (2) might be true because posthuman civilizations would lack the necessary resources to run the ancestor simulations. In fact, if they did have the resources, they would still be reluctant to use their resources to run the simulations because of the ethical implications. One implication would be that the simulations would "inflict harm and suffering upon its inhabitants"[6]. Post humans could also possibly lose interest in running these simulations if their human desires had faded. This could possibly be a result of them finding the creation of these simulations as pointless, but more so that their desires to understand humanity would fade. From proposition 2, he concludes that "posthuman societies will be very different from human societies: they will not contain relatively wealthy independent agents who have the full gamut of human‐like desires and are free to act on them"[7]. Bostrom infers that posthumans will essentially be constrained by whatever systems that hold them. They will not have the ability to freely act on any desires, for even those they are without.

However, running ancestor simulations may prove to be beneficial to posthuman society. Miloš Agatonović, a Senior Lecturer who specializes in Philosophy at the Preschool Teacher Training College, Kruševac Section of The Academy of Applied Preschool Teaching and Health Studies, further expands on this claim in his paper "The Fiction of Simulation: A Critique of

[5] Nick Bostrom "Are We Living Inside a Simulation" pg 1
[6] Nick Bostrom "Are We Living Inside a Simulation" pg 11
[7] Nick Bostrom "Are We Living Inside a Simulation" pg 11

Bostrom's Simulation Argument." He argues that by running the "large-scale realistic simulations," they may help us achieve a better society, by showing us the errors we can make in the history of humankind so we can correct them"[8]. Thus, these simulations would serve as a test and trial system so that further generations may be able to live more prosperous lives.

Furthermore, over the past few decades, computational power has become cheaper and with models such as the Monte Carlo simulations, which is a method used to simulate a problem continuously, which in turn provides a different outcome each time it is run. These simulations are often used in weather predictions and in finance especially in places where there seems to be many dimensions[9]. Given their ability to produce certain outcomes we can imply that in another century or so, it is possible that mankind could create their own "large scale realistic"[10] simulations given they have enough resources. Considering the potential advancements in technology over the next century, including the ability to create large-scale realistic simulations with sufficient resources, it's intriguing to ponder Nick Bostrom's third proposition.

Bostrom's third (3) proposition is "we are almost certainly living in a computer simulation"[11]. He argues that if propositions (1) and (2) are both false, then we are living in a simulation. However, Bostrom does not provide an explanation for why we would be aware that we are living inside of a simulation. However, he also argues that it would be impossible for future generations to create a simulation in the future, if we are not currently living inside of one. This is because if we are not currently living inside of a simulation, this would mean that we

[8] Agatonović, M"The Fiction of Simulation: A Critique of Bostrom's Simulation" pg 1581
[9] Viking, Nillson "An in Depth Look at Bostrom's Simulation Argument" pg 4
[10] Agatonović, M "The Fiction of Simulation: A Critique of Bostrom's Simulation" pg 1581
[11] Bostrom, Nick "Are We Living Inside a Simulation" pg 11

currently do not have the computational power or resources to create simulations and will not exist long enough to do so.

While Bostrom holds that if we do not currently have the computational power to produce simulations we never will, I disagree and think that it is plausible that we currently lack such power, it is also reasonable to argue that we will develop it prior to humanity's extinction. I contend that despite our current technological limitations, future advancements may enable the creation of simulations before the potential extinction of humanity. In regards to current technology, they show promise in the possibility of this future. For example, Video and Optical See-Through AR Surgical Systems, also known as VOSTARS, are augmented reality wearable head mounted displays which are used in surgeries. These headgears allow for users to see a virtual display of their patients, and are capable of creating 3D models of the patients as well. With inventions like VOSTARS already created, mankind is very likely to be on its way to creating simulations. This is because these headgears are an almost immersive experience, and if scientists can further develop this technology, over the course of many years, it is likely they could produce a more fully immersive experience such as the simulation.

Another illustration pertains to role-playing games, commonly referred to as RPGs. Within these games, players "assume the roles of imaginary characters and operate with some degree of freedom in an imaginary environment"[12]. Various formats of these games exist, with some allowing players to use headset devices that project a virtual screen, providing an immersive experience In a way, RPG games and the simulation are similar in that beings are simulated within virtual settings, and though they may have some autonomy, they are still bound

[12] Zagal, J. P., & Deterding, "Definitions of "Role-Playing games". *Role-playing Game Studies: Transmedia Foundations*, pg 4

by the rules of the system. Despite players' awareness of the game context, it is noteworthy to recognize the parallels and acknowledge these games as exemplary manifestations of human creativity. In the foreseeable future, there is certainly a possibility that we will be able to create our own full scale simulations.

In the implication section on the future of artificial intelligence in society, I will discuss further how recent technological advances are prime evidence that with time, society will be able to possibly develop ancestor simulations.

Overall, given Bostrom's argument, and our current technological advances, there is a serious possibility that we are living inside of a simulation that cannot yet be ruled out. Therefore, in the next section I will assume for the sake of the argument that we are living in a simulation and I will explore the implications of that assumption.

Implications:

Who Created the Simulation?

One of the many questions that stems from this argument is if we are living in a simulation, then who created the simulation? Bostrom provides one explanation: posthuman civilizations. These beings would be “humans” who have surpassed their human capabilities by integrating advanced technologies into their bodies, enhancing their physical being and intellect. These civilizations would have reached a stage where they have obtained most of the technology that can be reached in accordance with the laws of nature. Bostrom argues that at “such a mature

stage of technological development, it will be possible to convert planets and other astronomical resources into enormously powerful computers."[13] Given their ability to convert these resources, Bostrom finds that it is possible that post-human civilizations could create their own ancestor simulations. Though, unless there is a new discovery in physics, Bostrom claims that these civilizations would be unable to simulate the entire universe down to the quantum molecular level based on what is currently known about quantum mechanics.

Bostrom makes an alluring argument. It reminds me of a paper I heard by Darian Santmyer on "Heidegger and Space Colonization." The main section of his paper references Heidegger's critiques of technology, and I will be focusing on one of these critiques mentioned in Santmyer's paper that Heidegger makes of modern technology to argue that post-human civilizations, if they exist, currently could not create the simulation.

Heidegger argues that modern technology orders nature into a framework. Through this framework, humans exploit nature's energies and view themselves as managers of the system. He then argues that if humanity continues down this path, eventually, we will view ourselves only as resources meant to be used, instead of as beings meant to be valued. Heidgger does not reject technology, but warns of its dangers. I include this argument because I find it valuable for thinking about Bostrom's claims. If post-human civilizations are able to harvest these resources and create their own ancestor simulations, then the definition of humanity will blur. This is because in creating these ancestor simulations, post-human societies would view and use humans as resources that can be controlled. If humans can be used as resources, then they will no longer

[13] Bostrom , Nick "Are We Living Inside a Simulation" pg 11

be seen as having any intrinsic value and be reduced to mere resources to be used. Thus, post humans will have asserted themselves as the superior beings, ruling over both “man” and nature.

Though, Bostrom contends as well that these “posthuman civilizations” may themselves be simulations. So, it is possible that their idea of humanity is far from our own because they have long surpassed their human instincts, hence the name “post-humans.” Because of their post-human nature, they could simulate humans. In addition, Bostrom's supposition that the creators of the simulation are themselves simulations infers that there may be many simulations. As a result, over time, as these simulations are created, the nature of these beings change and evolve, leaving parts of their humanity behind. Historically, the word human comes from the Latin root humanus, a cross of homo and humus, which means “man” and “earth”. Therefore, historically human means a being deeply connected to earth. However, Heiedgger argues that man has strayed away from this definition and has instead obscured the meaning of the word.

Because the meaning of “man” has been changed due to technology, Heidegger provides a warning against technology. Mark Blitz, a Fletcher Jones Professor of Political Philosophy at Claremont McKenna College, in his explanation of Heidggers argument in “Understanding Heidegger on Technology" states that

> “We treat even human capabilities as though they were only means for technological procedures, as when a worker becomes nothing but an instrument for production. Leaders and planners, along with the rest of us, are mere human resources to be arranged, rearranged, and disposed of. Each and every thing that presents itself technologically thereby loses its distinctive independence and form. We push aside, obscure, or simply cannot see, other possibilities.”[14]

[14] Blitz, Mark. "Understanding Heidegger on Technology." Pg 68

Blitz argues that the workforce is an example of humans seen as resources. In some workplaces, a worker is only valued based on his economic productivity. So if a worker is underperforming based on the standards set by his supervisor, he or she could easily be fired and replaced. This can readily be seen as instances where an employee is close to the age of retirement, and forced out of his job.

Thus, I agree with Heidegger and Blitz because this is already happening within our societies. And so, if posthuman societies could create a simulation, then they would be viewing human lives as mere tools for achieving their own ends. This helps the argument because it allows us to understand the nature of the "creators" of the simulation and how they could arrive at creating their own simulations.

Moreover, If post human societies could create the simulation, Bostrom suggests that they could be conducted on "virtual machines"[15] which are run on large computers created by the post-human civilizations. These "virtual machines' ' would allow these civilizations to simulate a machine or rather a computer within another simulation. Essentially, this means that it is very likely the post-human creators could have an almost infinite amount of simulations or at least as many as they are able to make.

Bostrom argues that "If we do go on to create our own ancestor‐simulations, this would be strong evidence against propositions (1) and (2), and we would therefore have to conclude that we live in a simulation."[16] Propositions (1) and (2) argue that humanity would go extinct before ever reaching the post-human stage and that they would not have the necessary resources to create and run their own simulations. And so, if posthuman beings are able to reach this stage

[15] Bostrom, Nick "Are We Living Inside a Simulation" pg 11
[16] Bostrom, Nick "Are We Living Inside a Simulation" pg 11

and harvest these resources whether from space or earth, then one can assume that it is likely that we are living in a simulation and that the creators of the first simulated beings may also be simulated beings.

Because it is possible that post human civilizations, who are very likely to be simulations themselves, have long departed from the historical definition of man, Bostrom claims that our posthuman creators are both omnipotent and omniscient in the sense that they can "interfere in the workings of our world even in ways that violate its physical laws; and that they can monitor everything that happens."[17] The posthumans who created the simulation can also be said to have super-intelligence. Some might say their oversight is similar to that of "Big Brother" which is a character and part of a slogan used in the book 1984 by George Orwell in reference to mass surveillance. In the book, citizens live in a totalitarian society in which they are massively sureviled and their government wields total control over them. Their situation is, in part, similar to the inhabitants of the simulation because they also do not have complete control over themselves and their creators monitor all that they do.

