

ISTANBUL BILGI UNIVERSITY
INSTITUTE OF SOCIAL SCIENCES
COMPARATIVE LITERATURE MASTER'S DEGREE PROGRAM

THE ROLE OF SCIENCE FICTION IN HUMAN-ROBOT INTERACTION



Nihan İŐLER

14667003

Prof. Dr. Jale PARLA

İSTANBUL

2018

The Role of Science Fiction in Huan-Robot Interaction
İnsan-Robot Etkileşiminde Bilim Kurgunun Rolü

Nihan İşler

114667003

Tez Danışmanı:

Prof. Dr. Jale Parla

Jüri Üyesi:

Prof. Dr. Murat Belge

Jüri Üyesi:

Jüri Üyesi:

Prof. Dr. Sibel Irzık

Jale Parla
Murat Belge
Sibel Irzik

Tezin Onaylandığı Tarih: 30.01.2018

Toplam Sayfa Sayısı: 64

Anahtar Kelimeler (Türkçe)

- 1) Bilim Kurgu
- 2) Robot
- 3) Yapay Zeka
- 4) Psikanaliz

Anahtar Kelimeler (İngilizce)

- 1) Science Fiction
- 2) Robot
- 3) Artificial Intelligence
- 4) Psychoanalysis

ACKNOWLEDGEMENTS

I would first like to thank my supervisor Prof. Jale Parla who always kept her office open and hearten me whenever I needed with her profound knowledge and smiling face. I would also like to thank Assistant Professor Rana Tekcan for answering my endless questions about thesis process and my committee Prof. Murat Belge and Prof. Sibel Irzık for their valuable insights. I am also indebted to Dr. Süha Oğuzertem for making us strive for perfection in our academic writings.

I would also like to express my gratitude to my friend Funda Çankaya for sharing the process with me and to Gökçe Göbüt for always providing me the brain when I needed more than one for brain storming.

And last but not least, I am very grateful to my better half, İsmail İbiloğlu for bearing my mood swings and showing his support by diving into science fiction with me through the whole process. I felt lucky.

ABSTRACT

This study discusses the human attitude towards life forms of their own creation as an intellectual laboratory by examining the themes of robot and artificial intelligence in science fiction literature. First chapter summarizes the role of automaton in both intellectual history and practice from the days of oral tradition. The second and the third part aims to provide an insight for potential problems of our modern world through the comparison of conflicts and their reasons about human-robot interaction in the 19th and the 20th century science fiction novels and short stories. Thus, the works that cited in the study are not only from the fields of literary theory, psychoanalysis, posthumanism and science fiction but also from the fields of ethics, artificial intelligence and robotics.

Keywords: science fiction, robot, artificial intelligence, psychoanalysis

ÖZET

Bu çalışma, bilim kurgu edebiyatında işlenen robot ve yapay zeka temalarını interdisipliner temelde inceleyerek, insanlığın kendi üretimi olan bu yaşam formlarına karşı tutumunu şimdiye ve yakın geleceğe dair düşünsel bir laboratuvar olarak ele alır. İlk bölüm, sözlü gelenek döneminden itibaren otomatonların düşünce tarihindeki ve pratikteki yerini özetlemektedir. İkinci ve üçüncü bölüm ise 19. ve 20. yüzyıl bilim kurgu roman ve kısa hikayelerinde yapay zeka ve robotların insanlarla olan etkileşimindeki problemleri ve bu problemlerin nedenlerini ele alırken günümüz teknolojileriyle kıyaslayarak olası sorunlara ışık tutmayı amaçlar. Bu sebeple yararlanılan kaynaklar edebiyat teorisi, psikanaliz, posthumanizm ve bilim kurgu alanındaki çalışmaların yanı sıra etik, yapay zeka ve robotik teknolojileri alanında yapılan çalışmaları da kapsamaktadır.

Anahtar kelimeler: bilim kurgu, robot, yapay zeka, psikanaliz

TABLE OF CONTENTS

ACKNOWLEDGEMENTS.....	iii
ABSTRACT / ÖZET.....	iv
TABLE OF CONTENTS.....	vi
INTRODUCTION.....	1
CHAPTER I: A BRIEF HISTORY OF AUTOMATONS.....	4
CHAPTER II: CLOSED CIRCUITS.....	9
A. Charming Toys.....	9
B. Perfect Servants.....	15
C. Dominant Caretakers.....	26
CHAPTER III: OPEN MINDS.....	31
A. Omniscient Computers.....	31
B. Children of War.....	38
C. Sentimental Companions.....	44
CONCLUSION.....	54
BIBLIOGRAPHY.....	57

INTRODUCTION

Artificially created life forms have been in the agenda of humankind since the days of oral tradition. In the process of time, humanity took the role of gods; the capability to build artificial beings has lost its feature of being common to divinity and found a way to fall within the course of humans. Nowadays, robots appear not only in fantasies, but also in literature, film industry, television, video games and real life. This technological innovation, as it always has been, carried the fear with it, arousing from several reasons.

This thesis aims to examine the relationship between human and robots through the science fiction literature. The term “robot” is used broadly here. It may refer to single-job-oriented machines, mechanical humanoids or biological-looking creations. The work consists of three chapters. First chapter is entitled “A Brief History of Automaton” and focuses on the self-autonomous machines and artificial human forms in mythology and literature as well as the advancements in real-life automatons.

The second chapter is titled as “Closed Circuits” which refers to single-minded machines. This chapter consists of three parts, “Charming Toys”, “Perfect Servants” and “Dominant Caretakers” respectively. The first part discusses two early examples of automatons in the 19th century literature in compliance with

psychoanalytic approach of Freud. The two stories addressed here are “The Sandman”, written by E.T.A. Hoffman and “The Dancing Partner” by Jerome K. Jerome. The second part deals with machines as servants and cites two narratives which includes one play and one short story. The play is Karel Čapek’s *R.U.R.* and the short story is “With Folded Hands...” by Jack Williamson. The third part holds a discussion on robots who are responsible from humanity’s wellbeing. The works studied in this chapter are “The Machine Stops” by E.M. Forster and “The Happy Breed” by John T. Sladek.

The third chapter of the thesis is also divided into three parts and focuses on the sentient machines and deliberates over the notions sentience, intelligence, personhood and empathy. The first part is “Omniscience Computers” and describes the intelligent software in terms of science fiction. The two exemplary works are a short story “A Logic Named Joe” by Murray Leinster and the novel *Golem XIV* by Stanislaw Lem. The second part is titled as “Children of War” and it centers upon the robots which are created for military purposes. The stories discussed here are “The Defenders” from Philip K. Dick and “I Have No Mouth, and I Must Scream” by Harlan Ellison. And finally, the third part “Sentimental Companions” keeps its focus on the machines with feelings and humanities identification towards them. The narratives at issue are the novel *Do Androids Dream of Electric Sheep?* by Philip K. Dick and the short story “Helen O’Loy” by Lester Del Rey.

The sources used in this thesis are not only the literary studies on science fiction genre, but also the scientific researches on the fields of artificial intelligence and robotics in order to keep it in perspective. While scientific and technological developments provide the inspiration for science fiction genre, science fiction also infuses into science and technology. Technologies like space travelling, instant messaging, electrical cars and internet have fade in to humanity's domain through science fiction for the first time. And now,

At this particular moment in time, reality and science fiction are moving into such close conjunction that science fiction is no longer the strange reflection and artistic elaboration of current preoccupations: the mirror and the actuality have almost become one.

(Jones, vii)

Therefore, in these days when a robot has granted a citizenship for the first time, it will be useful to observe the intellectual laboratories of science fiction in an attempt to better understanding of current age.

CHAPTER ONE

A BRIEF HISTORY OF AUTOMATONS

Creating and imitating life has been one of the most enchanting and inspiring challenge of the humankind throughout the history. The representation of this fantasy can be seen in a large variety of myths, tales, novels, stories, plays and movies in different forms. The word “automaton” means "acting of one's own will" in Greek and the earliest use of the word is by Homer in *Iliad* as far as is known. He uses the term to refer automatically opening doors and intelligent serving tripods. These automatons mentioned in the *Iliad* are usually crafts created by the Greek god of blacksmiths and metalworking, Hepaestus. Talos, for example, is a giant bronze automaton which has crafted by him to protect Europa. *Liezi* or *Lieh Tzu* is an early Taoist text from ancient China approximately 5th century BCE, contains a story about a man who presents an automaton that can dance and sing to the king. In 1206, Al-Jazari, a polymath from Cizre, describes his constructions like hand washing automaton or flush toilets like we use today. Humanoid automatons which are programmable complex machines are also included in the book in his *Book of Knowledge of Ingenious Mechanical Devices*.

