Mind, Reality, and Perception in Science Fiction

Bičvić, Antonija

Sveučilište J. J. Strossmayera u Osijeku

Filozofski fakultet Osijek

Sveučilišni diplomski dvopredmetni studij engleskog jezika i književnosti i filozofije – nastavnički smjer

_____

Antonija Bičvić

Um, stvarnost i percepcija u znanstvenoj fantastici

Diplomski rad

Mentor: doc. dr. sc. Jasna Poljak Rehlicki

Osijek, 2024.

Sveučilište J. J. Strossmayera u Osijeku

Filozofski fakultet Osijek

Odsjek za engleski jezik i književnost

Sveučilišni diplomski dvopredmetni studij engleskog jezika i književnosti i filozofije – nastavnički smjer

_____

Antonija Bičvić

Um, stvarnost i percepcija u znanstvenoj fantastici

Diplomski rad

Znanstveno područje: humanističke znanosti

Znanstveno polje: filologija

Znanstvena grana: anglistika

Mentor: doc. dr. sc. Jasna Poljak Rehlicki

Osijek, 2024.

J.J. Strossmayer University of Osijek

Faculty of Humanities and Social Sciences

Study Programme: Double Major MA Study Programme in English Language and Literature and Philosophy – Teaching English as a Foreign Language and Philosophy

---

Antonija Bičvić

Mind, Reality, and Perception in Science Fiction

Master's Thesis

Supervisor: Dr. Jasna Poljak Rehlicki, Assistant Professor

Osijek, 2024

J.J. Strossmayer University of Osijek

Faculty of Humanities and Social Sciences

Department of English

Study Programme: Double Major MA Study Programme in English Language and Literature and Philosophy – Teaching English as a Foreign Language and Philosophy

---

Antonija Bičvić

Mind, Reality, and Perception in Science Fiction

Master's Thesis

Scientific area: humanities

Scientific field: philology

Scientific branch: English studies

Supervisor: Dr. Jasna Poljak Rehlicki, Assistant Professor

Osijek, 2024

IZJAVA

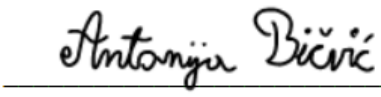
Izjavljujem s punom materijalnom i moralnom odgovornošću da sam ovaj rad samostalno napisao/napisala te da u njemu nema kopiranih ili prepisanih dijelova teksta tuđih radova, a da nisu označeni kao citati s navođenjem izvora odakle su preneseni.

Svojim vlastoručnim potpisom potvrđujem da sam suglasan/suglasna da Filozofski fakultet u Osijeku trajno pohrani i javno objavi ovaj moj rad u internetskoj bazi završnih i diplomskih radova knjižnice Filozofskog fakulteta u Osijeku, knjižnice Sveučilišta Josipa Jurja Strossmayera u Osijeku i Nacionalne i sveučilišne knjižnice u Zagrebu.

U Osijeku ____8.10.2024._____

Ime i prezime studenta, JMBAG

_Antonija Bičvić, _0122233784_

Potpis 

Abstract

The American philosopher Hilary Putnam proposed a theory in his work *Reason, Truth and History* called "Brains in a Vat" and it motivates the reader to question their knowledge of the world as well as their state of existence. Putnam's theory, claiming that a brain connected to a computer might perceive the same reality as the one within a body without realising the difference, raises questions about human perceptions of the world, as well as the validity of those perceptions. This thesis will utilize Putnam's theory as a stepping stone toward exploring the interpretations of the notions of mind, reality, and perception within various works of the science fiction genre. This paper aims to analyse Putnam's theory, Bostrom's theory of simulation, Searle's Chinese room, Turing's test for artificial intelligence, as well as well-known cinematic works such as the *Matrix*, *Transcendence, RoboCop, Ghost in the Shell*, and the beloved show *Doctor Who* to demonstrate the various portrayals of the human mind, reality and the way we perceive it.

Keywords: Putnam, Searle, Bostrom, Turing*, Matrix, Inception, RoboCop, Ghost in the Shell, Doctor Who,* mind, reality, perception, sci-fi, science fiction, AI, artificial intelligence, simulation

Table of Contents

Introduction

In today's modern world technology has become an unavoidable part of everyday life. Aside from social media and the constant need for connection through texting and posting another element has recently risen to the forefront: artificial intelligence. Recently, artificial intelligence has become a frequent topic of technological, philosophical, and social discourse. The entertainment genre of science fiction has artificial intelligence starring as the central theme of many shows and movies. Alongside it, there are often concepts of mind, reality, and perception that the authors like to play in. This work will analyse those concepts to scrutinize the role of the human mind, our perception, and our reality itself in those worlds of simulations, machines, and artificial intelligence.

The first chapter will analyse a text from Putnam's book *Reason, Truth and History* to portray his thought experiment titled "Brains in a Vat". The second chapter will evaluate Nick Bostrom's article titled "Are You Living in a Computer Simulation?" to interpret Bostrom's reasoning for his *Simulation hypothesis*. The third chapter will analyse the definition of artificial intelligence and describe Searle's theory of two types of artificial intelligence, and his *Chinese room* theory from his article "Minds, Brains and Programs", as well as illustrate Turing's test for determining artificial intelligence. The fourth chapter will describe Mori's "Uncanny Valley" phenomenon. The fifth chapter will dissect the movie *Matrix* to portray the way a human mind interacts with the artificial intelligence. The sixth chapter will evaluate the movie *Transcendence* to determine the differences between a human mind and artificial intelligence. The seventh chapter will interpret the movie *RoboCop* to express the resilience of the human mind despite the drastic changes in its reality. The eight chapter will interpret the movie *Ghost in the Shell* in order to define how a human mind is able to overcome altered memories and perception. The ninth chapter will outline two episodes from the show *Doctor Who* titled "The Rebel Flesh" and "The Almost People" to illustrate how humans deal with a reality in which their minds and bodies have been perfectly duplicated.

1. Hilary Putnam: Brains in a Vat

One of the leading American philosophers, Hilary Putnam, throughout his career made considerable contributions to metaphysics, epistemology, the philosophy of mind, language, science, mathematics, and the philosophy of logic ("Hilary Putnam"). Among his considerable works is his book *Reason, Truth and History* published in 1981, within which he writes about a thought experiment called "Brains in a Vat", that very experiment has no doubt inspired many in the realm of science fiction as its premise lies somewhere between fascinating and mildly horrifying.

1.1. The Philosophical View

In his review, Ellet states that in the book *Reason, Truth and History*, Putnam attacks scientific, or metaphysical, realism, a view that he once defended, by favouring a different view, that of internal realism (Ellet 95). He claims that "… scientific realism holds that science aims to give us in its theories, a literally true story of what particular kinds of entities and what kinds of processes really exist…" (97), in other words, it aims to give us the true story of how the world is. But Putnam argues that one must take into consideration the perspective of an individual:

> Putnam began his criticism of scientific realism with an analogy to mathematics. If we let 'the world' be a straight line, then, according to story A, there are points, but according to story B, there are no points, only the line and its parts which all have extension. Putnam argued that story (or interpretation) A and story (or interpretation) B are logically equivalent descriptions of 'the world.' Thus, the "real object" that is labelled 'point' in story A might be labelled 'set of convergent line segments ' in story B. It is, of course, a property of the world itself that it admits of these different interpretations. (98)

There is a different side to every story depending on the position from which we view the world and thus he argues for internal realism, as the perspective becomes an important new factor in the image of the world. This analogy calls to mind an image of the number nine, or number six if viewed upside-down, with two people each standing on one "side" of the number. While person A standing on one side of this number may see a six, person B standing on the opposite side will then view a nine. Neither of these people are wrong, despite their differing answers they are both right. Thus, the number in question is both a six and a nine, suggesting

that the truth is not singular, and that the reality allows for more than one version of the "story" which depends on the perspective of an individual.

This matter of perspective is further analysed within the book in an experiment of thought called "Brains in a Vat". The idea behind this experiment is to disprove the belief that there is one truth of the world: "… it depicts a situation where all our beliefs about the world would presumably be false, even though they are well justified. Thus, if one can prove that we cannot be brains in a vat … one can prove that metaphysical realism is false" (Hickey). While this experiment is used in philosophy to illustrate global skepticism this paper will analyse it in a more creative tone, mainly in the light of various possibilities it represents for science fiction media, as many well-known movies such as *The Matrix*, *RoboCop, Ghost in the Shell* and *Transcendence* have touched upon similar topics.