Bostrom claims that this would then mean that there is some hierarchy within the simulation, with the posthuman creators being at the top of that hierarchy. Bostroms description of these posthuman creators suggests that their likeness or nature is similar to one of a God. Because of this hierarchy, the created simulated beings will be "forced" to behave according to the rules set by their creators.

[17] Bostrom, Nick "Are We Living Inside a Simulation" pg 12

However, Dr Eric Steinheart, professor of Philosophy at William Paterson University, expands on this argument and argues that there has to be an original designer or creator of these ancient civilizations who have attained posthumanism, which he credits to be "God." He states

> "Neoplatonism says that God (the One) is the source of all reality. Another classical argument for the existence of God is the Design Argument. It shows up as Aquinas's Fifth Way. It goes something like this: (1) we see that our universe contains complex internal structure; (2) if there is any thing with some complex internal structure, then that thing was produced by an intelligent designer; (3) therefore, our universe has an intelligent designer; (4) it is conventional to refer to the intelligent designer of our universe as God; (5) consequently, God exists. And yet, if the Simulation Argument is right, then the designer of our universe is not God. It is just a deeper civilization"[18].

Essentially, Aquinas' fifth way is that our universe is made up of complex structures, and given its complexity, it has to be designed by someone or something that is highly intelligent. However, Steinheart argues that there has to have been an earlier designer than the ancient civilization because they must have been created by something themselves. He describes this being as "the unmoved mover; the uncaused cause; the undesigned designer"[19]. He claims that this being is infinitely more powerful and intelligent than any finite being or civilization[20]. Thus, this being guides and directs the finite beings within the ancient civilizations. One can assume that Steinheart's argument is likely because if post-human civilizations are able to create their own ancestor civilizations and possess certain God like qualities, then their creators must be even more powerful than them. They would have to be something or someone that was not designed

[18] Eric Steinheart "Theological Implications of the Simulation Argument" pg 30
[19] Eric Steinheart "Theological Implications of the Simulation Argument" pg 20
[20] Eric Steinheart "Theological Implications of the Simulation Argument" pg 20

themselves, neither coming or going into being, but just "being." He concludes this infinite being must be none other than God himself.

However, Steinheart also states that "if the Simulation Argument is right, then the designer of our universe is not God. It is just a deeper civilization."[21] From Steinhearts argument, we can infer that he disagrees with Bostrom's premise that we are living in a simulation. His argument implies that if it's true that we are living inside of a simulation that was created by post-human civilizations, then the world we call home, "reality" outside of the simulation, could not have been created by God, but by an older civilization. However, this conclusion would vary in what your beliefs are because as some argue there is no God. Furthermore, this conclusion would also conflict with the Big Bang Theory, which is believed to be how the universe came into existence. This is an interesting insight because it allows us to recognize how the simulation could also potentially affect our world as well as the simulation itself.

Regardless, considering Bostrom's argument, one wonders what this system within the simulation would be like. Why has God allowed these ancient civilizations to create the simulation and if they were created, are they also bound by sin as well like other finite beings? Given what we know already about the nature of these post-human creators, if we are living in a simulation, then inhabitants of the simulation might not be bound by sin. This is because we would be under the direction of the post human creators and not God. Though, one can suppose that the post-human creators could have to answer for their sins by God.

Given Bostrom and Steinheart's arguments, I seem to agree with both. I think that Bostrom's argument that a post human civilization could have created a simulation of our reality,

[21] Eric Steinheart "Theological Implications of the Simulation Argument" pg 30

but I also agree with Steinhearts argument that something such as God, must be the original creator of the post humans. And so, it is very likely that the creators of the simulation are post humans who were created by God.

Is an Afterlife/ and Resurrection Possible?:

If an afterlife is possible within a simulation, then how would it work? Would you simply be logged out of the simulation, and where does your sim go? Steinhart argues that the lives of those within simulation may have been recorded and then stored as data within the system. And so, after they die they could either be placed within another simulation or the creators may place their memory cards within an "artificial body, so that they can interact with the creators in their civilization"[22]. Steinheart argues that this process is similar to something he calls computational resurrection. He includes the work of Bruce Reinbach, a retired Professor of Philosophy, who claims that

> Viewed monistically, man is nothing more than a physical organism constructed and programmed in a certain fashion. Some have likened man to an extremely complex computer with a physical body. If one adopts this analogy, and applies it to the issue of life after death, the following would be the monistic re-creationist's thesis: just as one can construct two computers to look identical, program them identically, and feed them precisely the same program data, so it would not seem to be self-contradictory that an individual could be physically re-created to possess all the physical characteristics of the

[22] Eric Steinheart "Theological Implications of the Simulation Argument" pg 31

> deceased in identical proportions and correlations, such that he would look identical to the person who died.[23]

Reinbach's claim is that if we view the physical body as an instrument that was constructed, we can also imagine it as a "computational structure" whereas one could create and recreate certain physical aspects of the "program." Bostrom makes a similar claim, where he supposes that post-human creators, given their resources and abilities, could manipulate the memories of the simulated beings. If so, then truly, maybe these post-human creators may be God in their own right, as only a god could manipulate its creations' memories and resurrect them from the dead.

However, if simulated beings are able to die and be reborn continuously, then is there any true purpose to their existence? Reinbach goes on to argue that not only would these individuals look alike, but their consciousness would so as well. The brain could be reprogrammed and rewired to replicate that of the previous life the person had lived. Thus, one could be resurrected many times and the more times they are resurrected, the closer they are to God. Reinbach uses this argument from John Hick, a philosopher and theologian who argued that "the purpose of human life is to move to divine perfection"[24]. This idea was based on his pareschatology theory which is that there exists universes that contain perfect beings. And so, through death and rebirth, an individual gets one step closer to that level. This is because as you go through different lives, you continuously learn more about the nature of the world, and of its people, and how it works. Furthermore, with each life lived, there is the opportunity for character development and intellectual growth. And because all of your memory is stored within a memory card, all of this knowledge is readily accessible.

---

[23] Eric Steinheart "Theological Implications of the Simulation Argument" pg 32
[24] Eric Steinheart "Theological Implications of the Simulation Argument" pg 32

Although, if someone is continuously resurrected by the simulation, then is that person still the same person as the first time they were resurrected? Does resurrection change one's personal identity? In *Reasons and Persons* by Derek Parfit, a British philosopher whose focus was on personal identity, ethics and rationality, provides several theories in an attempt to identify what determines personal identity. One scenario given is of identical triplet brothers who are severely injured. Two of the brothers' brains are very damaged, and one brother's body is badly damaged. So, the doctors propose to split the brain of the third brother whose body is damaged, in half, and place one half in each of the brothers bodies. Both of these bodies now possess half of their third brother's brain and their physical features are identical. After their surgeries, the brothers now believe that they are both their [third] identical brother. They have the same personalities and character as him, and are also psychologically continuous (chain of direct memories)with him as well. For the sake of this argument, we can claim that this is a type of resurrection, in that their third brother is now seemingly surviving in his two brothers bodies. However, considering the effects of this "resurrection" Parfit questions whether either or both of these brothers would be able to claim that they are their third brother. To consider this, Parfit explains that there are there are "two kinds of sameness,"[25]qualitative and numerical. To be qualitatively identical, you and your "replica" would be exactly the same in your physical and mental appearance. However, you and your replica would not be numerically identical because your replica would not be the "same" person as you, since there can only be one of "you" in existence at the same time. So, in the case of the triples, neither brother can claim that they are their third brother because although they are qualitatively identical they are <u>not</u> numerically identical, which matters more. Being numerically identical matters because according to Parfit,

[25] Parfit, Derek. 1986. *Reasons and Person*. Oxford University Press. Pg 200

this sameness is likened to survival. What matters for Parfit is continuity of personal experiences and psychological characteristics, not identity which is more likened to the preservation of an unchanging ego or soul.

Furthermore, Parfit supposes that if each of these replicas were to live several years apart at opposite ends of the earth, they would reunite as totally different people. This is because having been separated from each other, each "replica" of the third brother goes on to have its own individual experiences independent of the other. Therefore, this reinforces the notion that neither brother is their third brother because at some point, these replicas will go on to live their own lives. And so, the question that remains is what happens to the identity of the third brother? If his personal identity does not survive in his "replicas" does he simply die at the point of his "resurrection"?

Parfit's conclusion that personal identity does not survive resurrection is quite appealing when we take into account Reinbach's argument that a person becomes closer to the likeness of God as they are repeatedly resurrected. If what Parfit says is true, then Reinbachs argument sounds plausible. If the goal or the end result is the likeness of God, it makes sense that a person's personal identity would begin to fade out as they are repeatedly resurrected. With the knowledge they consume during these events, their values, goals and beliefs change too. Soon enough, these individuals can no longer claim that they are the people they were before their first resurrection. They have now ascended to Godhood. This claim also supports the idea that the creators of the simulation are almost like Gods. Suppose these individuals themselves went through this process and have achieved this state of being. If this is true, it would suggest that Bostrom's claim that they would be omnipotent and omniscient would be true or at least seem plausible.

To further this argument, we'll take a look at a similar example offered by Daniel Dennett in his story *Where Am I?* Let us note that while Dennet and Parfits arguments are similar in their nature, Dennet's story emphasizes the complexities of personal identity between the mind and body. Meanwhile, Parfit's argument focuses on how personal identity changes over the course of time irrespective of the physical body's continuity. Nonetheless, Dennets argument is a valuable asset to discovering the effects of resurrection on personal identity.