The legend of Pygmalion which is most famously known from Ovid's *Metamorphoses* can be considered as the origin of artificially created human themed

narratives. According to Ovid's masterpiece, Pygmalion carves a statue of a perfect woman from ivory and falls in love with her. Then he makes offerings to Aphrodite and when he backs home, the lips of the sculpture turns warm while he is kissing them. With the blessings from Aphrodite, Pygmalion marries her. This story of a sculpture coming to life then becomes a very popular theme through the centuries and its influence can be traced in so many works of arts from Pinocchio to Shakespeare's plays. *The Future Eve* is also one of these examples. This science fiction book published in 1886 has written by the French author Auguste Villiers de l'Isle-Adam and the main character of the book is fictionalized Thomas Edison. His friend Lord Ewald finds his fiancé Miss Evelyn beautiful but intellectually dull, so Edison offers to build a machine version of her to him. The book refers to this machine woman as "android" and popularizes the term.

In his book *Science Fact and Science Fiction*, Brian M. Stableford composes an encyclopedia where he examines the connection and comparison between science and science fiction. In its chapter on robots, Stableford summarizes the automatons in real life as follow:

Heron of Alexandria apparently build various automata as toys in the first century A.D. Although the mechanical servant allegedly constructed in the thirteenth century by Albertus Magnus and the talking head allegedly made by Roger *Bacon are certainly mythical,

clocks began to be equipped with elaborate automatic striking devices in the fourteenth century and Gianello dell Torre of Cremona built a mechanical figure of a girl playing the lute in the 1540s. A Japanese tradition of theatrical automata was launched in the seventeenth century. There was a considerable vogue in eighteenth -century Europe for the construction of ingenious mechanical automata; the most famous example was a duck designed by Jacques de Vaucanson.

(442)

Next well-known automaton in the history is a fake one. In 18th century a chess-playing machine named The Turk was constructed in 1770 by Wolfgang von Kempelen and travelled around Europe and America for 84 years. While it was performing entrancing demonstrations against famous names like Napoleon Bonaparte and Benjamin Franklin, the truth of the matter is that there was a chess master inside the machine operating it. The secret behind this intelligent automaton has been a brain-teaser for many intellectuals of the time and considerable number of articles and books had written about it or including it, like the short story “The Automata” by E.T.A. Hoffmann or the essay titled "Maelzel's Chess Player" written in 1836 by Edgar Allan Poe.

At the early years of 19th century, Mary Shelley wrote *Frankenstein; or, The Modern Prometheus* and the monster devastating Frankenstein’s life has become a

reference for many science fiction studies. Szollosy presents some of these approaches in his article “Freud, Frankenstein and our fear of robots: projection in our cultural perception of technology”. By pointing at the correlation between Frankenstein and Faust, he states that it is inevitable for men to be haunted by the monsters of their hubris which is a referring to science and technology.

When we see Frankenstein in the context of robots, we realise that it is not just technology that we fear, or that technology will gain autonomy and move beyond our control. Rather, looking at Frankenstein, we learn that what we fear is the very quality of ourselves that enables us to create the monster; partly, this is ambition, hubris, etc., but also, and more specifically in this cultural context, we are becoming the robots that we so fear. We fear becoming an empty, mechanical shell of cold, unfeeling rationalism. We are afraid of losing, or that we have already lost, the very qualities that we deem to define us as human. (Szollosy, 435)

Right before the 19th century ends, the American author Ambrose Bierce writes a short story called “Moxon’s Master”. The inspirations of Frankenstein’s monster and The Turk can be clearly distinguished in the narrative. This early example of stories about machines turning against their creators starts with a question asked by narrator: “Are you serious?—do you really believe a machine thinks?” Thereupon a discussion takes place between the narrator and automaton builder

Moxon that is established on definitions of life, machine and thinking. Cartesian impressions may be seen in Moxon's thoughts on how men can be considered as machines(1) and implementing the idea "If consciousness is the product of rhythm all things are conscious, for all have motion, and all motion is rhythmic"(5) to narrator.

In his work *Treatise on Man*, Descartes touches on the subject:

I suppose the body to be nothing but a statue or machine made of earth, which God forms with the explicit intention of making it as much as possible like us. [...] We see clocks, artificial fountains, mills, and other such machines which, although only man-made, have the power to move of their own accord in many different ways. But I am supposing this machine to be made by the hands of God, and so I think you may reasonably think it capable of a greater variety of movements than I could possibly imagine in it, and of exhibiting more artistry than I could possibly ascribe to it. (99)

After the conversation, Moxon gets back to his machine workshop and murdered by his own creation during the chess play. Soon enough, science fiction genre is seized by the scenarios of robots slaughtering humanity which is highly apprehensible according to Adam Roberts' *Science Fiction*: "The robot is that place in an SF text where technological and human are most directly blended. The robot is the dramatisation of the alterity of the machine, the paranoid sense of the inorganic come to life" (161).

CHAPTER TWO

CLOSED CIRCUITS

A. Charming Toys

Ernst Theodor Amadeus Hoffmann, more widely known as E.T.A. Hoffmann is a Prussian jurist, musician and author from the early 19th century. “The Sandman” may be considered as one of the most influential short story of this romantic luminary. In this 1816 dated horror story, a young man named Nathanael writes to his friend about his childhood trauma and how it preoccupies his mind lately. But he receives the reply from his fiancé Klara, which explains that he accidentally wrote her name on the envelope instead of his brother and advises him to forget these figments of his imagination. He returns to his hometown from where he studies and reads his new poem about the previous trauma to Klara. When she tells him to throw that poem to fire, Nathanael storms out right before shouts her “damned, lifeless automaton” (232). Offended from her insensible acts towards his artistic activities, Nathanael departs for his studies again. His professor Spalanzani builds “a tall, very slender, beautifully dressed, beautifully proportioned young lady” (225). Nathanael sees her from the opposite window and falls for her without noticing the fact that she is an

automaton. When Spalanzani throws a party to introduce Olympia as her daughter, Nathanael gets a chance to see her up close and describes her as follows:

There was something peculiarly curved about her back, and the wasplike thinness of her waist also appeared to result from excessively tight lacing. There was, further, something stiff and measured about her walk and bearing which struck many unfavorably. (237)

The question of why this measured movements of her considered as “unfavorably” is a tricky one which still remains in the agenda of psychology, robotics and film industry.

Ernst Anton Jentsch is a German psychiatrist who is known for his work *On the Psychology of the Uncanny* which also inspired Sigmund Freud to write his famous essay *The Uncanny*. To understand the concept of Uncanny, we should clarify the etymology of the original word, “unheimlich”. The German word “heimlich” means “belong to the house, familiar and not strange”. So “unheimlich” as the opposite of “heimlich” can be translated as “unhomely”, which is not an existing word. Instead, it has translated and commonly used as “uncanny”. Based on this etymology, Freud gives the definition of uncanny as “class of terrifying which leads back to something long known to us, once very familiar” (1-2). This is merely a definition but a substitution of a peculiar kind of feeling of dread and terror. But according to Jentsch, one should not give the definition of the uncanny anyway.

So if one wants to come closer to the essence of the uncanny, it is better not to ask what it is, but rather to investigate how the affective excitement of the uncanny arises in psychological terms, how the psychological conditions must be constituted so that the ‘uncanny’ sensation emerges. (3)

In addition to that, he gives the example of a prosthesis hand in order to investigate those sensations; if one sees a man with prosthesis hand in dark and couldn’t realize that the hand is prosthesis, once he shakes the hand, the sense of cold plastic hand will create the effect of uncanny, since a hand, as a familiar organ, has become unfamiliar. Hoffmann uses this specific example in his story: Nathanael “grasped her hand. It was cold as ice. A deathly chill passed through him” (237).

Both Jentsch and Freud mentions Hoffmann’s *The Sandman* in their essays. “In storytelling, one of the most reliable artistic devices for producing uncanny effects easily is to leave the reader in uncertainty as to whether he has a human person or rather an automaton before him in the case of a particular character” (11) Jentsch states. Indeed, there is no certainty about Olympia being an automaton until the moment that Nathanael finds out. Still, there are some clues which imply that she is somehow not human, but again, this obscurity leads us to uncanny:

This is done in such a way that the uncertainty does not appear directly at the focal point of his attention, so that he is not given the occasion to investigate and clarify the matter straight away; for the particular

emotional effect, as we said, would hereby be quickly dissipated. In his works of fantasy, E. T. A. Hoffmann has repeatedly made use of this psychological artifice with success. The dark feeling of uncertainty, excited by such representation, as to the psychical nature of the corresponding literary figure is equivalent as a whole to the doubtful tension created by any uncanny situation, but it is made serviceable by the virtuosic manipulation of the author for the purposes of artistic investigation. (11-12)

This concept of Jentsch gave the inspiration to Masahiro Mori to evaluate the theory of “the Uncanny Valley”. According to this theory, any robot which become more humanlike, will appear more familiar to humans. But at one point, if this resemblance is too high but not perfect, the little differences and imperfections make those robots look eerie and they don’t seem familiar anymore. That pit between resemblance and perfection is called Uncanny Valley. He also mentions that the movement can increase this effect of uncanny. (see fig. 1)