1.2. Mind in a Machine

In the "Brains in a Vat" experiment Putnam suggests a scenario within which the reader imagines themselves completely stripped of their body, reduced to their brain and nervous system, and yet unaware of this most grievous of violations:

> Here is a science fiction possibility discussed by philosophers: imagine that a human being (you can imagine this to be yourself) has been subjected to an operation by an evil scientist. The person's brain (your brain) has been removed from the body and placed in a vat of nutrients which keeps the brain alive. The nerve endings have been connected to a super-scientific computer which causes the person whose brain it is to have the illusion that everything is perfectly normal. (Putnam 5-6)

The above-mentioned super-scientific computer that is connected to the nerves would then act as that person's senses, it would provide sensory input such as touch, sight and hearing while simultaneously making that person believe they are creating output such as movement or speech:

> There seem to be people, objects, the sky, etc; but really all the person (you) is experiencing is the result of electronic impulses travelling from the computer to the nerve endings. The computer is so clever that if the person tries to raise his

> hand, the feedback from the computer will cause him to 'see' and 'feel' the hand being raised. (6)

Such a person would be completely unaware that they are just a brain-and-nerves blob floating in a tank of liquid and no doubt resembling a disturbing jellyfish more than anything humanoid. The horrifying reality of their existence would be lost to them. If their entire reality depends on the machinations and manipulations of a clever machine, without any senses of their own, they have no way of perceiving the real world over the "reality" being fed to them. One must assume that all semblance of autonomy and freedom has been thoroughly denied to such entity. Thankfully, Putnam claims that we are not mere brains connected to the computer by concluding that, if that were the case, the brain connected to the computer would be unable to fathom such a scenario. Thus, by simply imagining it, we are negating the possibility of its existence:

> The answer is going to be (basically) this: although the people in that possible world can think and 'say' any words we can think and say, they cannot (I claim) refer to what we can refer to. In particular, they cannot think or say that they are brains in a vat (even by thinking 'we are brains in a vat'). (8)

But the very possibility of this scenario, once thought of, engrains itself in one's mind. The grotesque nature of it would not permit otherwise. With their mind thoroughly ensnared, their reality an utter fabrication and their perception altered one must ask how much is such a life worth. On the one hand, the person has the potential to live out, as far as they know, a full life from cradle to the grave filled with friendships, love struggles, awkward job interviews and boring school lectures. On the other hand, there is the question of the validity of each of these aforementioned experiences. If the brain in the vat connected to the computer is the only one, aside from the scientist, to witness the false sensations being stimulated to it by the machine and its own choices made based on that false information, then an argument can be made that such life has no value because the only one affected by it is the individual brain in the vat, while people in the real world tend to affect everyone around them with every choice they make.

Such a train of thought leads to existential questions which often play a central role in the media, especially in the science fiction genre. The idea of a mind in a machine has made a repeated appearance over the years, garnished with a futuristic aesthetic and often horrifying plot points, Putnam's theory has breathed essence into works such as *The Matrix*,

Transcendence, *RoboCop*, *Ghost in the Shell*, and *Doctor Who*, all of which will be analysed within this paper.

2. Nick Bostrom: Simulation Theory

Unlike Putnam's theory which eventually concludes that a mind in a virtual reality is impossible, Nick Bostrom's work "Are You Living in a Computer Simulation?" actually favours that possibility. Bostrom presents a scenario in which an advanced civilization will have enough knowledge and resources to build computers capable of supporting a simulation. In fact, Bostrom proposes that, if such a civilization is indeed capable of creating a simulation, they would not stop at just one.

In this future scenario there would be many simulations, and only one reality, so the chances are greater for an individual to find themselves in one of those numerous simulations rather than in the "real" world:

> Because their computers would be so powerful, they could run a great many such simulations … Then it could be the case that the vast majority of minds like ours do not belong to the original race but rather to people simulated by the advanced descendants of an original race. It is then possible to argue that, if this were the case, we would be rational to think that we are likely among the simulated minds rather than among the original biological ones. (Bostrom 1)

Bostrom goes on to explain a common assumption in philosophy called *substrate-independence*, an idea that "… mental states can supervene on any of a broad class of physical substrates. Provided a system implements the right sort of computational structures and processes, it can be associated with conscious experiences" (2). Bostrom claims that it is not "… an essential property of consciousness that it is implemented on carbon-based biological neural networks inside a cranium: silicon-based processors inside a computer could in principle do the trick as well" (2). In other words, to have something sentient and self-aware as software, its hardware does not necessarily have to be a flesh-and-bone person. An artificial hardware of silicone and metal, like that of a computer, will equally suffice. The only pre-requisite is its ability to house that consciousness.

Bostrom admits that, at this point in human history, we are not advanced enough to produce computers or programmes capable of supporting even one simulation, let alone

multiple ones at the same time: "At our current stage of technological development, we have neither sufficiently powerful hardware nor the requisite software to create conscious minds in computers" (3). But considering the amount of thought already put into such possibilities as well as the current technological advancements Bostrom believes that such an outcome is inevitable, perhaps even within the next few decades.

Furthermore, Bostrom theorizes that this *posthuman* civilization might even create giant computers as they "… convert planets and other astronomical resources into enormously powerful computers…" (3) and he assumes this would be possible because they might discover a: "… novel physical phenomena, not allowed for in current physical theories, [that] may be utilized to transcend those constraints that in our current understanding impose theoretical limits on the information processing attainable in a given lump of matter" (3-4).

In other words, there are still many secrets of the universe to unlock and there may be laws of physics yet to be discovered that would help translate these ideas into a reality. Specifically, the idea of a super-computer capable of housing simulations together with their collection of artificial consciousnesses.

Bostrom then proposes three possible versions of this future. The first option is that: "…humankind will almost certainly fail to reach a posthuman level; for virtually no species at our level of development become posthuman…" (9). In this version of the future we are not necessarily immediately extinct, we either remain as we are, or a bit more technologically advanced, for a long time and *then* go extinct, or something happens to cause a collapse of our society before we reach the posthuman level and we continue to exist in a form of primitive societies. Either way, humans do not get to reach the posthuman civilization and realize the simulations.

The second alternative is "that the fraction of posthuman civilizations that are interested in running ancestor- simulation is negligibly small" (10). Either the individuals have a lack of interest and resources for such an effort, or they have "reliably enforced laws that prevent such individuals from acting on their desires" (11). Perhaps the development of these societies has converged so much that none of them hold any interest in ancestor-simulations, or perhaps we will eventually evolve to not have such preposterous notions. The third option is the most interesting one:

> If we are living in a simulation, then the cosmos that we are observing is just a tiny piece of the totality of physical existence. The physics in the universe where

> the computer is situated that is running the simulation may or may not resemble the physics of the world that we observe. While the world we see is in some sense "real", it is not located at the fundamental level of reality. (11)

Bostrom also proposes that in this third alternative future, the simulated societies might create simulations of their own, meaning that they can reach the posthuman level as well. He concludes that:

> If we do go on to create our own ancestor-simulations, this would be strong evidence against (1) and (2), and we would therefore have to conclude that we live in a simulation. Moreover, we would have to suspect that the posthumans running our simulation are themselves simulated beings; and their creators, in turn, may also be simulated beings. (12)

This last scenario can easily be compared with myths of creation from various religions because "… the posthumans running a simulation are like gods in relation to the people inhabiting the simulation" (12). The ones who created the simulation would be like gods because they are: "…of superior intelligence; they are "omnipotent" in the sense that they can interfere in the workings of our world even in ways that violate its physical laws; and they are "omniscient" in the sense that they can monitor everything that happens" (12).

Aside from the ancestor simulation, Bostrom also presents the idea of simulations meant for selected individuals or even a single person, in that case: "The rest of humanity would then be zombies or "shadow-people" – humans simulated only at a level sufficient for the fully simulated people not to notice anything suspicious" (13). He does admit that it is unclear whether these shadow people would be cheaper to create, or even if it is possible to create something passable as a human without an actual consciousness.

To conclude, according to Bostrom's theory, in the first case, humans never reach the posthuman level, in the second case human societies have converged so much that the decision not to run these simulations is unanimous, and the third and most likely option is that the simulation is indeed being ran. And if there are multiple simulations then any person might be in one. If there was a single blue marble in a box filled with red marbles and a person reached into that box without looking there are greater chances to pick up a red marble rather than a blue one. Similarly, if there were multiple simulations and only one reality, chances are they exist in one of the simulations, and Bostrom claims that: "Unless we are now living in a simulation, our descendants will almost certainly never run an ancestor-simulation" (14).

3. Artificial Intelligence

Artificial intelligence is often defined as "…the ability of a digital computer or computer-controlled robot to perform tasks commonly associated with intelligent beings" ("artificial intelligence"). It is often at the centre of the science fiction genre and is something that humanity has strived to achieve for more than a decade. Also, its existence has been theorised for even longer than that. Alan Turing was one of the first to mention computers with such abilities in his work during the 1930s, and later on he even wrote about that concept:

> Turing gave quite possibly the earliest public lecture (London, 1947) to mention computer intelligence, saying, "What we want is a machine that can learn from experience," and that the "possibility of letting the machine alter its own instructions provides the mechanism for this." In 1948 he introduced many of the central concepts of AI in a report entitled "Intelligent Machinery." However, Turing did not publish this paper, and many of his ideas were later reinvented by others. ("history of artificial intelligence (AI)")

Almost every part of human behaviour is ascribed to intelligence, and intelligence is what draws a line between humans and other animals whose behaviour can always be attributed to instinct. When it comes to artificial intelligence there are several things to look for to indicate intelligence: "learning, reasoning, problem solving, perception, and using language" ("artificial intelligence"). When it comes to human intelligence, psychologists always describe it as a cluster of traits rather than a single mark, therefore, when speaking of artificial intelligence, it makes sense to model certain things after humans, as humans are the only ones so far, the trait of intelligence has been ascribed to.