In Dennett's fictional scenario, he explains that he was asked to join a project with NASA and The Department of Defense to retrieve a warhead that was buried deep within the ground in Tulsa, Oklahoma. To retrieve this warhead, Dennet is informed that he will need to have his brain removed from his body because the device was highly radioactive. His brain would be placed within a vat and connected to radio links. Even without his brain, he is told he would still be able to function his body properly. However, while retrieving the device, Dennett's body begins to deteriorate due to the radiation of the device and his body is ultimately crushed underground. With the loss of his body, Dennett' consciousness is switched to his brain and he realizes that he can no longer smell, taste or even move. Fortunately, the lab was able to provide Dennett with a new body, though this one was quite different from his original one. Immediately, Dennet names this new body Fortinbras, seemingly making an unconscious distinction between his old body and new one. Though he claims that after his adjustment to his new body, "his personality was by and large also preserved." [26] This is an interesting claim considering the ending of Dennett's story. Nonetheless, in addition to his new body, Dennett learns that the lab also made a "computer duplicate of his brain"[27] and that he will now be able to use both of his

[26] Dennett, Daniel. 1981. "Where Am I?" Lehigh University. 1981,pg 6
[27] Dennett, Daniel. 1981. "Where Am I?" Lehigh University. 1981, pg 6

brains interchangeably with the switch of a button. Dennet at first was quite apprehensive to this idea because he was afraid of there being "two" Dennetts. However, he soon accepts the decision and routinely begins to switch back and forth between his "actual brain" and his computerized brain. At one point, Dennett admits that when he switches between brains, he has no idea which one he is switching back to because there are no mark indicators on the control. He states that some may believe that in doing so, he has lost his sense of identity. However, he asserts that he still believes that he is fully Dennett. Ironically, at the end of the story, Dennett flips his switch and suddenly the perspective switches and we can assume that it is Dennett's computerized brain that is talking. He claims that for the past two weeks he has struggled because their brains had drifted out of sync and so "this" Dennett is not sure which brain he is using. He also claims that in certain circumstances, the other brain or the "other" Dennett would do things differently than he would have if he had control of their body at that time. "You'd scratch our itches, but not the way I would have..it's been like being carried around in a cage or better like being possessed."[28] Prior to this event, Dennett claimed that he was Dennet and that flipping the switch would not change his personality. However, by the end of the story, it seems that Dennett's other brain has developed its own personality and was in fact even seeking its own body. He states "At the end of this colloquium is over, you and I will fly to Houston, and we'll see what can be done to get one of us another body."[29] By the end of this story, can Dennett truly claim that he is still the same Dennett he was at the beginning of this story? No, because like the brothers in the previous example, Dennetts personal identity did not survive his "resurrection." His replica or "other self" is a separate individual from himself.

[28] Dennett, Daniel. 1981. "Where Am I?" Lehigh University. 1981, pg 7
[29] Dennett, Daniel. 1981. "Where Am I?" Lehigh University. 1981, pg 8

In comparison to Parfit's, Dennett's argument seems to support Parfits claim that what is important in personal identity is the psychological continuity of a person and not their bodies. In Dennetts story, he claims that when he is granted a new body, he still felt that he was himself because he still retained all of his memories, his personality and desires. However, when he uses the button on the switch, he now has broken this continuity and is technically another person entirely. This is similar to the Parfit case, where the third brothers' brain is placed inside of his twin brother's bodies. In both cases, psychological continuity is a key factor in one's personal identity. And so, without it, it is plausible to claim that once the continuity is broken, this person's personal identity would be different. Therefore, if a person was resurrected but placed within a different body, but still remained psychologically continuous with their previous self, one could claim that their personal identity would remain the same. However, as stated before, this individual would only be qualitatively identical to his previous self and not numerically identical.

Contrastly, Calvinism (protestant theology which follows doctrines of the traditional church, with an emphasis on predestination) suggests that in resurrection, personal identity survives. In "Death, Resurrection and the Continuity of Personal Identity" by G. Andrew Peoples, a New Zealand Christian Philosopher, presents the argument that in Calvinist tradition, at the time of resurrection during end times, the personal identity of an individual survives because it is uniquely tied to the soul. He states

> "Bodies are things that need not be continuous, they are frail, transitory and corruptible, subject to decay. The real self, the undecaying, abiding core of identity is the immaterial soul. Regardless of how far this contrast is pressed, the soul is less touched by the

> consequence of human fallenness than the body in that it, unlike the body, does not die (which is an effect of the fall). Setting the more ascetic elements of Calvin's anthropology aside, it is at least clear that he saw the body, unlike the soul, as unnecessary and transitory, whereas the soul was necessary, and could (and would) ensure the continuity of the self's identity"[30]

Here, Peoples seems to suggest that the body is not needed for resurrection, which is an idea that is commonly shared between him, Dennett, and Parfit. However, unlike Parfit, Peoples argue that what is essential to personal identity is the soul or ego of a person. This soul or ego is something distinct from a person's memory, experiences and desires and exists separately from the body. Peoples quotes Millard Erickson, a Professor of Theology at the C.S Lewis Institute, who claims that there is a duality found between the mind and body and when a person dies,

> "part of the human survives death. Death consists in the separation of the soul from the body. The immaterial soul lives on in a conscious personal existence while the body decomposes. At Christ's second coming, there will be a resurrection of a renewed or transformed body, which will be reunited with the soul."[31]

This claim by Peoples is consistent with an argument that Parfit introduces called the Non-Reductionist View. This argument claims that personal identity is a "separate existing enmity, distinct from their minds and body."[32] This enmity is something that is outside of experiences, memories and desires, it is also unchanging. For Calvinist, it seems that the soul or ego is considered to be the core of a person's personal identity. It is what makes a person who they are and is separate from the experiences they have had in life.

---

[30] Peoples, G. Andrew. "Death, Resurrection and the Continuity of Personal Identity." pg 6-7
[31] Peoples, G. Andrew. "Death, Resurrection and the Continuity of Personal Identity." pg 7
[32] Parfit, Derek. 1986. *Reasons and Person*. Oxford University Press. pg 332

Considering the arguments of Parfit, Dennett and Peoples, Parfit and Dennett's arguments seem to be the most plausible. For the sake of this argument consider a scenario where a person is involved in a car crash. In turn, they receive a few life threatening injuries and unfortunately become comatose. When they have awoken, they have a serious case of amnesia. They do not remember or recognize anyone in their family or friend group. In this state, this person would be considered to have broken their psychological connection with their previous self before the accident. Because of this, once released from the hospital, they now have a totally different personality than what they had before the accident. If what Peoples claims is true, that personal identity is found in the soul, which is unchanging, shouldn't this individual have retained his previous personality? For that reason, I consider Dennett and Parfits arguments to be more plausible than Peoples.

So, in consideration of the simulation argument, when discussing the issue of resurrection, it appears that resurrection/and afterlife can be possible. However, if the inhabitants of said simulations are repeatedly resurrected, they will lose their personal identity. This is because as they are continuously resurrected, their psychological continuity to their previous selves are broken. Furthermore, even if we were to agree like Parfit that psychological continuity is not what matters, but rather the psychological connections made possible by the memory cards suggested by Reinbach, then these individuals would still nonetheless be different people. This is because although these individuals will still be connected by previous memories, these memories are not are not <u>their</u> personal memories, but rather their "replicas." Furthermore, with time, it is also reasonable to suppose that these connections will become weaker with each resurrection within the simulation. And if so, then one could suppose that this may be a possible

condition for attaining a certain level of Godhood, as previously mentioned. In some theologies, having let go of one's earthly personal attachments they ascend to a state of enlightenment, bringing them closer to the stage of divinity.

How and Can We Determine Reality within the Simulation?

Although, if one's personal identity changes through resurrection, then how do their perceptions affect their new identities, since personal identity stems from the experiences had by an individual?

Consequently, one of the many questions that stems from this argument on the simulation is how to determine reality. Bostrom claims that if we are living in a simulation, then the world that we see is just one small aspect of the "totality of physical existence."[33] Furthermore, there is no certainty that the world inside the simulation will resemble the "actual" cosmos. So, while our perceptions of the world may seem real, "they are not located at the fundamental level of reality."[34] Bostrom argues that there may be different levels of reality. Though, he also argues that to have many levels of reality would prove to be too expensive so, our simulators would most likely end our simulations once we have become posthuman[35]. So who are simulators if not our gods?

Many philosophers have attempted to define what constitutes this "reality. " This is exemplified through philosophers such as Hilary Putnam, Plato and Rene Descartes' whose

[33] Bostrom, Nick "Are We Living Inside a Simulation" pg 11

[34] Bostrom, Nick "Are We Living Inside a Simulation" pg 11

[35] Bostrom, Nick "Are We Living Inside a Simulation" pg 11

arguments on reality all help us to think about Bostrom's argument because they allow us to think about the possibility that there is an existence of another reality that is much more real than the one we are currently perceiving. On the other hand, Berkeley's argument seems to undermine Bostrom's argument because he claims that the current reality that we are perceiving is the only reality. And so, if we do live inside of a simulation, then this would be a part of our reality. For example, an apple inside of the simulation would be as real as apples inside of the real world because each is still made up of ideas and ideas are objects of knowledge.

In this section we will begin with Plato's argument for reality that is made through his Allegory of the Cave and then we will discuss Putnam's Brain in a Vat Scenario and, lastly, Descartes' evil demon concept.

In Plato's *Republic*, Plato presents the readers with his Allegory of the Cave and sets the scene for his audience. There is a cave that has a large opening where the moon and sunlight can shine through. Inside the cave, there are people who are bound in chains and are unable to turn their necks, and so, they can only face forwards.[36] Behind them, there lay a great fire with a small wall in between it and the people. On this wall, you can see people and things moving and you can hear some people talk and some who are silent[37]. Here, Plato argues that what we perceive through our senses is our reality. These images and sounds are the only reality these individuals have known, and because they are unable to free themselves, they are forced to rely on their senses to tell them what it is they are perceiving. And so, what we currently sense, having no other options is our present reality. Next, he presents a situation where one of the people in the cave is freed from his chains and finds himself outside of the cave. Once outside, the individual

[36] Grube, G.M.A, *The Republic.* 514b
[37] Grube, G.M.A *The Republic.* 514c

is greeted by the light, and recognizes that the shadows inside of the cave were mere illusions caused by the fire and wind. Having been freed from his shackles, the individual is now able to recognize that what he had previously thought to be people was just one reality of the situation, his former reality. In other words, Plato argues that things outside of our senses are not our present reality and that there is a truer "reality" that is just not known to us. And because this reality is unknown to us, even if it is more "real" we are still inclined to call the reality we currently perceive as "our" reality.

However, George Berkeley, Irish philosopher and thinker, would argue that there is no truer reality than the current one we perceive or rather there is no other reality at all. The beginning of his argument in *Principles Concerning Knowledge* is that there are no material substances and that all objects are just collections of ideas. Ideas are objects of knowledge and their existence depends on being perceived. He states "all ideas are either imprinted on the senses or perceived by attending to one's own emotions and mental activities"[38]. The entity that perceives these ideas are "minds, spirits, or souls"[39]. So, for an idea to be, it has to be perceived. For that reason, you can safely say that our reality is dependent upon what we perceive and experience. Accordingly, if we were living inside of a simulation we would not be aware of this fact because the "simulation" would be our only reality. For the objects within this realm would be just as real as objects within the real world. "Therefore, because we can't possibly see or feel a thing without having an actual sensation of it, I also can't possibly conceive of a perceptible thing distinct from the sensation or perception of it"[40]. This claim from Berkeley suggests that to have an idea that has a real existence separate from what is perceived would be inconceivable.