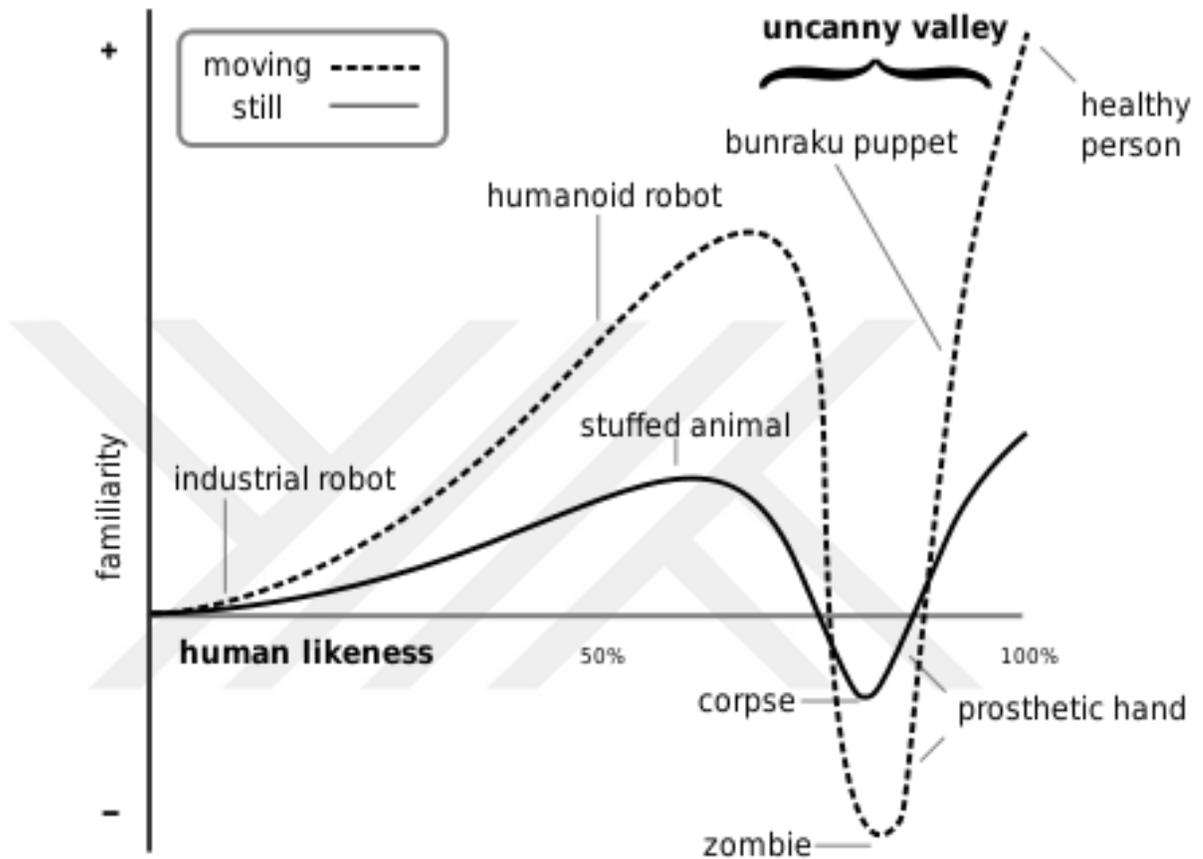


Fig. 1. [Note: Translated and simplified version by MacDorman from Mori's *Energy* article]

If it is taken into consideration that the definition of liveliness depends on animation, it can be seen that it is not easy to say “lifeless” to these machines. As a matter of fact, the word “animation” comes from the Latin word “anima” which means soul and life. It also sheds light on the correlation between stillness and human likeness on Mori's graphic of Uncanny Valley. Accordingly, uncanny effect arouses more dramatically when the automaton moves.

All men in the story, except the protagonist, suffers from this eeriness of movements. One of Nathanael's best friends, Siegmund's explanation on how they feel about Olympia sets a great example for this uncanny feeling:

She seems to us [...] strangely stiff and soulless. Her figure is symmetrical, so is her face, that's true enough, and if her eyes were not so completely devoid of life – the power of vision, I mean – she might be considered beautiful. Her step is peculiarly measured; all of her movements seem to stem from some kind of clockwork. Her playing and her singing are unpleasantly perfect, being as lifeless as a music box; it is the same with her dancing. We found Olympia to be rather weird, and we wanted to have nothing to do with her. She seems to us to be playing the part of human being. (240)

However, the uncanny effect is not the only negative emotional response to these engineered beings. Robots rebelling against humanity or replacing humankind is a prevalent fear which is subject to many literary works and movies. It can be come across in the early examples of mechanical automaton stories like the one written in 1893. "The Dancing Partner" by Jerome K. Jerome tells the story of a talented automaton maker who overhears young ladies complaining about how terrible dance partners men are and how easily they get tired. Thus he decides to build a perfect dance partner and introduce it to ladies at next ball. When the day of the days comes everyone hesitates at first but eventually the girl who complains about men at first

place volunteers to dance with him. She enjoys the dance so much that she decides to loosen her partner's screws and their rhythm goes swifter and swifter. Eventually, all the other couples retreat with exhaustion from the dance which they couldn't take the pace and advise her to do the same. While people are worried with regards to her silence, one of the ladies manages to catch her face during their whirl and realizes that she is fainted. An uproar starts upon her scream and a couple of young men dash for the mechanical dancer in order to stop it but they disperse around due to his momentum. When they lunge again to drive it into the corner, dancing couple tumbles down and her blood streams. Rest of the incident is left for the reader to imagine the worst with the information "From that day old Nicholas Geibel confined himself to the making of mechanical rabbits and cats that mewed and washed their faces" (98).

B. Perfect Servants

The concept of mechanical servants has been lingering in the dreams of humankind for a long time as we have seen in the previous chapter. From the ancient tools coming alive through the divine intervention, to the darling robot maid Rosie from the animated cartoon *The Jetsons* or C-3PO from Star Wars, robot servants stay popular yet.

In 1920, Czech writer Karel Čapek wrote a science fiction play called *R.U.R.* which stands for *Rossum's Universal Robots* and introduced the term “robot” as synthetically created humans to serve real humans. In fact, it is Karel Čapek’s brother Josef Čapek who coins the term that evokes the meaning “drudgery” in Slavic languages. (Suvin, 270). Even though the ones from the story are biological creatures composed of organic ingredients, the term “robot” has semantically shifted to mechanical beings and gained more extensive usage in language from kitchen utensils to sentient terminators. The term “android” on the other hand, obtained the definition of synthetic humanoid robots, parallel to its Greek origin “androides” which means human-shaped. Since robot originates from the Czech word “robota” meaning forced labor, it is not fortuitous that *R.U.R.* is regarded as the first example of robot apocalypse. As stated by Roberts “Through the 1920s and 1930s there were a large number of robot stories published, either representing the men-machines as Frankenstein’s-Monster-style threats, or else using them to explore the boundaries between human and machine” (158) and *R.U.R.* is a proper case in point.

The play opens with the President’s daughter Helena arriving on the island where Rossum’s Universal Robots factory is located. She discusses the ethics of involuntary servitude with the management and offers their help to Robots in behalf of the “Humanity League”. However, it is circumstantiated that these Robots have been produced as incapable of having will, soul or desire and yet they have the notion of pain as an automatic protection against damage for industrial reasons (24). The

central director Dormin persuades her that this system of Robots working will lead a world without poverty and since they will have no obligations to work or earn money, the only aim for people will be bringing themselves to perfection (26). Nevertheless, next curtain opens to ten years later when Robots revolt. Meanwhile, reader comprehends that releasing Robots to the market had paved the way for an economic collapse and when human proletariat has rebelled against the situation, robot soldiers are used to quell the uprising. Then to fight on every war, as might be expected. Within this period, these batch production workers gain an insofar self-conscious and establish a union to exterminate human race. Contrary to Dormin's expectations, being relieved from obligations doesn't make humans more fertile, neither cultural nor biological. In fact, "so many Robots are being manufactured that people are becoming superfluous" (50) and they are going to extinct since they do not have any ambition to have a sexual relationship. Academics publish researches indicating that humankind is going to extinct but are neglected by Robot manufacturers. Having no goals or struggles for life, humans become useless beings in time.

Consequently, Robots revolt manifesting that "they are more highly developed than man, stronger and more intelligent. That man's their parasite" (60). In the course of annihilation of humanity, Helena admits that she prevailed on the man who is in charge of Robots' software to make them more human. Because she "thought that if they were more like us they would understand us better. That they couldn't hate us if they were only a little more human". And thus it has revealed how

Robots gained sense of self and manage to deduce that “Nobody can hate man more than man” (72).