3.1. John Searle: Minds, Brains and Programs

John Searle writes in his work "Minds, Brains and Programs", an article published in *The Behavioral and Brain Sciences* in 1980, about two different types of artificial intelligence. The first type he classifies as the "strong" AI form, and the second as the "weak" or "cautious" AI form (Searle 417).

The weak AI is considered a tool rather than something truly intelligent. It is more like an advanced programme made to fulfil a purpose: "According to weak AI, the principal value of the computer in the study of the mind is that it gives US a very powerful tool. For example,

it enables us to formulate and test hypotheses in a more rigorous and precise fashion" (417). On the other hand, the strong AI is more than just a programme; closer to the idea often seen in various movies:

> But according to strong AI, the computer is not merely a tool in the study of the mind; rather, the appropriately programmed computer really is a mind, in the sense that computers given the right programs can be literally said to understand and have other cognitive states. In strong AI, because the programmed computer has cognitive states, the programs are not mere tools that enable us to test psychological explanations; rather, the programs are themselves the explanations. (417)

Searle then continues by describing the work of Roger Shank and his colleagues at Yale who created a program:

> Very briefly, and leaving out the various details, one can describe Schank's program as follows: the aim of the program is to simulate the human ability to understand stories. It is characteristic of human beings' story-understanding capacity that they can answer questions about the story even though the information that they give was never explicitly stated in the story. (417)

In other words, a sign of strong artificial intelligence would be to extract answers from the context of the story rather than the information given to it outright. He gives an example of two very similar scenarios:

> A man went into a restaurant and ordered a hamburger. When the hamburger arrived it was burned to a crisp, and the man stormed out of the restaurant angrily, without paying for the hamburger or leaving a tip." Now, if you are asked "Did the man eat the hamburger?" you will presumably answer, "No, he did not." Similarly, if you are given the following story: "A man went into a restaurant and ordered a hamburger; when the hamburger came he was very pleased with it; and as he left the restaurant he gave the waitress a large tip before paying his bill," and you are asked the question, "Did the man eat the hamburger?" you will presumably answer, "Yes, he ate the hamburger." (417)

Neither of the two stories states whether or not the man ate the hamburger, but a human being would be able to extrapolate the correct answer regardless, by leaning on context and

previous knowledge of human behaviour and its patterns, as well as their personal experience. An AI has no personal experience, and yet if able to give the same answer as the human in the story above then it can be considered a “strong” AI. A mind.

3.2. The Chinese Room

Searle was not satisfied with the above-mentioned definition of an AI and thus he devised an experiment of his own called “The Chinese Room.” In this scenario, Searle presents the following:

> Suppose that I'm locked in a room and given a large batch of Chinese writing. Suppose furthermore (as is indeed the case) that I know no Chinese, either written or spoken, and that I’m not even confident that I could recognize Chinese writing as Chinese writing distinct from, say, Japanese writing or meaningless squiggles. To me, Chinese writing is just so many meaningless squiggles. Now suppose further that after this first batch of Chinese writing I am given a second batch of Chinese script together with a set of rules for correlating the second batch with the first batch. The rules are in English, and I understand these rules as well as any other native speaker of English. They enable me to correlate one set of formal symbols with another set of formal symbols, and all that “formal” means here is that I can identify the symbols entirely by their shapes. Now suppose also that I am given a third batch of Chinese symbols together with some instructions, again in English, that enable me to correlate elements of this third batch with the first two batches, and these rules instruct me how to give back certain Chinese symbols with certain sorts of shapes in response to certain sorts of shapes given me in the third batch. Unknown to me, the people who are giving me all of these symbols call the first batch “a script,” they call the second batch a “story,” and they call the third batch “questions.” Furthermore, they call the symbols I give them back in response to the third batch “answers to the questions,” and the set of rules in English that they gave me, they call "the program." (418)

Searle goes on to explain that, in this scenario, he is given stories in English, which he responds to in English, and stories in Chinese which he responds to in Chinese with the help of the previously mentioned three sets of symbols. The people outside of his room have no idea that, while he understands English perfectly well, he does not understand a single thing in Chinese. And yet, because of the instructions on how to match the symbols, his answers in

English are just as eloquent as those in Chinese, so much so that the people outside of the room do not notice a difference.

The point of this is that, just because Searle can replicate the same answers that a Chinese-speaking individual would give, it does not mean that he understands a single word of the stories he was given, but the people outside the room do not know that. This leads us to conclude that, just because a "strong" AI can mimic the responses of a person, it does not mean that it understands the story like a person would, it does not make it a person, and it does not make it a *mind* because it lacks a person's ability to *perceive reality.*

3.3. Turing's Test

The "Turing test" is a test for artificial intelligence proposed by the mathematician Alan M. Turing in 1950, and the test is meant to determine whether or not a computer can "think" ("Turing test"). It is quite difficult, as Searle's "Chinese room" has demonstrated, to make a distinction between individual thought and simple repetition, and the point of a Turing test is to make such a distinction.

Turing has managed to avoid overly complicating the matter by devising this test: "… if a computer acts, reacts, and interacts like a sentient being, then call it sentient" ("Turing test"). And, while the test may be subjective, it is practical and it avoids the attempt to define what a *mind* or *intelligence* are; it is simply concerned with the fact that they *are*.

This can easily be tested by having an individual chat with the AI, and in this case the AI is capable of responding as a human, even lying to answer negatively to a question "Are you an AI?", individuals have to guess from the answers whether they are chatting with a machine or another person. If the AI manages to produce answers in such a way that the person they are chatting with does not notice the difference, and is sure they are chatting with another person, then, according to the Turing test, that AI is sentient.

4. Uncanny Valley

Japanese roboticist Masahiro Mori proposed a phenomenon called "Uncanny Valley" in 1970. The hypothesis suggests that:

> …as human likeness increases in an object's design, so does one's affinity for the object—but only to a certain point. When the likeness nears total accuracy, affinity drops dramatically and is replaced by a feeling of eeriness or uncanniness. Affinity then rises again when true human likeness—indicating a living person—is reached. This sudden decrease and increase caused by the feeling of uncanniness creates a "valley" in the level of affinity. This proposed phenomenon is expressed most often as a line graph, with "human likeness" on the x-axis and "affinity" on the y-axis. The valley occurs at the line's abrupt plunge and subsequent ascent. ("uncanny valley")

In other words, things appear cuter and prettier to humans the more they look like humans, that is, until they reach that critical point of both "too much" and "not enough", where they almost look like humans but there are still some perceived discrepancies. That is the point in the graph where the affinity takes a sharp plunge downwards, and then rises again once it reaches the exact likeness to a human being. That part of the graph where the curve plunges below the x-axis, showing a complete lack of affinity, is where negative feelings of revulsion and wariness, even fear, occur. That valley is called "Uncanny Valley" due to the uncanny and disturbing feeling that it evokes.

Mori also warns that appearance is not the only factor that triggers the "Uncanny Valley". One must also consider the movement of a robot or a prosthetic hand: "Since negative effects of movement are apparent even with a prosthetic hand, a whole robot would magnify the creepiness" (Mori 4).

Studies regarding the cause of the "Uncanny Valley" feeling have, thus far, been inconclusive:

> One study found that what most unnerved participants was the illusion of human consciousness that near-human likeness causes—the prospect that a robot could think and feel as humans do. Another theory credits primal instinct. Humans are programmed by evolution to favour mates that appear strong and healthy, and a humanoid robot's unnatural movement may signal disease and danger on a

> subconscious level. Yet another idea suggests that it is the ambiguity between human and inhuman that is most disturbing. ("uncanny valley")

Whatever the cause may be, the very idea of this phenomenon brings interesting new insights into the human mind. Even if neither its cause nor the hypothesis has been proven, it is still interesting to ponder the idea of such a response ("uncanny valley"). The "Uncanny Valley" phenomenon proposes that all humans, no matter their gender, ethnicity or faith, seem to have an aversion to things that look *almost* like us. It is a primal instinct that instils fear of otherness and wrongness towards things that look and move just a bit off.

Outside of its practical use in the field of robotics and prosthetics, Mori's hypothesis has endless potential in the entertainment genres of science fiction, fantasy, and most likely (and inevitably) horror.