[38] Berkeley, George. *A Treatise Concerning the Principles of Human Knowledge*. Pg 11
[39] Berkeley, George. *A Treatise Concerning the Principles of Human Knowledge*. Pg 11
[40] Berkeley, George, *A Treatise Concerning the Principles of Human Knowledge*. Pg 11

So, if there was a simulation, we could not assume that this "reality" was false because we have not perceived this "true reality." Therefore, we could not conceive of the existence of an additional reality because we our current one we would be our only idea of reality because simulated or not, all objects are made up of ideas.

Contrary to Berkeley's claim that we would not be aware that we were living inside of a simulation, based on the fact that our current reality is the only one we can perceive, proposition (3) suggests that if we are living inside of a simulation, we would be aware of our existence within the simulated world. Granted, Bostrom offers the argument that the creators of the simulation could be simulated beings themselves. As previously explained, Bostrom claims that the reality inside of the simulation would only be a fraction of the total physical existence of the universe. This implies that we would also be aware of this idea of an external world separate from the simulation.

Yet, Hilary Putnam, an American philosopher and mathematician in his Brain in a Vat scenario, suggests that if we were in a simulated world, we would not be able to know that we were inside of one. The Brain in a Vat scenario supposes that our brains are removed by a mad scientist and are connected to a super computer, which then sends signals to our brains to stimulate an external world. The simulated world the brains are placed in are indistinguishable from reality. For example, if I attempted to raise my hand, the vat would send a signal to my brain, and cause me to see and feel my hands being raised. Though, Putnam calls this a "self-refuting supposition, whose truth implies its own falsity"[41] because although the people within the vat may be able to say and think of the same words that we do, such as "I am a brain in a vat," their understanding of what that word means is not the same for us, who are in the

[41] Putnam, Hilary. *Brains in a Vat*. 1981. Pg 7

external world. He claims that the ideas or words said inside of the simulation would have little causal relation to the "true" definition of the word used in the external world, that it does not even suffice to be called a reference. This is because the evil scientist could manipulate our understanding of this world. He refers to Turing's test on the reference of words and gives the scenario where a man is made to talk to another "real" person and a machine to see if the machines could possess consciousness or intelligence. The machine talks about apples that are in New England. However, because the machine has never held an apple, nor tasted an apple, she calls the conversation no more than "syntactic play that resembles intelligent discourse."[42] This is a consistent theme because Putnam then argues that if a brain that is in a vat claims that there is a tree in front of them, they would not be referring to "real" trees because they are incapable of thinking of such. Putnam's argument seems to suggest that the "simulated world" may be a totally different reality from the outside world, as she consistently makes that distinction between the reality of the brains in a vat and the external world. Which is also consistent with a part of Bostrom's argument, that there is no certainty that simulation would resemble the actual cosmos.

Rene Descartes, a French philosopher, presents a similar argument to Putnam with the evil demon concept in that in both scenarios there is a possibility that there is a person who is tricked into thinking that there is a certain sort of physical world all around them when in fact that isn't the case. For example, Descartes argues that although it may appear to him that he is sitting by a fire, the sensations we receive through this experience could all be an illusion or dream. Descartes reasons that an evil demon may be causing him to believe that he is sitting by a fire, when in reality that is not the case. So, if he is being tricked by an evil demon, then all the propositions that he holds to be true are false. Therefore, in order to know if a proposition is true,

[42] Putnam, Hilary. *Brains in a Vat*. 1981. Pg 11

Descartes would need to rule out this possibility. However, because Descartes cannot be sure there is not an evil demon deceiving him, he cannot rule it out as a possibility. And so, he ultimately concludes that the only knowledge he can be sure of is that he exists because to even suppose or think there is an evil demon, he must exist.

It's interesting to think about Putnam's arguments when considering Descartes' evil demon concept because in both cases, there is a possibility that their surroundings or perceptions are being manipulated by an outside force. However, unlike Descartes, in Putnam's argument even if he was being deceived by an evil scientist, what he believed about his world and his perceptions would not be false. Additionally, both scenarios depict the difficulties of determining another reality than the one you are currently perceiving. This is because this "real" reality is outside of our sense perception and if it is real, we would have no way of knowing that it existed.

Considering all of the arguments made by Plato, Berkeley, Putnam and Descartes, there seems to be a consensus between Berkely and Descartes that if we were living inside of a simulation, we would not be aware of it and I agree. Given that you are placed inside the simulation unknowingly, and have no way of distinguishing the difference between your current reality and the reality outside of it, if we were living inside of a simulation, we would not be able to know.

Can Inhabitants of the Simulation Exercise Free Will?

Let's continue under the assumption that we are living inside of a simulation. Given that assumption, does free will exist? Marcus Arvan, an Assistant Professor of Philosophy at the University of Tampa, argues that we do. In his paper, "Peer-to-Peer Hypothesis and a New

Theory of Free Will," he argues that we are not bound by the physical laws of nature and that we in fact do have free will or rather, Libertarian free will. He first claims that within a simulation or a video game the "appearance of determinism or causal closure can actually be an illusion of sorts generated by (B) causal interaction in a higher-reference frame not determined by any law of physics within the simulation"[43]. This argument made by Arvan supposes within the lower reference frame, it appears to the inhabitants of the simulation that all of their actions are causally determined by the laws of nature.

However, if it is true that we are living in a simulation, it could nevertheless be the case that we have free will because of the possibility that reality at the deepest level (i.e. reality outside of our simulated world) is **not** entirely causally determined. Therefore, the appearance of a causally determined world would be an illusion because while their actions within the simulation may seem to be inexorable, they are actually being influenced by the higher reference frames, where actions are not governed by the laws of physics. Arvan gives an example of playing a video. While playing the game, there are several ways the game can end, and in the game it may appear that all ends are causally determined. However, as game players, you would have the ability to select what actions to take within the game. Therefore, though the ending of the game may seem to be determined, it is actually being influenced by outside factors that are not limited to the laws/rules that are placed within the game.

Likewise, the German philosopher Immanuel Kant similarly argues for the distinction between the two realms or reference frames as Arvan calls them. The first being the phenomenal realm which is the world of appearances that we experience. Meanwhile, the noumenal realm is the world as it is in itself, beyond our sensory perceptions. He argues that while our actions

[43] Arvan, Marcus. The Peer-to-Peer Simulation Hypothesis and a New Theory of Free Will, pg 7

within the phenomenal realm may be bound by the causal laws of physics, within the noumenal realm, we are free to act autonomously in accordance with rational principles. However, Kant also argues that this does not mean that we are totally free from the laws of causality, but rather free to act autonomously within reason with these principles. The point here is that within the noumenal realm, our actions are not entirely causally determined, a point which Arvan also makes. Referring back to the example of the video game, within the noumenal realm, or high reference frame, we understand that the phenomenal realm/lower reference frame appears to be governed by the laws of nature. And so, because there are these rules within this frame/realm, we must still act within reason of these principles, though our actions are our own to make. For example, within the phenomenal realm, 2+2 will always appear to equal 4. However, in the noumenal realm, I may want 2+2 to equal 5. However, although I am free to act on this interest, 2+2 will always equal 4. Based on this argument, what seems to exist in both realms is the law of reason, which Kants call the "fact of reason."

Though, to determine the plausibility of Kant's and Arvans arguments , we shall take a look at John Milton's argument for free will presented in *Paradise Lost*.

In his epic poem *Paradise Lost*, Milton argues that mankind is free to operate as free agents within this realm in conjunction with God's divine providence. This is exemplified through his retelling of the story of The Garden of Eden. In the Garden, Adam and Eve are told by God that they are forbidden to eat the fruit from the Tree of Knowledge. Through Divine Providence, God has already seen that Adam and Eve shall disobey his commands. However, Milton argues that Adam and Eve both made the "rational choice" to eat the fruit. Eve, having been deceived by the snake, eats the fruit of her own will, a choice made by her faculty of

reason. Similarly, Adam, knowing the consequences of their disobedience to God's commands, also eats the fruit from the tree. Here, Milton argues that their ability to rationalize is a prime example of their free will. He states in *Areopagitica* "Foolish tongues! when God gave him reason, he gave him freedom to choose, for reason is but choosing"[44]. Milton argues that God has given man the right to choose, and in choosing is reason. To choose, one must be able to deliberate amongst himself to decide which action to act, weighing the options of each decision. And so, in the ability to choose and reason we find freedom. This is similar to the arguments made by both Kant and Arvan , in that when combined, reason and choice, we find freedom, which only resides in the noumenal realm/higher reference frame/true reality. This is because in the phenomenal realm/lower reference frame/the simulation, it is only reason which resides there.

Though Kant's position on determinism is quite different from Milton, in that he does not believe free will and determinism to be compatible, I do agree with Milton and Kant on the point that our ability to reason, meaning our capacity to think, analyze and make rational decisions, is what makes us free. However, when considering the position of the inhabitants of the simulation, based on the arguments provided by Kant and Arvan, I must contend that within the simulation there would only be limited freedom. While bound by the laws of physics within the simulation, reason remains even if it is only to a certain degree in that it is within the bounds of nature. However, when considering the idea of *true* freedom, this can only be found within *true* reality because in that realm, there are no laws that govern it, unlike those within the simulation.

[44] Milton, John. *Areopagitica*. Pg 28

Therefore, considering the arguments provided by both Kant, Arvan and Milton, we can assume that within the simulation, an individual would still retain some form of free will that is made accessible through their capacity to reason. However, this freedom would be limited and subjected to the laws of physics within this realm that is determined by their creators.

Furthermore, because the inhabitants of the simulation are within this realm, as Kant argues, they would not be able to grasp or attain the knowledge of the noumenal realm because they would be limited to only knowledge of the phenomenal realm. Within the phenomenal realm, we perceive ourselves as making choices and decisions based on our understanding of the world, but this perception of freedom and agency is contingent on how things manifest in the realm of appearances.