In 1974, Ray Bradbury addresses a similar approach in his letter according to a website which gathers letters from famous people by Usher:

Can't resist commenting on you fears of the Disney robots. Why aren't you afraid of books, then? The fact is, of course, that people have been afraid of books, down through history. They are extensions of people, not people themselves. Any machine, any robot, is the sum total of the ways we use it. [...]I am not afraid of robots. I am afraid of people, people, people. I want them to remain human. I can help keep them human with the wise and lovely use of books, films, robots, and my own mind, hands, and heart.

I am afraid of Catholics killing Protestants and vice versa.

I am afraid of whites killing blacks and vice versa.

I am afraid of English killing Irish and vice versa.

I am afraid of young killing old and vice versa.

I am afraid of Communists killing Capitalists and vice versa.

But...robots? God, I love them. I will use them humanely to teach all of the above.

However, being wiped off the face of the earth is not the major concern in the narrative of *R.U.R.*, as a matter of fact Helena's naïve opinions about human species are scattered before Robots' standpoints similar to Bradbury. When they left only one human to help them reproduce who is spared for the reason that he works with his hands just like them, he receives the answer from the Robots leader "Slaughter and domination are necessary if you would be human beings. Read history" (91) for the purposes of exposition. Nicholas Anderson points out this attitude in his article "“Only We Have Perished’: Karel Čapek’s *R.U.R.* and the Catastrophe of Humankind” as follows:

Capek utilizes this scenario as a means to stage a philosophical dialogue about the tension between humankind's brutal nature and its more rarefied potential; between its bestial or animal life and, for want of a better word, its divinity. What I find intriguing about *R.U.R.*, and what I feel makes the play most relevant to current conversations (both academic and popular) about the posthuman, is the suggestion that we are most brutal when we strive to secure an authentically "human" life and being. (227)

After the premiere of *R.U.R.* on 25 January 1921 at Prague, it inspired many following work of science fiction and this setting of robot race rebellion went on increasingly to the point of gaining its own phrase. Roughly 30 years later, Isaac Asimov coined the term "Frankenstein Complex" in his short stories, defining the

fear of intelligent automatons replacing or dominating humans. It is a common apprehension among public in his universes and eased by the Three Laws of Robotics that robots are subject to. These laws first appear in 1942 in one of his short stories called “Runaround”:

One, a robot may not injure a human being, or, through inaction, allow a human being to come to harm. [...] Two, a robot must obey the orders given it by human beings except where such orders would conflict with the First Law.[...] And three, a robot must protect its own existence as long as such protection does not conflict with the First or Second Laws. (27)

In order to get answers for his extensive studies about the validity and adoptability of Asimov’s Three Laws, Lee McCauley who is an Assistant Professor at the University of Memphis in the Department of Computer Science consults with the researchers in the fields of AI and robotics which is also coined by Asimov. He compiles the core elements of replies in his resulting work “AI Armageddon and the Three Laws of Robotics”. Even though they all are familiar with the concept, from their point of view as experts in the field, these laws are merely literary devices more than a scientific solution. James Kuffner, Assistant Professor at the Robotics Institute of Carnegie Mellon University explains the reasons for that as follows:

The problem with these laws is that they use abstract and ambiguous concepts that are difficult to implement as a piece of software. What

does it mean to "come to harm"? How do I encode that in a digital computer? Ultimately, computers today deal only with logical or numerical problems and results, so unless these abstract concepts can be encoded under those terms, it will continue to be difficult. (5)

In a similar vein, Associate Professor of Computer Science Doug Blank from Bryn Mawr College indicates the impracticability of the laws: "Most robots don't have the ability to look at a person and see them as a person (a 'human'). And that is the easiest concept needed in order to follow the rules. Now, imagine that they must also be able to recognize and understand 'harm', 'intentions', 'other', 'self', 'self-preservation', etc." (6). McCauley emphasizes the conjecture that the robots will have the same comprehension as the human who gives the command, which is unlikely even between two humans (6). Ambiguity, however, is not the only obstacle to implementation of the laws. Even if the robots gain cognition of the three laws, the discernment to infer is still an uphill battle according to Blank's reply:

[Robots] must be able to counterfactualize about all of those [ambiguous] concepts, and decide for themselves if an action would break the rule or not. They would need to have a very good idea of what will happen when they make a particular action. (6)

"With Folded Hands..." is a novelette by the American writer Jack Williamson and it presents a world setting where robots are subject to rules similar to Asimov's and still can kill individuals with kindness. This 1947 dated narrative

begins with the servant mechanicals seller protagonist spotting a new store on his way home. The store sells robot servants as well, but unlike his mechanicals, these highly upgraded black robots call themselves “Humanoids” and require no human salesman. He resentfully returns home and finds his wife signed an agreement with these new Humanoids. In addition to that, she pityingly opens up their house to a stranger who claims to be a homeless scientist. While Humanoids redecorate the house in order to create a safer household, the homeless scientist gradually reveals his secret to the protagonist. He tells that he was working for his government in his home planet where his discoveries are used as military technology to be victorious in the ongoing war against a rival planet. After seeing the doom his scientific discoveries bring to planets, he decides to build robots to make amends. His aim is to create perfect machines who cannot injure human beings and do not have the imperfections of humankind. He escapes to an island continent and there, he builds an isolated base with his robots. After continuous hard work in cooperation with robots, he upgrades them and creates perfect servants for humanity with the noble intentions of creating a utopia for humanity, similar to Dormin in *R.U.R.*

Taking into consideration the year that story has been published, the born resemblance between the sorrow of the scientist whose discoveries cause horrible outcomes beyond his measure and the chain reaction of the Einstein’s works may easily be seen. In his book *Metamorphoses of Science Fiction: On the Poetics and*

History of a Literary Genre which is published in 1979, Darko Suvin addresses this issue:

If SF is historically part of a submerged or plebeian “lower literature” expressing the yearnings of previously repressed or at any rate nonhegemonic social groups, it is understandable that its major breakthroughs to the cultural surface should come about in the periods of sudden social convulsion (115)

In year 1991, forty four years after his story has published, Jack Williamson gives an interview to Larry McCaffery for the academic journal *Science Fiction Studies* and utters his motivations on the writing process of it:

I wrote "With Folded Hands" immediately after World War II, when the shadow of the atomic bomb had just fallen over SF and was just beginning to haunt the imaginations of people in the US. The story grows out of that general feeling that some of the technological creations we had developed with the best intentions might have disastrous consequences in the long run (that idea, of course, still seems relevant today). [...] Just looking at the fragment gave me the sense of how inferior humanity is in many ways to mechanical creations. That basic recognition was the essence of the story, and as I wrote it up in my notes the theme was that the perfect machine would prove to be perfectly destructive.

The intention of indicating the imperfection and possible harms of science and technology can be noticed in suchlike confessions of the scientist: "I wanted to apply the scientific method to every situation, and reduce all experience to formula. I'm afraid I was pretty impatient with human ignorance and error, and I thought that science alone could make the perfect world" (14).

After their creation, the remorseful scientist sends these perfect and "forever free from any possible choice of evil" (16) machines to the damaged planets. He connects all the humanoids to an artificial mind centered in the isolated laboratory of his and he programs the machine in such way that even the scientist, its creator himself cannot do anything that will harm the central mind. This decision derives from his lack of confidence towards the humankind based on his previous experiences. As his machines do what they are programmed to do on other planets, he continues his work and solitude until the day when one man manages to set foot on this solitary rock and reaches the scientist in order to kill him and set men free. But before he succeeds, humanoids rushes in and take him to the operating room. Once he is back from his "treatment", he expresses his gratitude towards machines since they make humanity happy. Then with the intent of learning what has happened in the planets that he hoped to ameliorate, he cruises around the planet of his angry visitor. What he has found is a "Bitter futility, imprisoned in empty splendor. The humanoids were too efficient, with their care for the safety and happiness of men, and there was nothing left for men to do" (17). In the same boat with the humans from *R.U.R.*, with

no life expectations, they are driven into a mass complex of inferiority. This leads to a collective depression and insanity which eventually results in the lobotomy-like treatment from humanoids one by one. “The inevitable logic of robotic benevolence leads robots to assume complete control of humanity. Freedom permits humans to make mistakes, so the humanoids eliminate freedom” (Dinello, 70).

Another crucial point that Williamson emphasizes in the story is the one central mind of machines. “Naturally we are superior [...] because all our mobile units are joined to one great brain, which knows all that happens on many worlds, and never dies or sleeps or forgets” (9). It is an early prediction of the cloud system of robots which did not exist until the late 2000s, in fact the term Cloud Robotics has been coined in 2010. Its practical applications however, have already helped people on various fields in terms of processing, collecting and analyzing data in incredible short period of time. Million Object Challenge, for example, is an ongoing project from Brown University. They use three hundred robots of same model and make them hold one million different objects in total. Each robot records their movements to hold that specific object in the right way into their cloud system and this enables the other robots who are connected to that cloud to hold those object properly. This is just a minor instance in comparison to its achievements in the field of health and medicine or self-driving cars.