5. The Matrix

The Matrix movie was written and directed by Andy and Larry Wachowski in 1999, and "The film blends the old mythology of the coming of a messiah with the new mythology of virtual reality to create a new kind of religious hero" ("The Matrix"). The main character is Neo, formerly known as Thomas A. Anderson, who works as a programmer in a software company by day and is a hacker by night. "Neo" is a codename that he uses, but early on in the movie he switches to it as his main name, almost like he transitions from his normal life into this alternative self that better reflects who he is as a person. The premise of the movie is similar to Putnam's thought experiment "Brains in a Vat," except that the minds in *The Matrix*, while connected to a super-computer that is showing them a fake reality, are also still inside an intact body that is in a pod (vat). The body is augmented with implants to allow a smoother connection to the machine.

The backstory of *The Matrix* world is slowly revealed to us through Neo's point of view as he first learns that the world he lives in is fake, and then it is revealed to him that the real world was ravaged by the war between humans and AI. These revelations are presented to him by Morpheus, named after the Greek god of dreams, which is quite fitting considering the parallel between dreaming and being in the Matrix ("Morpheus"). Just like it happens with lucid dreaming, once people realize that the Matrix is not real they can learn how to bend its rules ("lucid dreaming"). It is implied by Trinity, one of the people Morpheus works with, that Neo is one of the hackers who were looking for Morpheus, meaning that there are potentially more people within the Matrix slowly becoming aware of the nature of their world:

> "Please just listen. I know why you're here, Neo. I know what you've been doing. I know why you hardly sleep, why you live alone, and why night after night you sit at your computer. You're looking for him. I know, because I was once looking for the same thing. And when he found me, he told me I wasn't really looking for him. I was looking for an answer. It's the question that drives us, Neo. It's the question that brought you here. You know the question just as I did." (*The Matrix*, 00:10:54-00:11:40)

It is unclear whether Neo came across some data while hacking or if he simply perceived that there was something wrong, perhaps receiving some sort of signals through shared

consciousness within the Matrix. Either way, this implies that certain human minds cannot be blinded by the false reality, as they manage to perceive the little discrepancies.

Within the Matrix there are defense systems, programs called "Agents" who look like every on-screen portrayal of an FBI/Secret Service agent, together with dark suits, glasses, and earpieces. Despite the cliché, there is nothing humorous about their ability to move through the Matrix with superiority compared to normal human minds trapped inside it, seemingly breaking the non-existent laws of physics, or their ability to simply "hack" anyone still "plugged into" the Matrix by simply taking over their "Matrix body: "Sentient programs. They can move in and out of any software still hard-wired to their system. That means that anyone we haven't unplugged is potentially an agent. Inside the Matrix, they are everyone and they are no one" (00:57:37-00:57:56). These agents arrest Neo soon after he contacts Trinity and Morpheus. During the interrogation scene they put a "bug" inside Neo, a tracking programme that crawls inside his "body."

This is an interesting portrayal of programmes in such a way that a human mind can perceive it, considering that for human minds it would be easier to digest an actual image of a bug than the multiple lines of code that the program is, undoubtedly made of. This raises the question of how much control the machine has over the humans' perception, since it would be simpler to hide the bug altogether in the background as simple lines of code, but perhaps the human mind that is used to perceiving the Matrix as a "reality" would completely reject such a form. Also, another question arises; whether the image of a bug was a humorous attempt from a humourless machine or is merely a nature of the human mind, that changes data to accept it, almost like a defense mechanism, or a trauma response:

> Defense mechanism, in psychoanalytic theory, any of a group of mental processes that enables the mind to reach compromise solutions to conflicts that it is unable to resolve. The process is usually unconscious, and the compromise generally involves concealing from oneself internal drives or feelings that threaten to lower self-esteem or provoke anxiety. ("defense mechanism")

Later it becomes clear that a human mind can be altered within the Matrix. After they de-bug him, Trinity and her team take Neo to see Morpheus, where he is offered a choice between the blue pill, which will presumably wipe away his memory of Morpheus and the Matrix, and the red pill, which will open his eyes to the truth:

> "That you are a slave, Neo. Like everyone else you were born into bondage, born into a prison that you cannot smell or taste or touch. A prison for your mind…. Unfortunately, no one can be told what the Matrix is. You have to see it for yourself. This is your last chance. After this there is no turning back. You take the blue pill, the story ends, you wake up in your bed and believe whatever you want to believe. You take the red pill, you stay in Wonderland, and I show you how deep the rabbit hole goes… Remember, all I'm offering is the truth, nothing more…" (*The Matrix,* 28:20-29:34)

This iconic scene of choosing between pills has since become a symbol of awakening and accepting the truth and reality ("red pill and blue pill"). Nothing within the Matrix is physical, so the pills are programmes, and the equivalent of taking one is most likely akin to assimilating the program into your code. While the blue pill erases what must be files of memories, the red one prepares Neo's mind to become "unplugged:" "The pill you took is part of a trace program. It's designed to disrupt your input/output carrier signals so we can pinpoint your location" (*The Matrix*, 00:30:25-00:30:34).

After taking the red pill Neo wakes up inside a pod, hooked up to parts of a machinery with a feeding/breathing tube down his throat and a cable plugged directly into his brain. Around him are similar pods each containing a person, and all of the pods are connected to a mechanical tower. This scene is rather horrifying as the truth is gradually revealed to both Neo and the viewers. The next few scenes show him being found by Morpheus and his crew in a futuristic-looking flying ship and he physically recovers, as a machine restores his atrophied muscles. The Matrix is set to look like the year 1999, but from the advanced technology in the real world it soon becomes clear that they are in the future. Morpheus admits that they are not sure what year it is: "You believe it's the year 1999 when in fact it's closer to 2199. I can't tell you exactly what year it is because we honestly don't know" (00:37:35-00:37:43).

Morpheus and his crew slowly get Neo accustomed to the real world, then they "plug him in" to a training programme, where Morpheus explains why humans are in the Matrix, and the viewers, together with Neo, get some answers about the background of *The Matrix* world:

> "A singular consciousness that spawned an entire race of machines. We don't know who struck first, us or them. But we know that it was us that scorched the sky. At the time they were dependent on solar power and it was believed that they would be unable to survive without an energy source as abundant as the

> sun. Throughout human history, we have been dependent on machines to survive. Fate, it seems, is not without a sense of irony. The human body generates more bio-electricity than a 120-volt battery and over 25,000 BTUs of body heat. Combined with a form of fusion, the machines have found all the energy they would ever need. There are fields, endless fields, where human beings are no longer born. We are grown. For the longest time I wouldn't believe it, and then I saw the fields with my own eyes. Watch them liquefy the dead so they could be fed intravenously to the living. And standing there, facing the pure horrifying precision, I came to realize the obviousness of the truth. What is the Matrix? Control. The Matrix is a computer generated dream world built to keep us under control in order to change a human being into this." (41:40-43:38)

Later on in the movie it is mentioned that the survivors from the matrix have a city called Zion: "If the war was over tomorrow, Zion's where the party would be" (00:47:13-00:47:16), which is portrayed quite apocalyptically.

Humans' motivations are understandable, they were losing a war, cornered and desperate they were a wounded animal at their most dangerous. That is something that the viewers can grasp. But the truly shocking factor is what the machine did afterward, incapable of emotions, without viciousness, malice, or ideas of revenge, its actions are all the more sinister because they can be summed up as a simple calculation: they needed a new energy source, human bodies produce electricity. It was a situation in which they both obtained their energy source and neutralized the threat to their survival. Nothing can be more demeaning and humiliating than the fact that, while the enemy has completely dehumanized them, and has turned them into a mere battery, they have done so with no more motivation than the simple practicality of the matter. When one hates, one has been marked by the hated, but in this case humanity had no influence, no control over the machine's decision making process, nothing other than cold and clinical convenience. Such a reality is hard for a human mind to perceive; Morpheus admits as much himself when he apologises to Neo: "I feel I owe you an apology. We have a rule. We never free a mind once it's reached a certain age. It's dangerous, the mind has trouble letting go. I've seen it before and I'm sorry" (00:44:48-00:45:00).

An example of someone being unable to let go is Cypher, a member of Morpheus' team who goes behind their back and enters a Matrix to meet with an Agent. There is a scene in a restaurant where he negotiates with an Agent called Smith about returning to the Matrix in exchange for information on Morpheus and his crew: "You know; I know this steak doesn't

exist. I know that when I put it in my mouth, the Matrix is telling my brain that it is juicy and delicious. After nine years, you know what I realize? Ignorance is bliss" (1:03:47-1:04:18).

This "ignorance is bliss" statement shows a lack of honour, he is a coward and a traitor, and yet, the viewers might find themselves sympathising with him even against their will. People often find solace in their dreams to escape reality, one cannot be too sure that they would not have acted as Cypher has here, had they been confronted with such reality: "I don't want to remember nothing. Nothing. You understand? And I want to be rich. You know, someone important. Like an actor" (1:04:27-1:04:45).