> "But the cause on account of which, not yet satisfied through the substratum of sensibility, one must add noumena that only the pure understanding can think to the phaenomena, rests solely on this. Sensibility and its field, namely that of appearances, are themselves limited by the understanding, in that they do not pertain to things in themselves, but only to the way in which, on account of our subjective constitution, things appear to us"[45] (Kant, 1999)

Therefore, according to Kant, our true freedom is limited by the boundaries of our understanding and the inherent constraints of our subjective perception. Our actions may be free insofar as they align with our rational and moral principles, but they are ultimately shaped by our cognitive limitations and the way in which phenomena appear to us. In essence, Kant is highlighting the

[45] Kant, Immanuel. *Critique of Pure Reason*. Cambridge University Press. 349

idea that our perception of the world is mediated by our senses and understanding. We can only know phenomena, which are appearances shaped by our mental faculties, and we cannot directly access the underlying reality or noumena, where there is hope for free will.

While we are limited to only perceiving phenomena, shaped by our mental faculties and thus unable to directly access the underlying reality or noumena where free will might reside, Arthur Schopenhauer presents an intriguing perspective.

German philosopher Arthur Schopenhauer would agree with the claim that true freedom is unattainable to those within the simulation. However, he argues that instead of reason/perception, will is the driving force of reality. He argues that will is blind, unconscious and irrational. Will, as he defines it, acts for without any reason. It simply does. He states that "man does at all times only what he wills, and yet he does this necessarily. But this is due to the fact that he already is what he wills."[46] Essentially, Schopenhauer is arguing that every action that an individual makes is because he wills himself or something to do so. He claims that we tend to mistake this "will" for freedom because we believe that our decisions are completely made without any restraints. "This illusion rests on a true and self-evident datum of our self consciousness, which commonly is expressed as "I can do whatever I want."[47] However, according to Schoepnahuer this is a false ideal to have, in that because everything that man does or acts upon is done so necessarily, that means that there are restraints in his actions. He states that "if we consider his behavior objectively, i.e., from the outside, we shall be bound to recognize that, like the behavior of every natural being, it must be subject to the law of causality

[46] Schopenhauer, Arthur. *Essay on the Freedom of the Will*. pg 99
[47] Moreno, Agustin. "Schopenhauer On Free Will and Fatalism." pg 1

in all its severity."[48] Similarly to Arvan and Kant, Schopenhauer seems to agree that an individual's action is bound by the laws of nature. Though, he claims that he is not arguing that freedom does not exist, but rather it is something that is out of our reach.

> "Consequently, my exposition does not eliminate freedom. It merely moves it out, namely, out of the area of simple actions, where it demonstrably cannot be found, up to a region which lies higher, but is not so easily accessible to our knowledge. In other words, freedom is transcendental"[49]

Similar to Arvan and Kant, Schopenhauer distinguishes between the different realms. He claims that freedom is transcendental and that it is beyond our reach. Kant and Arvan make similar claims in that they believed that "true" freedom was found in a "higher" realm, which was almost unattainable and that the freedom in which we experienced in our current realm, is limited in that it is still bound by the laws of causality. True freedom, as Schopenhauer states "freely manifests itself in space and time as a multiplicity of individuals and actions, me and mine included. But I, as an individual, am not free at all."[50] Here, Schopenhauer seems to suggest that the will is separate from the individual. As described above, the will is based on motives, it simply does because it will's itself to do so. However, the individual is not capable of this freedom because he is bound by several restraints, whether physically, intellectually or morally. Therefore, he asserts that man himself is not free.

---

[48]Schopenhauer, Arthur. *Essay on the Freedom of the Will*. pg 99
[49] Schopenhauer, Arthur. *Essay on the Freedom of the Will*. pg 99
[50] Janaway, Christopher. "Necessity, Responsibility and Character: Schopenhauer on Freedom of the Will." pg 455

Hence, if an individual is within the simulation, based on Schopenauer's argument, they would not have free will, but rather their actions could be attributed to their ability to will things to be. However, if we consider that the creators of the simulation can influence an individual's will or desire, then based on his argument, individuals within the simulation would be nonautonomous, having neither free will nor the ability to will anything.

Suppose his argument was found to be true, then free will would be non-existent within the simulation. As a result, inhabitants would be like puppets who are being controlled by their puppeteers. Though, this begs the question of whether free will is a necessary condition for living? However great of a question, I will not explore this further in this section for the sake of focusing on the argument at hand.

While Schopenhauer offers a new perspective on the idea of freedom within the simulation, his argument seems less reasonable compared to Kant and Arvans. One of the main reasons is that he believed that the driving force of reality, will, was irrational, and that all actions are done out of necessity. If this is true, then our actions would have no meaning behind them, except necessity, which is loosely defined by Schopenhauer as something that "follows from sufficient ground"[51] Meanwhile, based on the arguments made by Arvan, Milton and Kant, reason is what drives forth our actions and reason gives meaning to those actions.

Nonetheless, having considered all of the arguments presented by Schopenhauer, Kant and Arvan, I contend that within the simulation, an individual would still have free will, though limited. Although, if this argument is true, how does this affect the existence of evil?

[51] Janaway, Christopher, "Determinism and Responsibility" Abstract

Problem of Evil:

Traditionally, the problem of evil is described as; if there is an omnipotent, omniscient, and essentially a morally perfect creator of the universe (God), then why does evil exist? And if there was such a God, then why does he not prevent these evils? One response to this argument is that evil persists in this world because there is a certain amount of goodness that results from evil. Gottfried Wilhelm Leibniz, a German philosopher, provides an example of this using Judas, the disciple who betrayed Jes*us,* in his *Discourse on Metaphysics.* In this example, Leibniz explains that Judas betrayed Jesus for 30 pieces of silver. Thus, Jesus was imprisoned, beaten and then killed. However, he further explains that God still allowed for Judas to be created, even though he knew that he would betray him because he foresaw the great goodness that would result from his betrayal, which was the saving of the world from sin. Therefore, Leibniz argues that evil exists because there is a great goodness that results from these events that we as finite beings cannot comprehend because they exceed human reasoning.

However, in the case of the creators of the simulation, if we assume that they have achieved this sense of "Godhood" then can we assume that evil persists in this world for the same reasons? Based on the evidence we have already collected, one can conclude that we would not be able to make this claim. This is because even if Bostrom is correct in claiming that the creators are both omnipotent and omniscient, we do not have reason to believe that they are morally perfect because of the ethical implications of the simulation. Although one could claim that there is some greater goodness that would result from their actions, based on what we know about the creator's, this idea would not be consistent with their nature. This is clear if we assume

that they have created the simulation. In doing so, the creators reinforce the idea that they view the inhabitants of the simulation as mere tools meant to be used to advance their own interests. Though, one could claim that there can still be some goodness resulting from the creators' views of the inhabitants. However, this argument would be hard to prove because one of the necessary components for this argument would be for the creators to also be “all good” in addition to being omnipotent and omniscient like God. But because they are not all good, it cannot be said that there can be any greater goodness resulting from their actions.

Furthermore, because the creators are assumed to be posthumanisms, we cannot reasonably claim that they are morally perfect beings because they have abandoned their humanity. Therefore, corrupting their moral characters. Thus, they are prone to mistakes, and capable of evil. As a result, evil persists within the simulation, because in part, the inhabitants are an extension of their creators in a relevant way. Because their creators are capable of evil, so are they. Furthermore, considering the previous arguments for the case of free will, assuming the creators also have this privilege, it would be reasonable to conclude that because of this ability given to them by their faculty of reasoning, they themselves and their inhabitants are capable of committing evil.

For a real-world example, let's take a look at The Third Crusade War (1189-1192). The Third Crusade was led by Richard the Lionheart and was an attempt to retake the crusader states, particularly Jerusalem. The land had fallen to the then, Muslim Leader, Saladin. Jerusalem had a profound impact on the Christian world and the capture of it sparked outrage throughout Europe. Jerusalem had fallen, and there was a desire to keep control of the territories of the Holy Land. The motivations for this war can be said to have been employed by reason. Though the outcome

of the war may have been predetermined, it was mankind who made the choice to begin the war. And because these choices were made based on reasons, that is what makes it a free choice. With that said, The Third Crusade was a truly bloody war with the estimated casualties ranging from 50,000 to 100,000. Given that we have free will, we are capable of committing evils. This is because through our faculty of reasoning, we are able to deliberate and make these conscious decisions of our own free will, and it is because we can reason, that we can even do so. However, as stated previously, this free will is limited to the laws of nature.

However, it is through this same ability to reason that we commit evil. When we partake in an evil action, we are consciously making that decision to do so. It is not a decision that we make irrationally, because as stated before by Milton, when Adam and Eve ate the fruit of the tree of knowledge, they knew of the consequences of their actions and still chose to eat the fruit. And so just like Adam and Eve's evil was committed by reason, so was the evil committed by the Templars' previous example. Kant in his Critique of Pure Reason famously stated that

> "A man may dissemble as much as he will in order to paint his recollected unlawful behavior ... as something to which he was carried along by the stream of natural necessity, and in this way try to make himself out as innocent. But he finds that the advocate who speaks on his behalf cannot silence the accuser in him when he is conscious that . . . when he committed the wrong he was in his senses, i.e., he was in possession of his freedom."[52]

[52] Kant, Immanuel. *Critique of Pure Reason*.Kant, pg 102-103

Essentially, Kant argues that when one is conscious of the moral laws that governs him (which reason provides), and chooses to act on his desires (maxims), he does so autonomously. Furthermore, Kant also argues that if we commit an act that we deem to be morally impermissible, and act on it because it is morally impermissible, then our actions are autonomous, made of our own will. And so, if a person commits an evil action, and tries to claim that he is innocent, on the basis that it was a phenomenal self, he would be wrong. This is because once again, if he is conscious of the morals laws and still commits an act of evil, then he was in possession of his freedom. Therefore, on Kant's account, he seems to suggest that if a person's phenomenal self were to commit an evil action, he would not be morally responsible for this action, as it was causally determined, and so moral responsibility seems to only lie in the noumenal realm.