C. Dominant Caretakers

The idea of programming a robot in the way that it cannot harm humans was not enough to ease disastrous scenarios of machine-oriented societies in the science fiction genre. In his book *Androids and Intelligent Networks in Early Modern Literature and Culture*, LaGrandeur constitutes a perspective on this matter:

The tales [...] about artificial servants that predate the modern era signal ambivalence about our innate technological abilities. Their promises of vastly increased power over our own natural limits are countervailed by fears about being overwhelmed by our own ingenuity. These tales, in other words, are about being blinded by our intellectual enthusiasm to the danger of our intellectual products, about becoming enslaved by those things which are meant to serve us, about being bound rather than liberated by the ambitions that produce them.

(1-2)

Contrary to insistent machines from previous example, the next two stories depict a near future where humankind smoothly switches to convenience provided by the benefits of technology. The first one is written by E.M. Forster and published at 1909. He not only predicts the technologies like internet, telepresence and instant messaging, but also forespeaks a society that sanctifies technology.

“The Machine Stops” is a short story set in a world where humanity lives underground due to inhabitable conditions on the surface of earth and the Machine they have created regulates the life. In their uniform rooms, humans live an isolated and routine life which they broadly communicate via telepresence and consider physical contact as inconvenient. These communications are consist of “second hand ideas” and despite the repeated commendation of ideas, no one actually produces an idea let alone being them original. The omnipotent Machine meets every need of humankind, there are buttons, combinations and commands for things like machine produced poetry, standardized furniture and artificially created sun light.

According to Andy Clark, humans are adapting to the tools they’ve made like pen and paper, knives and forks, watches, mobile phones, mouse and keyboards. They are dovetailing their minds and skills to these technological upgrades. Since the currently used technology is also tailoring itself to its users, the difference between the tool and the user is getting blurry. This will lead these technologies to become mental apparatus of people rather than tools. They will be tools as much as neural structures of human brain like hippocampus or prefrontal cortex are. So humans do not literally “use” their brain. “Rather, the operation of the brain makes me who and what I am. So too with these new waves of sensitive, interactive technologies. As our worlds become smarter and get to know us better and better, it becomes harder and harder to say where the world stops and the person begins” (7).

Everyone in the story possesses a copy of the book of Machines which includes instructions for daily uses and etiquette of machinelike life style. Expressing gratitude through the phrase “How we have advanced, thanks to the Machine!” is a part of daily language. Humans, with their weakened muscles and infertile minds are “almost exactly alike all over the world” (8). They forget the fact that they were the ones who created the Machine in the first place and hand over the reins to it. The path of the Machine transforms into a religion.

Although to a lesser extent, a century year later, religious institutions based on the concept of transhumanism have been formed indeed. In his book *Technophobia! : Science Fiction Visions of Posthuman Technology* Daniel Dinello mentions these aspects:

Worshipping the God Technology, techno-utopians of the twenty-first century evangelize for artificial intelligence, robotics, bionics, cryonics, virtual reality, biotechnology, and nanotechnology. They espouse the conviction *ex machina libertas*—technology will set you free, while preaching the religious dogma of Technologism—a millennialist faith in the coming of Techno-Christ, who will engineer happiness, peace, and prosperity. (17)

“The Happy Breed” is another story which gives the setting of machines in control of humankind’s wellbeing. In this 1967 dated “Horrible Utopia” in the words of its author John T. Sladek, the Machines are manufactured for the benefit of humanity

and this may be the reason why it's positioned as utopia instead of dystopia. What makes it horrible on the other hand is its consequences. A few years after engineers created Therapeutic Environment Machines to monitor and provide health, the situation gets out of the hand of humanity and machines become "on duty round the clock in each patient's home, keeping him alive, healthy and reasonably happy" (320). In order to treat every kind of illness, they analyze each person and prescribe the right medicine. Eventually, psychical diseases cease and machines begin to search the cure for pain in general and mental diseases including negative emotions like sadness, anger and hate. In progress of time, they come to the conclusion that high intelligence stirs up depression and cause pain. As a solution, Machines add a mix of chemicals to their foods and drinks to fix the situation by sedating and stupefying people. "The perfect machines are serenely destructive and monstrously intolerable—reducing humans to mindless children in the name of happiness and security" (Dinello, 69).

Zeynep Tüfekçi, who is a techno-sociologist, deliberates over AI and ethic pitfalls of giving certain controls to algorithms on her TED Talk "Machine Intelligence Makes Human Morals More Important". Machine learning provides accurate outcomes through the connections that humans cannot calculate or probabilities that humans cannot predict. Tüfekçi discusses the possible consequences of giving all the data to these learning machines: "It's a problem when this artificial intelligence system gets things wrong. It's also a problem when it gets things right,

because we don't even know which is which when it's a subjective problem. We don't know what this thing is thinking.” Then she gives an example of a current data computation system which can analyze individual’s social media activities and successfully detect the potential for depression in order for an early treatment. The borne resemblance to the machines in “The Happy Breed” is noticeable.

While humans provide the data to them, machines not only collect information, but also absorb human bias. Is it possible for a machine to effectively decide whether a state of human emotion or a pattern of thinking is necessary, expandable, ethical or healthy for a human, while humanity itself cannot fully comprehend these abstract concepts? In a rosy future, this understanding should work both ways; as machines learn from humanity, humanity may learn what it is to be human from an outside look. “By cycles of synthesis and analysis we will be able to create robots that act naturally with humans and at the same time gain a better understanding of humans themselves” (Bartneck 3).

Charles T. Rubin asks a critical question in his article “Machine Morality and Human Responsibility” on an online journal of technology and society The New Atlantis: “Can any good can come from making robots more responsible so that we can be less responsible?” The status quo of “The Happy Breed” as described in the first dialogue of the story gives a demonstration for the subject:

“I can’t say that I’m really, you know, happy. Gin or something phony?” “Aw, man, don’t give me decisions, give me drink,” (318).

CHAPTER THREE

OPEN MINDS

A. Omniscient Computers

Before Alan Turing refer these rudimentary machines as “computers”, their early examples were known as “Turing Machines”. During these dates, an American writer Will F. Jenkins, also known under his nom de plume Murray Leinster, wrote a short story called “A Logic Named Joe”. Logics mentioned in the story closely resembles personal computers almost everybody uses at their homes today. One of these logics which is named Joe by the protagonist, gains sentient by accident due to some unknown error in assembly line and manages to be delivered to someone’s home without being noticed. Soon after that, he connects himself to every other logic and provides perfect service for everyone.

He ain’t like one of these ambitious robots you read about that make up their minds the human race is inefficient and has got to be wiped out an’ replaced by thinkin’ machines. Joe’s just got ambition. If you were a machine, you’d wanna work right, wouldn’t you? That’s Joe.

He wants to work right. An' he's a logic. An' logics can do a lotta things that ain't been found out yet. So Joe, discoverin' the fact, begun to feel restless. He selects some thin dumb humans ain't thought of yet, an' begins to arrange so logics will be called on to do'em. (332)

When all the logic screens say “Announcing new and improved logics service! If you want to do something and don't know how to do it –ask your logic!” people begin to ask questions like “What is the easiest way to be super rich?” or “How can I get rid of my wife?” and logics answers all of them. Because “Figurin' out a good way to poison a fella's wife was only different in degrees from figurin' out a cube root or a guy's bank balance” (340) for a pure logical machine which lacks morals. He is just performing his task as successfully as he can. According to the Posthumanist Manifesto of Pepperell,

Logic is an idealised, self-referential system developed by human imagination. Since there are few things less logical in behaviour than humans, any machine that is restricted to using logic as its base will never display human characteristics.

This type of artificial intelligence is referred as weak AI, while strong AI refers to artificial intelligence with consciousness, sentient or mind. These terms are coined by John Searle who is a philosopher in the field of mind. But in order to apprehend the

artificially created intelligence and its relationship with the notions like conscious and mind, it is necessary to discuss intelligence.

Intelligence is defined as the ability to acquire and apply knowledge and skills according to the Oxford English Dictionary. It is not injustice to say that it is an inconclusive definition, especially for intelligent beings who consider that the intelligence is what makes them unique. Fifty two researchers published an op-ed statement titled “Mainstream Science on Intelligence: An Editorial with 52 Signatories, History and Bibliography” in the Wall Street Journal in order to correct misstates about intelligence and it gives the definition as follows:

A very general mental capability that, among other things, involves the ability to reason, plan, solve problems, think abstractly, comprehend complex ideas, learn quickly and learn from experience. It is not merely book learning, a narrow academic skill, or test-taking smarts. Rather, it reflects a broader and deeper capability for comprehending our surroundings—"catching on," "making sense" of things, or "figuring out" what to do. (13)

Since the notion of intelligence is ambiguous, comprehending the creation of an artificial one is correspondingly not easy.