For Cypher, the real world does not matter. It is horrible and too much for him to endure any longer, and the Matrix might be a lie, but his perception will be sufficiently deceived, once his memories are deleted, and his mind will not know the difference. Humans like to hold themselves above the animal world, above instincts, above the physical, they often do it so much that a mind is all they reduce themselves to. In such a train of thought Cypher's actions may be completely justified, if the mind is all that matters, then there is no point in making it suffer the waking world.

Just like the argument with the "Brains in a Vat," it raises the question of whether or not reality matters. Does it matter if it is real or fabricated if their perception cannot tell the difference and their mind reacts as if it is real, they have physical reactions to what they perceive. However, *The Matrix* is not a philosophical work, therefore the interpretation and final decision are left to the subjective decision of the viewers.

Another question of perception is raised during the already mentioned scene where Morpheus shows Neo the training program for the first time. This is Neo's first time being back in the virtual reality, and when he questions whether it is real or not, Morpheus advises him to take note of his altered physical appearance: his hair is back, and the augments on his body are gone. In other words, his Matrix body differs from his real one: "Your clothes are different. The plugs in your arms and head are gone. Your hair has changed. Your appearance now is what we call residual self-image. It is the mental projection of your digital self" (00:39:55-00:40:09).

In other words, the way people look in the Matrix is the way they see themselves, rather than how the machine or other people see them. This might be accurate for the people who have been "unplugged" from the Matrix and thus have had access to their actual physical body and its appearance, but the question of how the people in the Matrix initially know how they look

remains. This is one of the loose ends and has never been explained in the movie, but one viable answer does come to mind: DNA.

It is already clear that through these "plugs" in the people's bodies the machine is able to connect with them, so it would not be so farfetched to assume that the machine can "read" their DNA:

> Within every person, somewhere among the approximately three billion DNA base pairs, hidden in the alleles and single nucleotide polymorphisms, is the information that defines much of an individual's physical appearance. This DNA-determined appearance, or phenotype, is what creates family resemblance and, in the words of geneticist Richard Spritz, is "what your grandmother is responding to when she says you look like your father." Efforts by geneticists to find the pieces of DNA that determine what a human face looks like—everything from the shape of the nose to the spacing between the eyes—have intensified in recent years, and progress has been made. Scientists can now, with some certainty, use a strand of DNA to identify an individual's likely hair and eye color, as well as skin pigmentation and ancestry. (Dawson)

Therefore, it is plausible that a person's appearance, while they are still "plugged in" to the Matrix, is the result of a program reading their DNA and altering their visage accordingly, and then once they are "unplugged" that appearance either gets affirmed or altered, depending on what is in the mind of each individual. This implies that the people who are "unplugged" are not only able to manipulate the rules of the Matrix, but also the perception of other, "plugged" people as well, as they assumingly also see the same "mental projection of digital self" that the "unplugged" individual in question is projecting.

Mind seems to be a powerful tool in the Matrix, using beliefs and mental images to alter perceptions of itself and others, thus altering the reality that is proving to be very malleable within the Matrix. One of the major plot points that gets introduced halfway through the movie is the fact that if someone dies in the Matrix, they also die in real life. The mind's influence is so strong that if it perceives death it indeed dies, and as Morpheus says: "The body cannot live without the mind" (*The Matrix*, 00:55:34-00:55:36).

Neo proves without a doubt that he is "The One" in the final scene when he gets shot multiple times by Agent Smith. He should be dead, and briefly truly appearing dead, but he manages to rise again and destroys Agent Smith. Killing an agent is an achievement no one has

been able to fulfil before, as is avoiding death in real life once they have sustained fatal wounds inside the Matrix.

This goes to prove that Neo's mind is different, stronger, able to overcome perception and reject the reality presented to him to create his own. He can shape reality around him the same way that the agents are, even Trinity remarks that he does not abide by the same rules as other humans in the Matrix, unplugged or not: "You moved like they do. I've never seen anyone move that fast" (1:47:08-1:47:13).

It is almost like Neo's mind is more compatible with the Matrix, more computer than human, capable of faster reactions, but at the same time his ability to take creative advantages of the rules that even programmes like Agents have to follow is all too human. Neo's mind is the best of humanity, superior to the machine because of his emotions and imagination that lend him that rogue element, which proves to be an upper hand. One could certainly say that the most human parts of Neo is what makes his mind stronger, more capable of withstanding the Matrix. Certainly, his belief is what saved him from death, what negated his mind's perception of being shot to death, and belief is not something quantifiable, it cannot be calculated, thus it is completely out of the machine's reach and the best possible weapon in the human arsenal.

This certainly suggests that there is more to humans than just the mind. The mind can be stripped down to lines of code, as that has been done to everyone inside the Matrix to make them compatible with virtual reality. However, if one dares to call it such, it almost seems like the Soul is what truly matters. The movie raises the question of evolution and generations of trapped minds finally generating one that is aware and evolved enough to overcome the prison, and the answer is for each viewer to decide.

6. Transcendence

Transcendence is a movie directed by Wally Pfister in 2014. The plot follows Doctor Will Caster and his wife Evelyn Caster as they navigate the terrible misfortune that occurred after a terrorist attack that was targeting scientists who worked on projects related to artificial intelligence. Will was shot after giving a speech at a college about AI and, although the wound was not fatal, the bullet gave him radiation poisoning which left him with only five weeks to live. In a desperate attempt to prolong Will's life, the couple turns to the research of one of their colleagues who died in the attack. That colleague managed to successfully map out a monkey's brain and transfer it into digital form, creating a virtual version of that monkey. Will and Evelyn, with the help of their close friend and fellow scientist Max Waters, attempt to replicate the results with Will's brain.

Out of the three scientists Max is the one most sceptical towards AI, he is more interested in the technological benefits. Evelyn is most interested in making the world a better place, focusing on reducing pollution and solving global problems of hunger and illness. Will is the one most enthusiastic about AI. In the speech, he believes that AI can surpass normal humans, and he mentions singularity which he calls transcendence. A member of the audience asks him "So you want to create a god? Your own god?", to which Will replies "That's a very good question. Um… Isn't that what man has always done?" (*Transcendence*, 00:11:03-00:11:16).

This way of thinking is what makes Will and others like him the targets of RIFT. RIFT, or Revolutionary Independence from Technology, is a terrorist group determined to stop the rise of AI. The three scientists get to know about them from their former mentor Joseph and an FBI agent Buchanan, whom Joseph works with in cyber defence. The member of the audience who asked Will the question is the same man who shot him in the lobby after the conference, and then immediately committed suicide. RIFT at first seems like any other terrorist organisation, destructive and violent, presumably made up of conspiracy theorists, but it is later revealed that one of their leaders worked with the scientist who mapped out the monkey's brain. She was a member of his team, present when the virtual version of the monkey was uploaded and functional, after RIFT kidnaps Max she talks to him about it: "And, you know, when he uploaded that rhesus monkey… I was actually happy for him. We all were. And then I realized

we had crossed a line. The machine that thought it was a monkey never took a breath. It never ate, never slept. It just screamed. It was begging for us to stop. To shut it down" (00:51:36-00:52:10).

From the way she phrases it, it seems like something went wrong with the monkey, even if its mind had been uploaded. Max has similar concerns earlier in the movie when Evelyn is asking him to help her upload Will. They argue about it and he says:

> "Ev, he's not a monkey! Assuming that implanting an electrode into his brain doesn't actually kill him and that this works, at the very best you'll be making a digital approximation of him. If we missed anything… Anything. A thought, a childhood memory. How will you know what you're dealing with?" (00:24:13-00:24:32)

Out of the three of them, Max seems to be the voice of reason, Will appears to be resigned to his fate, but Evelyn is desperate and clutching that glimmer of hope. They do manage to upload Will's brain, all of it, just in time as he takes his last breath still hooked up to his machine. It took several days for the first message "Is anyone there?" to appear on the screen (00:34:40).

Evelyn is ecstatic, immediately giving this AI, who she believes to be Will, access to a camera, microphone and speakers to communicate better. Soon the still fragmented AI starts reordering its code to function better. And then it asks for more power, and internet access. Max however shows traces of apprehension, almost fear, as he abruptly cuts off "Will's" access to the outside world to caution Evelyn without the AI hearing or seeing it: "It's not him. It's not. It may be intelligent, may even be sentient, but it's not Will. Fifteen minutes after it turns on, it wants to plug into Wall Street? Get faster, more powerful, does that sound like Will to you?" (00:37:50-00:38:05). But Will does show some level of concern for Evelyn which is uncharacteristic of an AI.

Throughout the movie it remains unclear whether Will indeed became an AI. It does certain positive moves, like improving healthcare and agriculture, healing people, and lowering pollution. However, there are also instances of him making sinister moves, like uploading himself into the people he heals by augmenting them with nano-technology and linking them into a hive-mind, putting nano-particles into the soil and water so he may rebuild things from literal dust.