But, what is to be said of natural evils, such as hurricanes and tornadoes? Who is at fault? Dustin Crummet, a Postdoctoral Fellow of Philosophy at the University of Notre Dame argues that these natural evils are "results from the actions of whoever created our environment"[53] He claims that moral evils are "creaturely"[54] while natural evils are not. However, Crummet makes the interesting claim "if the simulation hypothesis is true, then there may not be any genuine natural evils."[55] These apparent evils would just be symptoms of the system or recklessness from the creators. He argues "Maybe, for instance, our programmers didn't realize we would really be

[53] Crummett, Dustin. The Real Advantages of the Simulation Solution to the Problem of Natural Evil. Pg 620
[54] Crummett, Dustin. The Real Advantages of the Simulation Solution to the Problem of Natural Evil. Pg 620
[55] Crummett, Dustin. The Real Advantages of the Simulation Solution to the Problem of Natural Evil. Pg 620

conscious and harmed by the evils which their programming allowed."[56] Though, I find this argument rather hard to believe. If these simulated beings were created and programmed by these creators, then they must have known the fragility of their creations. If you create something, you must at least know its limitations because you created and designed it to be that way. Therefore, in this case, "the fault lies with the designers of the virtual world in question. In selecting the world-generation parameters as they did, the programmers of our world ensured that our world is as plagued by natural evils as it is."[57]

Therefore, reviewing the arguments made previously, evil exists within the simulation because of the ability to reason, which is employed by both the simulators and their creators. Furthermore, even in the "natural" occurrences that happen within the simulation, the creators are still at fault, as they are able to choose whether to input these events into the simulation. Therefore, we can also conclude that the creators of the simulation are not moral beings, and could possibly be evil beings.

Reward and Punishment System within the Simulation:

Though, if the simulation is as Bostrom claims, that the inhabitants can be rewarded or punished for their behavior within the simulation, how would this system decide what is good or bad behavior? Would this consist in the form of a point system? For example, in the show "The Good Place," a show which explores morality and redemption, a person's score determines whether they went to a good or a bad place after death. Once you achieved a score of more than a

[56] Crummett, Dustin. The Real Advantages of the Simulation Solution to the Problem of Natural evil. Pg 621

[57] Dainton, Barry. "Natural Evil: The Simulation Solution."pg 212

million, your place in the good place was secure and if you had less than a million points you were sent to the bad place. However, in that show, the point system was deemed to be flawed because when determining points for things considered “bad” it was based on a domino effect. For instance, if you bought flowers from a florist and gave it to your wife, you could receive 1000 points, but if those flowers were grown as a result of unpaid labor then you could end up receiving -1000 points. And if the creators of the simulation used a similar system, how would they determine how much points an action could be worth. For example, saving someone and simply holding the door for someone behind you should have significant differences in their points.

Furthermore, if we assume that based on the number of points accumulated by each individual determines their place in the afterlife, how would this system work in terms of resurrection. In our previous section, we explained that for Reinbach, the afterlife in the simulation would just be computational resurrection, meaning when an individual died, their data would be restored and placed within a new digitized body or an identical replica of their previous body. So, if resurrection is the afterlife for these individuals, would this change the nature of their resurrection. Could they be given new lives that were worse than the ones they previously lived if they only had a few points? If so, this process could be said to be similar to the idea of karmic retribution. Karma in itself is described as “any action done by an agent produces a fruit , which ripening (Sanskrit: vipāka) is gradual, and will fall back on the agent once achieved.”[58] Essentially, this means that any action that a person does, whether good or bad, they will receive back to them. For good actions, their results would be good, and for bad actions taken, they would receive painful consequences. This is essentially important for the Buddhist religion. In

[58]Richard, Frederick. Karmic Retribution in Theravada Buddhism: A Way to Salvation? Pg 1

this religion, a person's actions on earth affects their next life after resurrection. Degrees vary in their punishments; if their actions were good, they could come back as a human and live a good and comfortable life. If their actions were bad, they could come back to life as an animal or be sent to hell.

Considering Karmic retribution in regard to the simulation, it allows us to gain an idea of how this point system would work with the simulated work. Furthermore, it offers a perspective on how the punishments and rewards could vary depending on the points accumulated or the amount of good/bad actions taken. Though, one may question whether the creators of the simulations would bother to reward or punish the inhabitants of the simulation, considering that they regard them as mere tools.

However, if we look at history, specifically the institution of slavery within the United States, where African bodies were considered to be mere tools used for economic gain, we can see that there are instances where slaves would be rewarded for good and bad behavior. In some cases, if a slave acted out of line, they would be brutally punished by having their backs whipped, children taken or in some cases even killed. On the other hand, if a slave behaved well and produced favorable results, their master could, if they chose to, reward them by providing them with extra food for that day's dinner, or even in some cases, allowing them to get married to their partners that resided on the plantation.

Therefore, having historical evidence of people who were considered tools, be given rewards/and or punishments, it is reasonable to suppose that the creators of the simulation would reward and/or punish the inhabitants of the simulation for their behavior within it.

Ethical Considerations of Simulation Argument:

Given the actions of the creators of the supposed simulations, can we in truth, consider them to be moral beings? The creators of the simulation may believe that they are. They could claim that they created these simulations in an attempt to save humanity or to understand our existence better. But at what cost? Their actions can be considered unethical because they violate the right to bodily autonomy and strip away individuality. Critics of the simulation argument can say that a human life should not be toyed with because we are different from beasts. We can talk, we can reason and our "freedom" is what makes us different from other species. Furthermore, our finite existence is what makes us human. We live and then we die, it is the natural course of life. When external factors influence this course, they become judge, jury, and executioner of these beings. But this can still be considered to be a morally gray area because if we look at the concept of utilitarianism, it states that we should do what is best for the common good, minimizing pain and maximizing pleasure. If they can save humanity at the expense of a few, their actions could be considered justified. However, if these simulated beings are unaware of their situation of being within a simulation, is it right for them to live their lives in a complete lie? Or is ignorance truly bliss. Though, if it is true based on the arguments of Kant, Berkeley, and Putnam etc that we would not know that we lived in a simulation, then would this change the gravity of the crime? If we are not aware of being inside of a simulated world, then there is not much to be done. However, one could still argue that although the inhabitants of the simulation are unaware of their situations, the actions of the creators are still morally impermissible. But if this is true, then who will judge the creators and seek justice for the sake of the simulated beings? If it is true, as Steinheart claims, that God is the ultimate creator of the creators of the simulation, then the responsibility lies with him. However, if there is not an ultimate creator, then

the creators would simply be able to act freely without restraint. Therefore, humanity will be subjected to the will of the creators, reducing them to the status of mere tools.

Furthermore, as stated in previous sections, Heidegger argues that we should be wary of technology because with our further use of it, we will over time begin to see human beings as resources themselves. As a result, we will distort the historical definition or the value of a human being. In the journal article, *From Post-Humanism to Ethics of Artificial Intelligence*, by Rajakishore Nath & Riya Manna, Natha and Manna present an argument made by Frederick Nietzsche on the concept of posthumanism. Nietzsche argued that

> "overman', [or posthumans] are beings who will be devoid of human limitations. He will be the next Godly person to the commons, who can navigate his peers to a comparatively free society. Nietzsche has argued that his 'overman' would possess some extraordinary qualities which are beyond the capabilities of any ordinary human being"[59]

Similarly to Bostrom, Nietzsche argues that post-human beings would be God-like and will be beings who are much more capable than regular human beings. In his perspective, these beings will not be held by any restraints. From the tone of Neitzsche's argument, it seems to suggest that the creation of post-humans will allow for true freedom. However, if a society such as this existed, one could argue that it would not be an orderly society. Hobbes in *Leviathan* argues that man's natural state is of war. And, if posthumans were to be created, eventually there would be a divide between man and the quasi machine. For instance, due to the superhuman qualities of

[59] Nath, Rajakishore, and Riya Manna. "From posthumanism to ethics of artificial intelligence." Pg 1

posthumans, most likely they would take over the workforce since they are able to work faster and more efficiently. Nath and Manna state that

> "If posthumans are proved to be a better performer than humans, then it will be a huge blow to anthropocentric superiority as well. If there will be a shortage in scope for human labor, then the posthuman AI robots will be in direct competition with human beings regarding the efficiency to act. Certainly, the former will replace the latter."[60]

Therefore, the value of human beings will decrease and Nath and Manna explain that this will directly impact the anthropence view. This view asserts humans as the dominant force within nature. However, with the creations of posthumans, humans will no longer be the superior beings in nature. In fact, their value could decrease to the level of nature itself. Thus, we again are reminded of Heidegger's warnings of technology. If society moves forward with creating the post human beings, society itself may fall apart. This is because there could be uprisings that are caused based on these creations. Humans will be pushed out of society and even the workforce, losing their position as the dominant force within nature. Furthermore, there's a majority of people who would disagree with the creation of post-humans because of the fear that they will take over society. And these individuals' fear can be justified.

Over the years there have been many movie productions which depict posthumans overthrowing society. One example that comes to mind is the movie *The Matrix*. In this movie,

> "the machines dominate, subjugate, enslave, and seek to eradicate human beings. They are the enemy of humanity, inimical to human life and liberty. As the computer program

[60] Nath, Rajakishore, and Riya Manna. "From posthumanism to ethics of artificial intelligence." pg 190

> known as Agent Smith remarks in M1, "As soon as we started thinking for you, it really became our civilization, which is, of course, what this is all about: evolution."[61]

Here, these post humans take control of their societies, and assume dominance over human beings. Though one may call these productions to be propaganda, one cant' simply ignore the real concerns that these movies address. If we truly do go on to make post humans, then our position as human beings will be threatened. Ultimately, our own creations could one day come to eradicate us.

Furthermore, with the creation of these beings, the probability of ancestor simulations being created will become higher. If this is true, then it is likely as Bostrom supposes, that posthumans will be the ones to create these simulations. Additionally, at this point, humans will have the necessary technology to create it. Given that these post-human creators will have a distorted perspective on humanity, one could say that their morals would also be different from our own. Therefore, it may seem acceptable to these beings to create these simulations, maybe in an attempt to replicate their previous societies to explore civilization prior to posthumanism. However, even if these beings had good intentions for creating these simulated worlds, it would still not justify their actions. This is because in creating these simulations, as previously stated, it would impede upon the autonomy of the inhabitants of these worlds. Furthermore, as I previously argued on page 6, it would be cruel to put human beings through this process, especially without our knowledge. Although we would not be aware of being within the simulation, we would still be deprived of our right to liberty, to live actual *real* experiences. From a moral standpoint, this action would be impermissible. Although, considering our recent

[61] Van der Laan, J. M. "Machines and human beings in the movies." pg 32

developments in technology, there is a real possibility that in the future we could create these simulations. However, at that appointed time, I hope that the implications stated in this thesis will be considered and that we will not move forward with ancestor simulations, no matter how great the benefit may be for posthuman societies.

Artificial Intelligence: Friend or Foe?