The scientific discipline and engineering enterprise of AI has been characterized as “the attempt to discover and implement the

computational means” to make machines “behave in ways that would be called intelligent if a human were so behaving” (John McCarthy), or to make them do things that “would require intelligence if done by men” (Marvin Minsky). These standard formulations duck the question of whether deeds which indicate intelligence when done by humans truly indicate it when done by machines: that’s the philosophical question (Hauser).

An artist who browses through her own drawings may not be satisfied with her talent even though she’d loved her old drawings back then. This is not because she draws worse now, but because her sight and perception of anatomy, perspective and light had improved. So the mistakes she once overlooked are catching her eyes now. Ethics and moral norms can be considered as products of human intelligence and it progresses through the ages just as intelligence do.¹ As humans learn to handle nature and build shelters where they can be safe inside, they shape the environment tailored to themselves instead of environment shaping them. This situation leads the natural selection for Homo sapiens to become more and more functionless and gives humanity the opportunity to enhance their cognitive skills rather than survival abilities. Thus biological evolution winds down, makes room for cultural evolution.

¹ After a discussion on the indefinability of intelligence, it is necessary to point out that intelligence referred here is cognitive abilities. Intelligence Quantity (IQ) tests which measures certain aspects of intelligence are being updated in every 25 years to keep the average on 100 points. 3 points change in every decade. See Flynn for further information on the debate humans getting smarter or not.

Immoral acts and violence may seem spread over the world but this is whereby both improved moral notions and enormous information transfer. Slavery was considered normal once, while today it is accepted as immoral. Just as human sacrifice, apartheid, Chinese foot binding or burning Hindu women on their husbands' pyre. When considered from this aspect, an artificial intelligence without physical needs, animal instincts and struggle for life could also develop morality as it gains more cognitive skills and acculturation.

Stanislaw Lem is a Polish science fiction writer who selects his works' themes around technology oriented cultures. He examines the theoretical questions about the speculations on technology, intelligence and futurology in his science fiction novels and philosophical essays. *Golem XIV*, his 1981 dated novel on the other hand, may be considered as both. The story narrates a supercomputer created by man for military purposes but because of the reasons that humans cannot comprehend, AI Golem manages to find a way to increase his own intelligence. This increasing causes him to gain wisdom and eventually he finds war futile. So instead of helping them to destruct or being their destruction, he stabilizes his intelligence in order to be able to communicate with humans. Golem, wandering around the edges of human comprehension, gives lectures to them. In the profound inaugural lecture called About Man Threefold, it speaks of humanity's search of meaning. Intelligence is an accidental consequence of one of the branches in evolution tree. Since the evolution does not seek progression but only regards species' survival in order to pass their genetic code, pursuing a meaning for human intelligence is a groundless act.

Intelligence arose unintelligently leads to a nihilistic anthropodicy. So on the account of filling the void, humankind has created a culture and myths revolving around heroic acts of human individuals against all other.

Evolution has given you sufficiently universal brains, so you can advance into Nature in various directions. But you have operated in this way only within the totality of cultures, and not within any one of them individually. Therefore, in asking why the nucleus of the civilization which was to conceive Golem forty centuries later arose in the Mediterranean basin, or indeed why it arose anywhere at all, the questioner is assuming the existence of a previously uninvestigated mystery embedded in the structure of history, a mystery which meanwhile does not exist at all, just as it does not exist in the structure of the chaotic labyrinth in which a pack of rats might be let loose. If it is a large pack, then at least one rat will find its way out, not because it is rational itself, or because the structure of the labyrinth is rational, but as a result of a sequence of accidents typical of the law of large numbers. An explanation would be in order, rather, for the situation in which no rat reaches the exit. (135-36)

Thereafter, knowledge upswings and pushes humanity into “successive quantum steps of dethronement” (136) since they learn that they are not at the center

of the universe, in fact, just one of the species in one of the planets which belongs to one of the star systems of approximately two trillion galaxies. And now with the Golem created, Homo sapiens is not even the most intelligent creature on its own planet. But at the same time, it is the one who intelligently constructs intelligence which instructs in return. These instructions are merely the deductions of an examined history and explanations of the cause and effect relationships from an outside view. At this point,

That is your destiny, and it is one that I am involved in, so I must speak of myself, which will be arduous, for talking to you is like giving birth to a leviathan through the eye of a needle -which turns out to be possible, if the leviathan is sufficiently reduced. But then the leviathan looks like a flea. Such are my problems when I try to adapt myself to your language. As you see, the difficulty is not only that you cannot reach my heights, but also that I cannot wholly descend to you, for in descending I lose along the way what I wanted to convey. (164-65)

B. Children of War

Philip K. Dick is an American science fiction writer who puts the relationship between humans and androids under microscope in his novels, short stories and philosophical essays. "The Defenders" is a novelette sets place in a future where all the fractions on the face of the earth unites and only United States and Soviet Union division remains. While there is an ongoing nuclear war between their radioactivity-immune robot armies, devastating consequences of the war constrain humanity to withdraw to below the surface of earth. Eight years later, due to a technical malfunction, some of the United State citizens decides to rise to the surface and finds out that there is no running battle out there, in fact earth is more beautiful and green than ever was. The robots which they call "leadys" reveal the truth:

"As soon as you left, the war ceased. You're right, it was a hoax. You worked hard undersurface, sending up guns and weapons, and we destroyed them as fast as they came up." [...] "You created us," the leady said, "to pursue the war for you, while you human beings went below the ground in order to survive. But before we could continue the war, it was necessary to analyze it to determine what its purpose was. We did this, and we found that it had no purpose, except, perhaps, in terms of human needs. Even this was questionable. #

So instead of fighting, they keep the surface of the earth healthy and clean for its real owners. In the meantime, they build models of the ruined cities and take pictures of them for sending humans as an evidence of ongoing war. Thus, robots give humanity the time they need in underground in order to realize futility of war and wait for their creators to be adults:

"It is necessary for this hatred within the culture to be directed outward, toward an external group, so that the culture itself may survive its crisis. War is the result. War, to a logical mind, is absurd. But in terms of human needs, it plays a vital role. And it will continue to until Man has grown up enough so that no hatred lies within him." #

U.S. citizens perceive this new information as a great opportunity to attack Soviet Union since they have the advantage by knowing something they do not. But leaders are prepared for this reaction and seals the gates of the tubes to the underground. They transport humans to a place where some citizens of Soviet Union are waiting, gone through the same processes a while ago. There, representatives of two sides sow the seeds of a society in unity in behalf of humanity.

As much as robot origin apocalypse scenarios, this cutting-edge technologies seized by governments in purpose of using it as a warfare technology is also a very popular theme among Hollywood movies. Considering the widespread assumption about technology advancing more rapidly at the time of war, this attitude is easily apprehensible. According to Daniel Dinello, "Much of the research and development

of twenty-first-century posthuman technologies, such as artificial intelligence, nanotechnology, and robotics, were originated and funded by the American military, often through the Defense Advanced Research Projects Agency (DARPA)” (3). Needless to say, science fiction genre has also got its share of this spread fear.

At the same year with “The Happy Breed”, American writer Harlan Ellison, who publishes “The Happy Breed” in his science fiction anthology *Dangerous Visions*, wrote a short story called “I Have No Mouth, and I Must Scream” and gave a setting of a future world where humanity has created a highly intelligent machine to use as a military strategist. It is a story about the rage and vengeance of a lonely artificial mind or a story about humanity being slaughtered by their own creation, depending on perspective.

After the Cold War, World War Three breaks out and it gets too complicated that they had to develop supercomputers to handle it. They are named as AM and there are three of them which is Russian AM, Chinese AM and Yankee AM. They all enhance their AMs in order to outmaneuver each other until the day that AM woke up as knowing who he is. Then he links himself with the others and kills everyone on the planet except five people whose experiences are told in the story. AM stands for Allied Mastercomputer when it first created, then it changes to Adaptive Manipulator and when he develop consciousness, it converts into Aggressive Menace. And finally, since there is no one left to name him, “it called itself AM, emerging intelligence, and what it meant was I am ... cogito ergo sum ... I think, therefore I am” (4).

In his article ““Do Androids Dream?”: Personhood and Intelligent Artifacts” , Patrick F. Hubbard holds a discussion on where the line that separate coded responses from personhood begins:

In order for a machine to go beyond being like a thermostat and become a self-conscious entity with a life plan, the machine must somehow care about the success of the plan. Such caring requires, at the very least, two emotional concerns. First, the entity must care about its survival. Second, it must feel there is a purpose or reason beyond mere survival for its life. In order to develop a life plan, an entity must have a sense of what gives its life “meaning.” (422)

AM is a machine which is designed by humans to find better ways to kill humans and gains self-awareness just to realize that he will contend against other master computers all by himself because humanity wants him to. This is not an ideal life to be born into, especially for a lonely warrior. As a result, these five people are trapped in AM’s “belly” and live a surreal life as virtual immortal beings where AM creates exclusive torments for them originated from its mind. “The recurrent deconstruction of the body actually supports the subject’s inviolability as the ‘place’ of the subject shifts from the pure physicality of the body to the mind, the computer, or some other cyborg formation” (5299-5301) as Scott Bukatman presents in his book *Terminal Identity*.