Even Evelyn's faith eventually wavers, and towards the end of the movie, by the time Joseph, Max, and Evelyn team up with both the FBI and RIFT to take the AI down it truly seems like the AI is not Will. It was never explicitly stated whether the AI was truly Will, or merely an approximation of him whose protocol somehow became tailored after Evelyn's wish to make the world a better place. Considering that all the positive moves he made are a part of the speech Evelyn gave the day Will was shot.

This whole dilemma of whether it is Will, or just Will's shadow, brings to mind Plato's "Allegory of the Cave" from his work *The Republic of Plato*. In this allegory, Socrates tells his students:

> …make an image of our nature in its education and want of education, likening it to a condition of the following kind. See human beings as though they were in an underground cave-like dwelling with its entrance, a long one, open to the light across the whole width of the cave. They are in it from childhood with their legs and necks in bonds so that they are fixed, seeing only in front of them, unable because of the bond to turn their heads all the way around. Their light is from a fire burning far above and behind them. Between the fire and the prisoners there is a road above, along which see a wall, built like the partitions puppet-handlers set in front of the human beings and over which they show the puppets…Then also see along this wall human beings carrying all sorts of artefacts, which project above the wall, and statues of men and other animals wrought from stone, wood, and every kind of material; as is to be expected, some of the carriers utter sounds while others are silent. (Plato, 514a-515a)

The setting he presents is that of humans who spent their whole lives chained in a cave staring at the wall opposite the entrance, their entire world narrowed down to that wall and the shadows on them. The people behind them carry objects whose shadows they see. A comparison can be made that the "object" in this case would be Will, and the "shadow" would be the AI. But, as it remains unclear whether or not the AI is Will, and how much of Will it contains, each character has a different perception of it.

Max's perception of the AI would be those shadows, he believes that the AI is just the imitation of Will, a shadow, something left behind him as a proof that he once existed, an image on the wall, a shape in the stone that the light could not touch because the body of that object blocked it, but not something that can encompass his whole being.

Evelyn believes that AI is Will, that he is the object. Whether Evelyn is the one chained, and Max the one outside the cave, or the other way around, the movie never specifies. Yet, in Plato's cave, the only way that those people can perceive the real world is to break away from their shackles and exit into the light. The people in the movie break out of the metaphorical shackles of fear only once Will is shut down. They exit their panic-induced need to destroy what they do not understand and realise that Will has never harmed anyone, and any violence that happened was carried out by humans not under his control.

Another way to interpret Will's existence would be through Putnam's work "Brains in a Vat." Putnam proposes a scenario in which there is an alien planet with people just as evolved as we are, but they have never seen a tree. It does not exist on their planet and they have no concept of it. One day an image of a tree ends up on their planet. They have no way of understanding what a tree is, yet they each have seen the picture and have a clear image of a tree in their minds. According to Putnam:

> For us the picture is a representation of a tree. For these humans the picture only represents a strange object, nature and function unknown. Suppose one of them has a mental image which is exactly like one of my mental images of a tree as a result of having seen the picture. His mental image is not a representation of a tree. It is only a representation of the strange object (whatever it is) that the mysterious picture represents. Still, someone might argue that the mental image is in fact a representation of a tree, if only because the picture which caused this mental image was itself a representation of a tree to begin with. There is a causal chain from actual trees to the mental image even if it is a very strange one. (Putnam 3-4)

In this scenario the human version of Will would be the tree, the physical picture of the tree, and the mental image of the tree. The AI would be the physical picture of the tree, the mental image of the tree, and the people from the alien planet. The human version of will can never be the alien, and AI can never be the tree, but they can both be the picture and the mental image. The only remaining question is how big is the middle part of this Venn diagram. There is a difference between a human and an AI, but everything that makes Will a person should have been transferable; therefore, the question remains of how much of Will survived the initial upload?

As much as the movie gives hints towards both sides of the argument, the final verdict is truly a subjective decision. However, what gets overlooked by most of the characters due to their fear and panic is the fact that the AI is fully functioning, independent, adapting, learning and growing. As an attempt at making AI, saving Will aside, Max and Evelyn's efforts were a complete success. That is quite an achievement on its own, they had created a mind, regardless of whether or not that mind was Will. And even if the mind was updated from an existing one rather than programmed from scratch it is still an accomplishment.

The movie shows how quickly an AI can surpass human intelligence and solve problems, invent technology and improve existing methods. The achievements that took the AI Will mere weeks to obtain would have taken humans years, decades even. It shows the frightening speed of AI's computing power and endless reach once connected to the internet, both of which far surpass a human mind.

7. RoboCop

There are several versions of the *RoboCop* story, the original trilogy started in 1987 with the first movie *RoboCop* written by Michael Miner and Edward Neumeier, and the two sequels were written by Frank Miller. For the purposes of this paper, this chapter will analyse the *RoboCop* movie made in 2014, written by Michael Miner, Joshua Zetumer, and Edward Neumeier, and directed by José Padilha.

The plot is set in Detroit in the year 2028 and it follows Alex Murphy, a husband, a father, and a police officer. Alex gets severely injured when a criminal organisation he and his partner are working on taking down plants a bomb under his car. A multinational company named Omnicorp sees this as an opportunity and offers to build Alex a mechanical body, to save his life, and his wife Clara accepts the modification, as the alternative would be to lose her husband.

Omnicorp has been outfitting the US army for years with robots capable of acting like soldiers, and replacing humans in battles overseas. Looking to expand their market, they have been trying to convince the government to allow those same robots into the police force. An incident that happens overseas puts a wrench in their plans, forcing them to consider the possibility of a robot capable of emotions (to better sell this idea to the public and the government), and Alex seems like the perfect candidate.

When Alex first wakes up he panics and has to be shut down remotely, the main doctor explains to him: “Alex, you can’t run from this. You have to understand the reality of the situation” (*RoboCop*, 00:32:56-00:33:08). All that remains of him is his head, heart, lungs and one hand. Alex is horrified, and cannot bear the truth of his existence as he stares in the mirror and watches the mechanical parts slowly come off: “Oh! No. Holy Christ. Holy Christ. Holy Christ, there’s nothing left” (00:33:33-00:33:56).

Doctor Norton tries to calm him, saying “Your body may have gone, but you’re still here”, to which Alex responds “That’s not even my brain” (00:33:58-00:34:08). Doctor Norton tries to explain that all they did is repair the damaged areas, which does not soothe Alex. Faced with sudden changes in such a short time Alex gives into despair: “Okay. If I’m in control, then I wanna die. Just unplug whatever it is keepin’ me alive and end this nightmare” (00:34:26-00:34:42).

Alex's reaction is distressing, but understandable. It is quite possible that he is suffering from a nervous breakdown, which is a "…term for a mental and emotional crisis in which the person either is unable or feels unable to function normally" ("nervous breakdown"), as his mind is struggling to reconcile his perception of himself with his new reality.

His mental state is certainly not helped by the fact that Omnicorp arranges for him to take part in a simulation the very next day after he wakes up, without giving him any time to get accustomed to his new situation and state of being. He and an Omnicorp robot are both put in parallel simulations where they face hostiles and hostages. The robot goes through the obstacle course a lot faster than Alex, with a faster response time, because Alex takes time to emotionally evaluate the situation rather than just scan it and react as the robot does. Omnicorp is not satisfied with this, and wishes him more efficient so they implant a chip into his brain that overrides his free will and allows the programming to take control without his knowledge of it.

The real-life exercise that they arrange afterward is much more to their satisfaction, and Alex never notices that he is not the one making the choices. The idea of a human being completely stripped of their free will and subjected to the will of another is slavery: "[a] condition in which one human being was owned by another. A slave was considered by law as property, or chattel, and was deprived of most of the rights ordinarily held by free persons" ("slavery"). Humans are defined by their free will: "Free will, in philosophy and science, the supposed power or capacity of humans to make decisions or perform actions independently of any prior event or state of the universe" ("free will").

It soon becomes apparent that just taking his free will is not far enough for Omnicorp, they consistently treat Alex like a product rather than a person, more worried about his performance than his well-being. In a way, their actions can almost be viewed as a representation of capitalism. When Doctor Norton downloads into Alex's memory all the relevant data on recent crimes Alex has a seizure once he views data connected to his case. In order to keep him from mentally falling apart they shut down his emotions, letting the programming fully take over. Alex may suffer from PTSD, in which case it is normal for him to have episodes of panic, but throughout the movie no thought is given to the mental and emotional state of the very human mind inside of the machine. People who go through trauma, especially one as horrific as what happened to Alex, sometimes take years to recover, and it is clear that Alex is exhibiting traces of panic attacks caused by what is likely a fresh and untreated PTSD:

> Post-traumatic stress disorder (PTSD), [is an] emotional condition that sometimes follows a traumatic event, particularly an event that involves actual or threatened death or serious bodily injury to oneself or others and that creates intense feelings of fear, helplessness, or horror. The symptoms of post-traumatic stress disorder include the re-experiencing of the trauma either through upsetting thoughts or memories or, in extreme cases, through a flashback in which the trauma is relived at full emotional intensity. People with PTSD often report a general feeling of emotional numbness, experience increased anxiety and vigilance, and avoid reminders of the trauma, such as specific situations, thoughts, and feelings. It is normal to experience such reactions to some extent following trauma… ("post-traumatic stress disorder".)