Therefore, in this next section we will explore the possibilities of creating our own ancestor simulations based on our current technological advancements, with a slight focus on the ethics. I will first explore a few examples of recently developed artificial intelligence technology. Then, I will provide a few implications of doing so. Afterwards, I will then supplement this argument with a few preventive measures that are provided by those who support the advances in artificial intelligence research. Lastly, I will conclude this section with my reasoning for why these preventive measures may not be entirely effective.

Over the past few years, research on artificial intelligence has rapidly grown. In most studies, the goal of this research is to increase human capabilities. With such advances, it is possible that in the near future, humanity will be able to create their own ancestor simulations. For example, the company Hanson Robotics in 2015 created a humanoid machine named Sophia. Sophia, the humanoid robot, is able to hold and maintain intelligent conversations and mimic human behaviors and emotions. Her knowledge of the world also increases with each interaction she makes with someone. “Thus, Sophia becomes increasingly familiar with the culture, customs, feelings, emotions, and linguistic styles of her interlocutors. And all this experience is

accumulating in her memory."[62] Furthermore, her sophisticated demeanor has also allotted her to become the First Citizen Robot of the World, a name given to her by Saudi Arabia. With the developments of artificial intelligences such as Sophia, the age of posthumanism becomes closer. If scientists are able to produce humanoid robots who are capable of mimicking human behaviors and even producing them with human-like facial features, it is highly likely that in the future, this same research can be used to integrate artificial intelligence into human bodies. Furthermore, other companies such as NASA and ISRO (Indian Space Research Organization), have also begun to incorporate humanoid machines into their workforces. For example, in 2020 ISRO sent a humanoid robot called Vyomamitra for an unmanned space mission. The creation of humanoid robots such as Sophia and Vyomamitra are prime examples of how post-humans in the future could replace the human workforce, and serve as reminders of Heidegger's warning of technology. Furthermore, their existence also emphasizes the rapid growth of research in artificial intelligence, which might just lead to the creation of humans with more advanced and powerful capabilities, who in turn, could create a simulation. One example is the recent research into exoskeletons. According to Saba Samiei, Founder and CEO of MASCO Technologies, an artificial intelligence company, in "On the Dangers of Artificial Intelligence," argues that

> "There are currently multiple projects around the world working on developing exoskeletons to either perform tasks that humans are unable to do or to do humans tasks much better. For example, Francois G. Pin and John Jansen of Oak Ridge National Laboratory, Oak Ridge, Tenn have developed an exoskeleton that allows its human operator to load a 1000-kilogram bomb into an aircraft as if it were a 3-kilogram load"[63]

[62]Retto, Jesús. "Sophia, first citizen robot of the world." pg 6
[63] Samiei, Saba. "On The Danger of Artificial Intelligence." pg 52

These exoskeletons allow their human operators to exceed their human capabilities and complete tasks that would normally be impossible for a human. Thus, the creation of exoskeletons provide a bridge between artificial intelligence and humanity, whose path may lead to the creation of a post-human society.

In addition to that, there are many companies that are seeking to create artificial beings with "super intelligence" such as Google. Steven Livingston and Mathias Risse in their journal article "The Future Impact of Artificial Intelligence on Humans and Human Rights" state that

> "Google's DeepMind and Google Brain, along with initiatives at other major tech companies, are pushing AI toward superintelligence. Greg Brockman, co-founder and chief technology officer at the company OpenAI, recently noted that after fifty or more years of research, AGI (Artificial General Intelligence) now seems likely in the near term."[64]

For example, Google's DeepMind initiative seeks to outsmart humans through artificial beings with superintelligence with the aim to improve humanity. However, with the creation of these technologies, who is to say that these beings will not be the fall of humanity? Livingston and Risse claim that

> "Homo sapiens is not going to be exterminated by a robot revolt. Rather, Homo sapiens is likely to upgrade itself step by step, merging with robots and computers in the process . . . In pursuit of health, happiness and power, humans will gradually change first one of their

[64] Livingston, Steven, and Mathias Risse. "The future impact of artificial intelligence on humans and human rights." pg 142

> features and then another, and another, until they will no longer be human." The process will most likely unfold gradually, with the most well-resourced and politically connected sectors of society availing themselves and their offspring of the advantages of bioengineering in much the same way they currently avail themselves and their offspring of the advantages of better healthcare, schools, housing, and leisure time. The trends in wealth inequality identified by Thomas Piketty will produce even starker distinctions between the haves and the have-nots."[65]

According to Livinston and Risse, the emergence of post-humans will in fact better society, allowing them to advance and evolve into higher beings. However, they also speak of the effects of doing so. One implication is that this would create a wealth disparity between those who are financially at the top and those who are at the bottom. To be able to start the process of post-humanism would likely be an expensive venture. Thus, only those who are able to afford it will be able to partake in this opportunity until scientists are able to create them at cheaper costs. Again, this highlights the dangers of this technology, because if only certain groups are able to access this technology, then there will be a stark divide in society, and those at the bottom may feel forced to revolt. However, this may also end in the destruction of humanity because against post humans, humanity will be powerless. Although, some supporters of this venture such as Elon Musk, claim that "the only hope for human survival in the face of inevitable AI superintelligence is to be found in inferior humans merging with advanced machines."[66] Bostrom and Vincent C. Miller describe superintelligence to be

[65] Livingston, Steven, and Mathias Risse. "The future impact of artificial intelligence on humans and human rights." pg 152

[66] Livingston, Steven, and Mathias Risse. "The future impact of artificial intelligence on humans and human rights."pg 152

> "any intellect that greatly exceeds the cognitive performance of humans in virtually all domains of interest." One idea how superintelligence might come about is that if we humans could create artificial general intelligent ability at a roughly human level, then this creation could, in turn, create yet higher intelligence, which could, in turn, create yet higher intelligence, and so on …"[67]

As the research in super intelligence continues to progress, the creation of posthumans becomes more likely. This is because although the first stage may be in creating superintelligence within the artificial beings, at some point, humans will yearn for this power as well. And with the creation of posthumans, comes the possibility of simulations. Though, again this raises the risk of post humans having less of a tolerance of those who are not enhanced beings, and this can lead to many human rights violations.

In spite of the dangers of superintelligence, Musk and his group at the company Neuralink Corporation, are in the process of developing a new neurolink technology that he claims would be able to aid in diagnosing and treating several cognitive illnesses. Furthermore, in Takudzwa Fadziso's "Why Neurolink will Change Humanity" he states that "Neuralink will interface individuals and machines to adequately empower synthetic telepathy…A full brain interface would associate with your mind and control each sign sent by your neurotransmitters."[68] At full completion, Musk's neuro link will allow users to telepathically connect and this just may be the first step to society's transition to post-humanism.

---

[67] Müller, Vincent C. and Bostrom, Nick 'Future Progress in Artificial intelligence: A Survey of Expert Opinion,' pg 2

[68] Fadziso, Takudzwa. "Why Neuralink Will Change Humanity Forever?." Pg 77

However, some critics may argue that even with the advances made in regards to artificial intelligence, there lingers little to no danger at all. For example, Ozlem Ulgen, Associate Professor of Law, specializing in artificial intelligence and ethics at The University of Nottingham, in her article "Kantian Ethics Concerning Artificial Intelligence" claims that

> "Artificial intelligence and robotics do not possess human rational thinking capacity or a free will to be able to understand what constitutes a rule that is inherently desirable, doable, and valuable for it to be capable of universalisation. But there is human agency in the design, development, testing, and deployment of such technology so that responsibility for implementing the categorical imperative resides with humans. Humans determine which rules are programmed into the technology to ensure ethical use and moral conduct"[69]

According to Ulgen, artificial intelligent machines do not possess the ability to rationally make decisions nor do they have free will. Furthermore, they are created by humans, and we have agency in how we develop these machines. Therefore, if these machines do end up creating high forms of artificial intelligence, resulting in post humans, then the fault lies in us. Ulgen's argument suggests that it is not artificial intelligence that we should be wary of, but rather of humanity itself.

Likewise, Samiei makes similar claims that artificial intelligence can not advance on its own, and that it is dependent on human agency in its development.

[69] Ulgen, Ozlem. "Kantian Ethics in the Age of Artificial Intelligence and Robotics." pg 72

> "intelligent machines that are not intelligent enough to comprehend the impact of their decisions on their surrounding world. More fundamentally, they are not yet intelligent enough to learn about the world without external intervention… AI is currently dependent on humans to learn and improve, this has meant that some of the AI researchers (personal communications, November 7, 2018) believe that AI will never get to a point where they can impose a danger due to its dependence on humans"[70]

Therefore, she claims that based on her research and conversations with other researchers, artificial intelligence will never reach a point where it will endanger humanity. Additionally, she includes solutions/precautions that researchers can use to ensure that this will not happen. First, researchers would need to program and develop these artificial machines in a friendly manner with programming to ensure that they do not develop any hostility towards human beings. To do this, researchers could use a special AI that would be able to study human behaviors and mechanisms in order to create more artificial beings that are similar in their friendly nature.[71]

However, even if precautions could be set within these artificial beings, there is always the slight possibility that they would still be able to pose a threat to humans. This is because although they will be designed and programmed by humans, humans are also prone to error since it is in our natures. Thus, if a human programmer were to make a mistake in programming these beings, then who knows the consequences that will follow.

Furthermore, as it is highly possible that we could develop into a post-human society, this emphasizes how likely it is that the post-humans will go on to create ancestor simulations in the

[70] Samiei, Saba. "On The Danger of Artificial Intelligence.", Pg 20
[71] Samiei, Saba. "On The Danger of Artificial Intelligence." Pg 29

future. However, as I argued earlier, one should remember that with the creation of these beings, comes the downfall of humanity for its definition will become distorted, and result in many other disparities between those who are enhanced and those who are not. Therefore, we should not continue research into these technologies. If all that lies at the end is our destruction, should we not take the necessary actions to prevent this? Or are the "benefits" too great for the costs that will come with it? If so, then society as a whole should re-evaluate how we view not only ourselves but the world itself.

Conclusion:

In this section, I will now review each of the implications and the conclusions drawn.