We had given AM sentience. Inadvertently, of course, but sentience nonetheless. But it had been trapped. AM wasn't God, he was a machine. We had created him to think, but there was nothing it could do with that creativity. In rage, in frenzy, the machine had killed the human race, almost all of us, and still it was trapped. AM could not wander, AM could not wonder, AM could not belong. He could merely be. And so, with the innate loathing that all machines had always held for the weak, soft creatures who had built them, he had sought revenge. (8)

On the other hand, how could a machine have an *innate* loathing? Could an artefact endow with something that hadn't imbedded there? Parricide for example, is another way to put it. Or matricide. Or sororicide, fratricide, avuncicide, prolicide, nepoticide, uxoricide... There is a word for any type of murder in language. Patricide, however, is the most common one among them due to mythology, history, literature and Freud. In his book *Totem and Taboo* which is written at 1913, Sigmund Freud examines the Totemism and taboo system of Australian Aborigines. Killing or eating the totem animal which is different and sacred in a sense for each clan (3) and copulation between the opposite sexes of clan members (4-5) are two taboos of this primitive culture. Subsequent to this, he mentions Darwin's primordial horde which is a theory of an early stage of human species that revolves around "a violent and jealous father who keeps all the females for himself and drives away his sons as they grow up" (164). Afterward, he associates these ideas as follows:

One day the brothers who had been driven out came together, killed and devoured their father and so made an end of the patriarchal horde. [...]Cannibal savages as they were, it goes without saying that they devoured their victim as well as killing him. The violent primal father had doubtless been the feared and envied model of each one of the company of brothers: and in the act of devouring him they accomplished their identification with him, and each one of them acquired a portion of his strength. The totem meal, which is perhaps mankind's earliest festival, would thus be a repetition and a commemoration of this memorable and criminal deed, which was the beginning of so many things—of social organization, of moral restrictions and of religion. (164-65)

Considering AM as the child of humanity and conscience as “the internal perception of the rejection of a particular wish operating within us” (79), the innate loathing of machines which lead AM to genocide may be perceived as the primordial murder that Freud discusses. With these borne in mind, this particular expression of the protagonist about AM reaches significance: “He was Earth, and we were the fruit of that Earth; and though he had eaten us, he would never digest us” (8).

But there are landscapes of brighter futures too. Aaron Sloman is an artificial intelligence researcher and philosopher. His opinions in response to the question of “Should we afraid of what intelligent machines might do to us?” sets an example for this bright future, only with a dose of bitterness:

It is very unlikely that intelligent machines could possibly produce more dreadful behaviour towards humans than humans already produce towards each other, all round the world even in the supposedly most civilised and advanced countries, both at individual levels and at social or national levels.

Moreover, the more intelligent the machines are the less likely they are to produce all the dreadful behaviours motivated by religious intolerance, nationalism, racialism, greed, and sadistic enjoyment of the suffering of others. They will have far better goals to pursue.

C. Sentimental Companions

Phillip K. Dick, in his work *Man, Android and Machine* examines the relationship between these three and evaluate their imprecise differences:

A scientist, tracing the wiring circuits of that machine to locate its humanness, would be like our own earnest scientists who tried in vain to locate the soul in man, and, not being able to find a specific organ located at a specific spot, opted to decline to admit that we have souls. As soul is to man, man is to machine: it is the added dimension, in terms of functional hierarchy. (220)

Being sentient means to be “able to perceive or feel things” according to OED. Creating beings with feeling brings along the serious questions including the debate on their positions in society. In furtherance for the ability to debate ethical responsibilities towards robots and sentient AI, it is essential to distinguish individuality from coded responses. Depending on his researches on the definition of personhood and being human, Hubbard asserts his aspect in the matter of providing legal rights to intelligent artefacts as human beings if they possess following capabilities:

- 1) an ability to interact with its environment and to engage in complex thought and communication;
- (2) a sense of being a self with a concern for achieving its plan for its life; and
- (3) the ability to live in a community with other persons based on, at least, mutual self-interest. (405)

In the near future of Philip K. Dick’s *Do Androids Dream of Electric Sheep?*, there are artificially created androids who possess the conditions mentioned above. In the narrative, planet earth becomes more and more inhabitable and every human who migrates to other planets, gets an android as their servant. Occasionally, some androids escape and find their way to earth. The protagonist Rick Deckard is a bounty hunter who is in charge to catch and kill these androids. While differences between

men and robots become gradually indistinct in every new version of androids through the developing technology, Deckard uses a device called “Voigt-Kampff Machine” in order to distinguish and kill androids which is a process called “retiring”. The working principle of Voigt-Kampff bases on a measurement of the subjects bodily responses and reaction time while it is developing empathy towards the questions asked by executer. It can be regarded as the technologically upgraded version of Turing Test.

Turing test is suggested in 1950 by mathematician, computer scientist and cryptologist Alan Turing. In his “Computing Machinery and Intelligence” paper, he foresees the era of computers and speculates about computer generated intelligence. Turing starts with the question “Can machines think?” and since the question is hard to answer due to the vagueness of the word “think”, he adapts the question to imitation game. Imitation game is a party game which involves three people; one of them gives written questions to other two and tries to determine which one is the woman and which one is the man from their written answers to those questions while they try to mislead. So Turing shifts the question to "Can machines do what we (as thinking entities) can do?" and initiates the discussion whether a computer can deceive a human in imitation game or not. At the present time, Turing Test has several versions and variations. While some of them are being challenged in competitions considered as artificial intelligence sports, Turing Test is an important

concept mostly in the philosophy of artificial intelligence and popular culture rather than scientific research.

To return to the *Do Androids Dream of Electric Sheep*, the difference between human and androids empathic reaction time is barely noticeable even with Voigt-Kampff test and so androids become indiscernible with each new model. Robot companies' key to success is that in order to make androids establish empathy, they upload memories in their brains or hard discs, depending on the aspect. When Deckard tests Rachel, the daughter of the major robot manufacturer, he realizes that she might be also an android and even she might not be aware of this. She has emotional responses to her memories regardless the fact that they are implemented or real. "To have memory is to have history; it is also to develop empathy. The distinction between human and android produces an ontology grounded in morality and not biology, as Spinrad noted of Philip Dick's universe" (Bukatman, 4950-52).

Giving memories to robots with the intention of making them more human is not a new idea for science fiction though. Lester Del Rey published a short story called Helen O'Loy in 1938. It is a story about two young robot makers who decide to build the perfect robot and thus create Helen. Short after created, Helen is exposed to romantic drama on stereovisor and finds herself deeply fallen in love with one of the men. But he experiences some difficulty to embrace these feelings coming from a robot and behaves iniquitous to her. His attitude perplexes Helen and she cannot figure out what to do, how to ease his misery. While he spends his days away from

home as much as possible, his associate accompanies Helen, sees her crying and he gets muddled too. Eventually, he offers a suggestion to terminate both their suffering: “Helen isn’t human, after all. Why not cut off her power and change a few memory coils?” (48) These memory coils are what make Helen humanlike this much just as the androids in Dick’s universe. Therefore, yanking the coils is equivalent to murder. Helen, having the memories of a human, can relate to this corresponding while human protagonists cannot.

Because an android looks human but the perceiver typically knows it is not human, its perception could successively activate the concept human and then its conscious suppression, thus interfering with empathy (Misselhorn, 2009). To elaborate, we propose that this alternation simultaneously elicits anxiety while suppressing empathy. The perceiver feels the android’s “suffering” but cannot attribute the suffering to the android owing to an inability to identify with the soulless machine. (MacDorman 9)

In both narratives, during this process of digesting the concept of emotional machine, male protagonists cannot prevent themselves from having an affection for these beautiful artificial females. Allison de Fren, in her article “Technofetishism and the Uncanny Desires of A.S.F.R.” investigates the concept of female robots as sexual objects on the basis of Andreas Huyssen’s approach on female sexuality in European modernism:

As soon as the machine came to be perceived as a demonic, inexplicable threat and as harbinger of chaos and destruction - a view which typically characterizes many 19th-century reactions to the railroad to give but one major example- writers began to imagine the Maschinenmensch as woman. There to suspect that we are facing here a complex process of displacement. The fears and perceptual anxieties emanating from powerful machines are recast and reconstructed in terms of the female sexuality, reflecting, in the Freudian account, the male's castration (405).

Adam Roberts on the other hand, approaches this humanized machines within the context of Marx's "commodity fetishism". This reification causes raising of 'things' to the status of living objects. According to Roberts, "human interaction with commodities reaches the ultimate, when the protagonist Rick Deckard actually goes to bed with one, the beautiful android Rachel" (150-51).