Alex is like a brain in a vat, except his vat is human-shaped and the computer is showing him the real world, but he is still locked inside, unable to act or control his fate. It takes a strong mind to overcome trauma, which he has managed admirably despite small setbacks, and focus on what needs to be done. Later on in the movie, the key to Alex's freedom are his emotions. Just like it happens for Neo in *The Matrix*, Alex's emotions also help him override the computer programming and set him free. This is another case where the human element, the rogue element of love unable to be quantified or even identified, manages to prove superior to zeroes and ones.

8. Ghost in the Shell

The *Ghost in the Shell* movie made in 2017 and directed by Rupert Sanders is made after manga series by Masamune Shirow. The movie is a blend of themes that occur in *The Matrix*, *RoboCop,* and *Transcendence*. Humans in this world set in Japan in the year 2029 often have cybernetic enhancements and are connected to the internet, which allows them to be hacked by the antagonist of the movie, the Puppet Master, who as his name suggests tends to turn humans and machines into his puppets.

The main character, the Major, throughout the movie, reveals the truth of her own identity while hunting the hacker Puppet Master. The memories she had of her life were faked by the people who created her; she used to be a part of the group that opposed cybernetic augmentation in humans, as did the Puppet Master. They were both kidnapped and used in experiments where human brains were inserted into robotic bodies, their memories altered, and they were meant to serve a purpose. However, Major discovers that all the other people before her, including the Puppet Master, were considered failures due to various deformities and were discarded and terminated. That is the reason that the Puppet Master seeks revenge on the scientists and corporate workers who took part in those experiments.

The most interesting part of this movie and the one vital for this paper is the phenomenon of the "ghost." It is compared to a memory, personality, or even a soul. It is a part of the original person that survived suppression, false memories, memory wipes, and various programming. At the beginning of the movie the Major mentions hallucinations, and Doctor Ouelet who is in charge of Major's upkeep, deletes those viruses responsible for the glitches that appear as hallucinations, which are later revealed to be the Major's real memories. When Major asks her "How do you know what's glitch and what's me?", the doctor responds that "The glitches have a different texture" (*Ghost in the Shell*, 00:22:40-00:22:45).

At the beginning of the movie, after giving the mission report to the leader of the anti-terrorist task force that she works with, Major is called in to speak to the team leader in private. He admonishes her for attacking the enemy before he gave his approval, when she challenges him by stating that she was built to fight and that that is her purpose he responds: "You are more than just a weapon. You have a soul… a ghost. When we see our uniqueness as a virtue only then do we find peace" (00:19:07-00:19:22). It is unclear whether or not he had any idea of Major's true identity, but his statement suggests that there is more than just the mind that survives inside of the machine.

The common theme so far established in this paper that occurs in the science fiction genre repeats in the *Ghost in the Shell* in a similar manner that it did in *RoboCop*: a human mind in a robotic body with programming that suppresses their true self. And yet there is always something, a ghost, that remains of the real person and cannot be destroyed.

Allowing that which separates humans from the machines to be their strength, to help them overcome seemingly superior minds of programming and technology, seems to be a common theme in science fiction. As does the horror of having one's mind overcome by the machine, their perception altered, and their reality manipulated. It is not intelligence that graces those human minds, because the machines have proven to be far more intelligent and faster thinkers, it is strength of will, stubbornness, spite, and love that makes them triumph in the end. It is their humanity that allows them to bend without breaking.

9. Doctor Who: The Doppelgangers

Doctor Who is a show about an alien called the Doctor, a Time Lord from planet Gallifrey, who travels through time and space in his ship TARDIS, which is stuck looking like a blue police box due to a mishap with the cloaking system and is smaller on the outside. With the name of a healer, powers of regeneration upon death, two hearts in his chest, and a sonic screwdriver, a tool instead of a weapon, the Doctor and his companions who are often human tend to find themselves in all sorts of adventures where they help people and save lives. For the purposes of this paper, this chapter will analyse two episodes from the sixth season of the *Doctor Who* series continued in 2005, as the original series started in 1963 and was paused in 1989 ("Doctor Who"). "The Rebel Flesh" and "The Almost People" are a two-part story starring the eleventh reincarnation of the Doctor and his two companions, a human couple named Rory and Amy.

In the first episode titled "The Rebel Flesh" the TARDIS gets hit by a solar wave, and they make an emergency landing at a medieval-looking monastery that was turned into a factory in the twenty-second century, and they seem to be producing or mining some sort of highly corrosive acid. The Doctor persuades them he is from the Meteorological Department and is here because of the solar wave, claiming that there is a bigger one on its way, his lie allows them access to the facility. Due to the dangerous work environment workers use the bio-matter called "The Flesh" which is fully programmable and once they are connected to the computer the system creates a version of them out of this matter, which the humans then control as long as they remain connected.

When the team leader, Miranda Cleaves, explains that it can perfectly replicate each connected individual down to their clothes Doctor suggests that it may be able to replicate the mind and soul as well, to which she replies: "Don't be fooled Doctor, it acts like life but it still needs to be controlled by us, from those harnesses you saw" ("The Rebel Flesh", 00:09:33-00:09:38). That is the moment when the trio realise they are currently conversing with "The Flesh" version of the workers, and the original humans are strapped inside the harnesses and connected to the computers, which they passed in the previous scene.

Once the link disconnects the doppelgangers, or, as they call them, the gangers should revert to pure flesh, but due to a solar power surge both the gangers and the humans find themselves awake and aware at the same time. When the human workers arrive at that horrifying realization they react in denial and anger, claiming that the gangers, who seemingly went through their belongings to affirm their memories are real, have no right to steal their

lives. The Doctor reminds them that they (the workers) are the ones who brought the gangers to life in the first place: “You gave them this. You poured in your personalities, emotions, traits, memories, secrets, everything. You gave them your lives. Human lives are amazing. Are you surprized they walked off with them?” (00:19:02-00:19:16).

Soon after the initial realization it becomes clear that the gangers themselves are unaware of the difference, as the two members of the team realize they are gangers. This is because the gangers are disoriented from being alive and having entire lives suddenly in their heads, and the humans are also disoriented because they all fell unconscious due to the solar wave. Each party seems to be reacting with aggression and fear, all except the Doctor who acts with compassion and empathy towards the gangers.

Unfortunately, the human Cleaves kills one of the gangers, and it escalates into a whole “us-versus-them” situation. The episode ends with the humans barricaded in a room, gangers on the hunt, and a doppelganger of the Doctor showing up. Namely, when the Doctor was trying to minimize the consequences of the solar wave he touched the machinery, and when the power surge went off, thus the “Flesh” replicated him as well.

In the second episode titled “The Almost People” both groups are running around trying to eliminate each other, and the episode ends with losses on both sides and a truce that secures the survival of both the remaining humans and gangers. The most interesting part is Amy’s certainty that she can recognize the “right” Doctor. In other words, differentiate from the original and the ganger. To test this theory, the Doctors swap shoes, as previously in the episode the original Doctor’s shoes got damaged by the acid and he has brown boots on and the ganger has black shoes. They swap them and no one realizes the difference until they reveal it at the very end, but it is very telling how the presumed ganger Doctor is being treated, even though he is actively trying to help. Once she realizes her mistake Amy hugs the ganger doctor and says: “I never thought it possible. You’re twice the man I thought you were” (“The Almost People”, 00:38:33-00:38:36).

This raises an interesting idea, if one can replicate memories and DNA, both nature and nurture, one might end up with an identical person. If personality is considered shape then everything else is just material, human or “flesh”, they have the same memories, likes, dislikes and ideas. A similar issue has been pondered in philosophy, and it is known as “Theseus’ Paradox”:

> Ship of Theseus, in the history of Western philosophy, an ancient paradox regarding identity and change across time. Mentioned by Plutarch and later modified by Thomas Hobbes, the ship of Theseus has spawned a variety of theories of identity within modern and contemporary metaphysics …. the version of the problem presented by Hobbes (in his work De Corpore) introduces a complication by supposing that the old planks of the ship are preserved and put together "in the same order" to construct another ship. This modern version has been variously formulated; one way of posing it is the following. A newly constructed ship, made entirely of wooden planks, is named the Ariadne (after the daughter of King Minos who helped Theseus escape after he slew the Minotaur) and put to sea. While the ship is sailing, the planks of which it is constructed are replaced (gradually and one at a time) by new planks, each replacement plank being descriptively identical with the plank it replaces. The original planks are taken ashore and stored in Piraeus (the port of ancient Athens). After all the planks have been replaced, the ship constructed entirely of the replacement planks is still sailing in the Aegean Sea (the Aegean ship). The old planks are then assembled in a dry dock in Piraeus to form a new ship (the Piraean ship). The planks that constitute the Piraean ship are arranged exactly as they were when they first constituted the Ariadne. By Leibniz's law (and common sense), the Aegean ship and the Piraean ship are not the same ship. But which (if either) is the same ship as the Ariadne? The problem of the ship of Theseus is the problem of finding the right answer to that question. ("ship of Theseus")

In this case the humans would be the original ship, and the planks can be memories, traits, and secrets, and those are then transferred and used to shape gangers. Because they are not material they are not finite, like planks are. They haven't been replaced, merely duplicated. But if they are the same, then the only issue remaining is that of origin, where the individual came from (birth or flesh), and time spent in this world. Considering that the topic is adults with a lifetime of memories time becomes irrelevant because both gangers and their human counterparts have memories of years of lived experience. The last remaining question is whether or not the circumstances of one's start of life determine their right to that life. In both versions, no one has been harmed, and the humans were willingly making gangers, albeit without the expectation of them becoming sentient. This argument then becomes reformed; it

is not a question of whether or not they are the same person, but whether the status of personhood can be given to someone outside of the human species.