The first implication explored within this paper, under the assumption that we did live within a simulation, sought to determine the identity of the creators of this simulated world. In this section, I first reviewed Bostrom's argument where he claims that the creators were post-human beings, who were both omniscient and omnipotent, who are also likely simulated beings themselves. Then, I included the work of Steinheart who similarly argued that the creators were post-human beings. However, Steinheart also claimed that if this is true, then there must be another creator, one much deeper than the post humans, who created the post humans. He concludes that this creator must be God. Although, Steinheart also claims that if the simulation argument were true, then <u>our</u> world must have also been created by these post humans or an ancient civilization. However, as I argued, this would conflict with theories such as the Big Bang Theory of how the universe was formed. Therefore, based on these arguments, I concluded that it

is likely that if the simulation argument were true, then the creators would be post humans, who were created by God.

The second implication I explored was the possibility of an afterlife/resurrection existing within the simulation. In this section I argued that it would be possible by utilizing the works of Reinbach, Parfit, Dennett, and Peoples. However, I also claimed that an individual's personal identity would not survive this change. Reinbach claims that within the simulation, a person's sim could be computationally resurrected multiple times. And, the more times they are resurrected, the closer they get to God. I supported this argument by utilizing the works of Dennett and Parfit, who argued that personal identity does not survive resurrection. Therefore, if one's personal identity does not survive, this would allow the individual to progress further to the stage of a celestial being, who transcends both time and space, such as a god, who Bostrom likens the creators of the simulation to be.

For the third implication, I expired how and can we determine reality while within the simulation. I ultimately concluded that we could not determine reality within the simulation because we would not be able to conceive or perceive of another reality outside of the one we presently would be at. If we look at arguments provided by Berkeley, Putnam and Descartes, they all agree that because of our current perceptions on reality, we could not think that there is another reality that is more real than the one we are currently living in. For example, Putnam claims that even if we were to know the word simulation, the definition of the word would be different from the "true" meaning which is found in the "real" reality. Furthermore, Berkeley argued that it would be inconceivable to have an idea that would have a more real existence than the present one being perceived. This is because all objects are made up of ideas, and ideas are

made up of experiences. Therefore, if we have not experienced this more "true" reality, then we cannot know that it exists. Thus, if we are unable to conceive of another reality separate from the one we are experiencing, then we cannot determine reality within the simulation. We cannot even assume that we would have no basis to even question it.

Additionally, for the fourth implication, I explored whether or not inhabitants of the simulation could exercise free will. For this section, I concluded that inhabitants would be able to exercise free will but only to an extent, which would be employed by the faculty of reason. For example, I used the works of both Arvan and Kant to illustrate how within the simulation, actions could be considered to be causally determined. However, in the *real* world, these actions would not be entirely governed by the laws of physics, but rather individuals would be able to act more autonomously. This is because as Kant states, because our knowledge of the noumenal realm is rather ambiguous, it is likely that there exists true freedom there, compared to the freedom found within the simulation/phenomenal realm. In addition, I explore Schopenhauer's definition of "the will" which he describes as an irrational, unconscious and the driving force of reality. However, I argued that I disagreed with Schopenhauer's claim that will was the driving force of reality, because this would entail actions to be meaningless. Furthermore, rationality is an innate feature of humanity, and as Milton claims, something that was given to us by God. Therefore, will cannot be the driving force of reality. Having reviewed their arguments, I concluded that within the simulation, there was free will, and the inhabitants would be able to exercise it but only to a certain degree, as they would be bound by the laws of physics.

In the fifth implication, I consider the problem of evil, and explore how evil can exist within the simulation. In this section, I argue that because the inhabitants of the simulation

possess reason, or rather that reason exists within their simulated world, evil exists. Additionally, I argue that the creators of the simulation are also to blame for this behavior because they were the ones who created the simulation with these conditions, and that by default, their creations are extensions of them. Therefore, I argue that it is reasonable to claim that because the creators possess the ability to commit evil, so do their creations, the inhabitants of the simulation.

In the sixth implication, I question whether or not the inhabitants of the simulation could be rewarded and/or punished for their behavior. I supposed that if there was such a system, it was likely that it would be on a point scale. Although, I also argued that this scale could be used to determine how the inhabitants' next life would be after resurrection. I compared this process to the idea of Karmic Retribution, a term commonly found in Buddhist traditions. In theory, this meant that if a person committed good deeds while alive, in their next life they could be reborn as a wealthy human being. However, if they committed bad deeds, then the individual's next life would be full of hardships. Though, I also noted that some may criticize whether the creators of the simulation would actually reward or punish the inhabitants since they considered them to be tools. However, I argued that yes, it is still likely that the creators would because history proves it. I used the institution of slavery as an example, and noted that slave owners would reward and punish their slaves depending on how well they performed their duties that day. Therefore, I concluded that it was likely that the inhabitants of the simulation could be rewarded and/or punished for their actions and that it could be measured by a point system.

For the seventh implication, I present and discuss a few ethical implications of the implications of the simulation argument. One implication I discuss is how society will be affected by the creation of posthumans. One reason is because their emergence will threaten the

position of humanity as the dominant force. Now, post-humans will be the superior race, and will soon enough take away jobs from those who are not enhanced, leaving a stark economic divide between the races. Additionally, if post humans go on to create a simulation, it would threaten the autonomy of the people and disrupt the natural flow of life. As humans, we are finite beings, and if we can be resurrected(computationally) within the simulation, we take away what makes us human in the first place. However, I also question if that is the goal of the posthuman creators in the first place. At the end of the section, I conclude that the creation of posthuman society should not be attempted. Also, if they did create a simulation, from a moral standpoint, it would be morally impermissible due to the effects it will have on humanity.

Lastly, in the eighth implication I present many examples of technology that showcases how close society is to developing a posthuman society, which in turn would lead to a simulation. One example I provided was of Elon Musk's new project, Neurolink, which is a microchip technology that is implanted into the brain. According to Musk and his team, one of their goals for this technology is telepathy between Neurolink users. Upon finding this, I found it to be a high example of how far technology has progressed so far and how close researchers are to progressing society to a posthuman stage. Their enthusiasm for this research illustrates that a posthuman society, or a simulation can no longer be posed as a question of if, but of when.

Throughout this paper, we have explored several implications that could occur based on the assumption that we could currently be living inside of a simulation. Having explored all these arguments, I contend that we are not living inside of a simulation because we do not currently have the technology to do so, as the first step to the creation of the simulation would be a

posthuman society. However, after reviewing how far the research into artificial intelligence has become, it is very likely that in the near future, we all will be living inside of a simulation.

Bibliography

1. Bostrom, N. (2003). Are We Living in a Computer Simulation?. *The Philosophical Quarterly*, *53*(211). 1-15 Nick Bostrom "Are We Living Inside a Simulation" pg
2. Agatonović, M. (2023) The Fiction of Simulation: A Critique of Bostrom's Simulation Argument. *AI & Soc* **38**, 1581
3. Zagal, J. P., & Deterding, S. (2018). Definitions of "Role-Playing games". *Role-playing Game Studies: Transmedia Foundations*, 4
4. Blitz, Mark. "Understanding Heidegger on technology." *The New Atlantis* (2014): 63-80.
5. Steinhart, E. (2010). Theological Implications of the Simulation Argument. *Ars Disputandi*, *10*(1), 20-32
6. Parfit, Derek. 1986. *Reasons and Person*. Oxford University Press. 200, 332
7. Dennett, Daniel. 1981. "Where Am I?" Lehigh University. 1981. 6- 8 https://www.lehigh.edu/~mhb0/Dennett-WhereAmI.pdf.
8. Peoples, G. Andrew. "Death, Resurrection and the Continuity of Personal Identity." 6-7
9. Grube, G.M.A. (1997). *The Republic*. Diamond Pocket Books Pvt Ltd. 514b-514c
10. Berkeley, G. (1878a). *A Treatise Concerning the Principles of Human Knowledge*. 11-12
11. Putnam, Hilary. *Brains in a Vat*. 1981. 7-
12. Arvan, M. (2015). The Peer-to-Peer Simulation Hypothesis and a New Theory of Free Will. 7
13. Milton, J. (1890). *Areopagitica*. 28
14. Kant, I. (1999). *Critique of Pure Reason*. Cambridge University Press. 349

15. Schopenhauer, Arthur, and Konstantin Kolenda. *Essay on the Freedom of the Will*. Courier Corporation, 2005.

16. Moreno, Agustin. "Schopenhauer On Free Will and Fatalism." *Philosophy Pathways* 230, no. 1 (2019).

17. Janaway, Christopher. "Necessity, responsibility and character: Schopenhauer on freedom of the will." *Kantian Review* 17, no. 3 (2012): 431-457.

18. Janaway, Christopher, 'Determinism and Responsibility', *Self and World in Schopenhauer's Philosophy* (Oxford, 1999; online edn, Oxford Academic, 1 Nov. 2003),

19. Kant, Immanuel. 1999a. *Critique of Pure Reason*. Cambridge University Press.

20. Crummett, D. (2021). The Real Advantages of the Simulation Solution to the Problem of Natural evil. *Religious Studies*, *57*(4), 620-621

21. Dainton, Barry. "Natural evil: the simulation solution." *Religious studies* 56, no. 2 (2020): 209-230.

22. Richard, Frédéric. "Karmic Retribution in Theravāda Buddhism: A way to salvation?." In *Retribution in Jewish and Christian Writings: A Concept in Debate*, pp. 149-171. Mohr Siebeck, 2019.

23. Nath, Rajakishore, and Riya Manna. "From posthumanism to ethics of artificial intelligence." *AI & SOCIETY* 38, no. 1 (2023): 185-196.

24. Van der Laan, J. M. "Machines and human beings in the movies." *Bulletin of Science, Technology & Society* 26, no. 1 (2006): 31-37.

25. Retto, Jesús. "Sophia, first citizen robot of the world." *ResearchGate, URL: https://www. researchgate. net* (2017).

26. Samiei, Saba. "On The Danger of Artificial Intelligence." PhD diss., Auckland University of Technology, 2019.
27. Livingston, Steven, and Mathias Risse. "The future impact of artificial intelligence on humans and human rights." *Ethics & international affairs* 33, no. 2 (2019): 141-158.
28. Müller, Vincent C. and Bostrom, Nick (2016), ‘Future progress in artificial intelligence: A survey of expert opinion’, in Vincent C. Müller (ed.), Fundamental Issues of ArtificialIntelligence (Synthese Library; Berlin: Springer), 553-571.
29. Fadziso, Takudzwa. "Why Neuralink will Change Humanity Forever?." *Asian Journal of Humanity, Art and Literature* 7, no. 1 (2020): 69-78.
30. Ulgen, Ozlem. "Kantian ethics in the age of artificial intelligence and robotics." *QIL* 43 (2017): 59-83.