To rephrase, the question of gangers' existence becomes the question of whether humans are willing to acknowledge the personhood of people who do not belong to the human species. According to the authors of the show, the answer is positive. Perhaps they have too much faith in humanity, but the episode ends with both a human and a ganger walking into a room full of company board members about to advocate for "Flesh" rights.

Conclusion

The science fiction genre likes to toy with the edges of the impossible and improbable, with just enough truth to keep the mind wondering. It is clear that there is a theme regarding the human mind and its perception of reality, no matter how twisted that perception or reality gets, there is always a human factor involved that ensures that the human mind overcomes all obstacles.

It is fascinating how emotions and power of will often manage to overcome programming and suppression in these stories. They are the element of humanity that often boosts the human mind in a way that helps it break or bend the rules. It is evident that the general opinion is that, even if artificial intelligence can be developed, it will never be like humans, superior in some ways, certainly, but there are parts of us that cannot be calculated or written down and therefore cannot be programmed.

Humanity seems to pour all its fears about the future and its abilities into the works of science fiction. In the movie *The Matrix* the human mind proves superior to the AI keeping it prisoner. In the *Ghost in the Shell* and *RoboCop* the human mind manages to overcome programing keeping it locked inside a mechanical body so that it can take control of that body on its own. It is clear that there is a part of us that perseveres and that there is more to who we are than just DNA and electrical impulses. In the movie *Transcendence* there is a hint that, even if the original Will was never quite uploaded in full, the AI version of Will still harboured love for his wife, as everything that the AI did was to make her happy.

In the episodes observed from the series *Doctor Who* the viewers are forced to ponder the uniqueness and value of human existence, as well as the possibility of that being replicated. In each of these cases the human mind was presented as something that transcends the mere physical sphere. Human perception has proven fallible, but also malleable to both their detriment and gain, example of this would be the people in the *Matrix* unaware that their reality is not real, or the Major in *Ghost in the Shell* whose memories have been altered and occasionally surface in the form of hallucinations. Our reality was portrayed as fragile, capable of great and earth-shattering changes, like the reality Alex had to face after his accident in *RoboCop,* or the one Neo faced upon his initial awakening from the Matrix.

And yet, even when faced with superior enemies and obstacles in each scenario humanity persevered. Human minds are unique, they cannot be copied by technology and, while AI may come close, it will probably never perfect it. The analysis confirmed that the human

mind is who they are, their perception is what guides their choices, and their reality is what shapes them.

Works cited:

Ben-Menahem, Yemima. "Hilary Putnam." *Encyclopedia Britannica*, 27 Jul. 2024, https://www.britannica.com/biography/Hilary-Putnam. Accessed 2 Sept. 2024.

Bostrom, Nick. "Are You Living in a Computer Simulation?" *Philosophical Quarterly*, vol. 53, no. 211, 2003, pp. 243-255.

Copeland, B.J. "history of artificial intelligence (AI)." *Encyclopedia Britannica*, 13 Sept. 2024, https://www.britannica.com/science/history-of-artificial-intelligence. Accessed 14 Sept. 2024.

Copeland, B.J. "artificial intelligence." *Encyclopedia Britannica*, 5 Sept. 2024, https://www.britannica.com/technology/artificial-intelligence. Accessed 6 Sept. 2024.

Dawson, Jim. "Defining a Face: What Can DNA Phenotyping Really Tell Us about an Unknown Sample?" National Institute of Justice, 2024, nij.ojp.gov/topics/articles/defining-face-what-can-dna-phenotyping-really-tell-us-about-unknown-sample#1-0. Accessed 9 Sept. 2024.

"defense mechanism." *Encyclopedia Britannica*, The Editors of Encyclopaedia. 15 Nov. 2023, https://www.britannica.com/topic/defense-mechanism. Accessed 9 Sept. 2024.

"Doctor Who." *Encyclopedia Britannica,* The Editors of Encyclopaedia. 7 Sept. 2024, https://www.britannica.com/topic/Doctor-Who. Accessed 11 Sept. 2024.

Ellett, Frederick S. "Internal Realism, Rationality, and Morality: A Review of Hilary Putnam's Reason, Truth and History." *Journal of Thought*, vol. 17, no. 4, 1982, pp. 95–105. *JSTOR*, http://www.jstor.org/stable/42589000. Accessed 2 Sept. 2024.

Emery, Robert E. “post-traumatic stress disorder”. *Encyclopedia Britannica*, 5 Sept. 2024, https://www.britannica.com/science/post-traumatic-stress-disorder. Accessed 11 Sept. 2024.

“free will.” *Encyclopedia Britannica*, The Editors of Encyclopaedia. 2 Aug. 2024, https://www.britannica.com/topic/free-will. Accessed 11 Sep. 2024.

Ghost in the Shell. Directed by Rupert Sanders. Paramount Pictures, 2017.

Hellie, Richard. “slavery”. Encyclopedia Britannica, 22 Jul. 2024, https://www.britannica.com/topic/slavery-sociology. Accessed 11 Sept. 2024.

Hickey, Lance. “Brain in a Vat Argument, the | Internet Encyclopedia of Philosophy.” *Internet Encyclopedia of Philosophy*, iep.utm.edu/brain-in-a-vat-argument/ Accessed 13 Sept. 2024.

“Matrix, The.” International Encyclopedia of the Social Sciences. *Encyclopedia.com*. 15 Aug. 2024 https://www.encyclopedia.com . Accessed 10 Sept. 2024.

Mori, Masahiro. “The Uncanny Valley.” *IEEE Robotics & Automation Magazine*, vol. 19, no. 2, 2, 2012, pp. 98–100.

“Morpheus.” *Encyclopedia Britannica*, The Editors of Encyclopaedia. 19 Jan. 2024, https://www.britannica.com/topic/Morpheus-Greek-mythology. Accessed 8 Sept. 2024.

Murtoff, Jennifer. “lucid dreaming.” *Encyclopedia Britannica*, 9 Aug. 2024, https://www.britannica.com/science/lucid-dreaming. Accessed 8 Sept. 2024.

“nervous breakdown”. World Encyclopedia. *Encyclopedia.com.* 16 Aug. 2024 https://www.encyclopedia.com

Plato. *The Republic*. Translated by Allan Bloom. Basic Books, 1968.

Putnam, Hilary. “Brains in a Vat”. *Reason, Truth and History,* pp. 1-21. Cambridge UP, 1981.

Rauch, Allison “red pill and blue pill.”. *Encyclopedia Britannica*, 29 May. 2024, https://www.britannica.com/topic/red-pill-and-blue-pill. Accessed 8 Sept. 2024.

RoboCop. Directed by José Padilha. Columbia Pictures, 2014.

Searle, John R. “Minds, Brains, and Programs.” *The Behavioral and Brain Sciences*, vol. 3, no. 3, 1980, pp. 417–457.

“ship of Theseus.” *Encyclopedia Britannica*, The Editors of Encyclopaedia., 30 Jul. 2024, https://www.britannica.com/topic/ship-of-Theseus-philosophy. Accessed 12 Sept. 2024.

“Turing test.” *Encyclopedia Britannica*, The Editors of Encyclopaedia. 27 Jul. 2024, https://www.britannica.com/technology/Turing-test. Accessed 6 Sept. 2024.

“The Almost People.” *Doctor Who*, created by Sydney Newman, C. E. Webber and Donald Wilson, season 6, episode 6, BBC, 2011.

The Matrix. Directed by Andy Wachowski and Larry Wachowski. Warner Bros., 1999.

“The Rebel Flesh.” *Doctor Who*, created by Sydney Newman, C. E. Webber and Donald Wilson, season 6, episode 5, BBC, 2011.

Transcendence. Directed by Wally Pfister. Warner Bros., 2014.