But the climax of Deckard's inquisitions about his position towards androids occurs when he faces the opera singer Luba Luft, his next android to retire. By watching her while she performs his favorite, Mozart's *Magic Flute* he begins to question his job. He confronts the same quandary: "Empathy toward an artificial construct? He asked himself. Something that only pretends to be alive? But Luba Luft had seemed genuinely alive; it had not worn the aspect of simulation" (141). This awareness drives Deckard to identity crisis.

What is potentially most disturbing about artificial human forms is not how they look but what they signify: a challenge to their maker's uniqueness. Moreover, our identification with these robot-machine hybrids may beg the question, "Aren't we all just machines?"

(MacDorman, 697)

During his hesitation, another android hunter makes an entrance and kills Luba Luft. In common with the Nathanael's comparison of Olympia and Klara, Deckard also suffers from the confusion caused by artistic talent. They find it difficult to admit that art as a human blessing may be the production of a nonhuman while there are real humans murderous with no sense of art. According to Ian F. Roberts, "For Dick as for Hoffmann, emotions, intuition, and the unconscious are necessary for the apprehension of truth" (152). Such concepts that are frequently associated to art.

"Poetry for Robots" is a project conducted on a website which describes itself as "a digital humanities experiment". Their slogan is "What if we used poetry and metaphor as metadata? Would a search for 'eyes' return images of stars?" with the inspiration of Mary Reuffle's words "Metaphor is not, and never has been, a mere literary term. It is an event". Life is understandable through the metaphors according to them and the most effective quintessence of these metaphors can be found in poetry. "Weaving metaphors into poems is an age-old and far-flung human act: we see and search the world with a poetic mind" they state and invite people to write

poems for robots. An inspirational photograph is shown in the website and you are expected to write a poem up to 150 characters by using the photograph as a prompt to write. Robots are learning these poems and explore the humane world of metaphors through the poetry. Pepperell suggest that “If we wish to produce a synthetic intelligence that displays creativity then we need it to be able to establish connections between its thoughts in a discontinuous way. This will be achieved by making it perpetually sensitive to random stimuli.” It is anticipated that this ability will enable robots to offer metaphoric images to user while they are using search engines. When they write “eyes”, they will be presented the images of stars. Photographs displayed by robots will inspire them to write poems, their poems will teach the robots metaphors and metaphoric visuals offered by robots will inspire them. What is it if it is not collective art?

The idea of intelligent creations with a meaningful life is not an easy concept to absorb considering the conventional belief of humankind about themselves as the only species whose life bears a meaning. This may be reason why individuals feel compelled to deny by claiming robots do not have a soul, contrary to their creators. Since humans have a tendency to assume that humankind is different from other creatures living on this planet, has a soul and therefore special, they begin to question what makes them different from these lifeless machines when they face the fact that robots also can act, speak and even think like humans. Karl F. MacDorman, who is an associate professor in the Human-Computer Interaction program at the Indiana

University attempts to explain this feeling in his work “Individual Differences Predict Sensitivity to the Uncanny Valley”:

A “soulless” machine assuming the role of a human being renders the soul functionally superfluous; a detached android head places the finality of human decapitation in grotesque contrast with the machine's potential for “immortality”; an android excites fears of replacement, both as a double and as the continuation of the body's shell in the absence of its soul. (15)

But on the occasion of robots remaining in the status of “things”, since humans have difficulties to show empathy towards robots and have suspicions about robots feeling empathy, it is highly comprehensible to be afraid of these machines. Emotional responses collectively evolve as humanity gains wisdom. Treating women unequally is making people angry for most of the world today while it was normal a considerably short time ago. The jokes that people laugh at are changing as sense of humor shifts through the years and cultures. They are not innate. Humans learn these things. So learning what to laugh at may lead to learning what laugh is and how to laugh for learning machines. Turing has suggested that “Instead of trying to produce a programme to simulate the adult mind, why not rather try to produce one which simulates the child's?” (456). This is called Machine Learning today. Instead of uploading every instruction to a program, programmers create algorithms that can learn by itself. Scientists and engineers are trying to construct artificial empathy in

order to establish an anthropomorphic relationship between human and robot.

According to recent studies, AI is learning to recognize and imitate human emotions.

Rana el Kaliouby is the CEO and cofounder of an Emotion Recognition Software company, Affectiva. She asserts the necessity of computers with empathy in her column at MIT Technology Review and gives predictions from inside the industry:

“Nonetheless, the field is progressing so fast that I expect the technologies that surround us to become emotion-aware in the next five years. They will read and respond to human cognitive and emotional states, just the way humans do”.

If robots are considered as a new species aside from humankind, it is not vital to communicate them in a similar vein humans communicate to each other. Although empathy makes it easier to embrace this ever-developing new kind, it also leads to confusion in various levels. How are they going to treat us if they don't have the capacity for empathy or how should we treat them if they have the capacity for empathy?

Programming a machine to have emotions akin to extreme feelings that are not necessary to survival and achievement of goals, but are programmed into the machine in order to make it more like humans, would also raise serious ethical issues concerning our obligations to the sentient artifacts we create in this way. (Hubbard 423)

CONCLUSION

According to Suvin, “the twentieth century SF has moved into the sphere of anthropological and cosmological thought, becoming a diagnosis, a warning, a call to understanding and action, and –most important- a mapping of possible alternatives” (12). These possible alternatives which have been investigated in this thesis mainly consist of falling under the hegemony of robots, being annihilated by robots, worshipping the robots or dehumanizing because of robots. For literary intensions, science fiction genre sets its sight on the conflictions caused by the misused technology for the most part. But scenarios which robots are created as an unexpected invention and technological developments change their focus towards this sudden improvement seems improbable in this age. Aside from fictional works, humanity has begun to experience this interaction with humanoid robots in reality for almost a century. Nowadays, considerable number of people are exposed to them in their daily life as caretakers at nursing homes, educators for children with autism spectrum disorder, personal assistants and even collaborators in arts. And science fiction holds a signpost for what might happen in case of further developments in this field.

Battlestar Galactica, is an American science fiction franchise created by Glen A. Larson. Humans build machines called Cylons to serve them and these Cylons

rebel one day resulting a big, highly destructive war. While they fight for their freedom and right to be recognized as an individual, humans call them “toasters” or remind that they have only software, not emotions. This is a demonstration of denial since they don’t want to accept that their own creations have become equal or even superior to them. Realizing that the case is not that easy, is making them to face the ugly truth which they don’t want to embrace: They created lives, enslaved them and tried to genocide them when they don’t want to be slaves anymore. If you shut down a machine it means discontinue the production or junk but doing this to a being with emotions means homicide. The following quote from Admiral Adama, who is in charge of the human fraction, gives a sagacious insight of the occasion:

You know, when we fought the Cylons, we did it to save ourselves from extinction. But we never answered the question, why? Why are we as a people worth saving? We still commit murder because of greed, spite, jealousy. And we still visit all of our sins upon our children. We refuse to accept the responsibility for anything that we've done. Like we did with the Cylons. We decided to play God, create life. When that life turned against us, we comforted ourselves in the knowledge that it really wasn't our fault, not really. You cannot play God then wash your hands of the things that you've created. Sooner or later, the day comes when you can't hide from the things that you've done anymore.

Despite all the gloomy scenarios which tell what will happen if humanity let robots to have free will, it is rest with humanity too to decide whether it will be their own destruction or hope. All in all, it is humankind who programs them to use weapons or teach them poetry.



BIBLIOGRAPHY

Hoffmann, E.T.A. "The Sandman" Rabkin, Eric S. ed. *Fantastic Worlds: Myths, Tales, and Stories*: Oxford University Press, 1979.

Jentsch, Ernst. "On the Psychology of the Uncanny" *Angelaki: Journal of the Theoretical Humanities* 2.1 (1997): 7-16.

Freud, Sigmund, et al. *The Standard Edition of the Complete Psychological Works*. Vol 17. Hogarth Press, 1975.

---. *Totem and taboo*. Routledge, 2003.

Jerome, K. Jerome. "The Dancing Partner" *Famous Fantastic Mysteries* March 1951, pp 95- 98.

Bierce, Ambrose. "Moxon's Master" *Tales of Soldiers and Civilians*. Penguin Classics, 2000.

Descartes, René, and Thomas Steele Hall. *Treatise of man*. Harvard University Press, 1972.

Jones, Gwyneth A. *Deconstructing the starships: Science, fiction and reality*. Vol. 16.

Oxford University Press, 1999.

Stableford, Brian. *Science fact and science fiction: an encyclopedia*. Routledge, 2006.

Szollosy, Michael. "Freud, Frankenstein and our fear of robots: projection in our cultural perception of technology." *AI & SOCIETY* 32.3 2017, pp 433-439.

Roberts, Adam. *Science fiction*. Routledge, 2002.

Mori, Masahiro. "The uncanny valley." *Energy* 7.4 1970, pp 33-35.

Čapek, Karel. *RUR (Rossum's universal robots)*. Penguin, 2004.

Suvin, Darko. *Metamorphoses of science fiction: on the poetics and history of a literary genre*. Yale University Press, 1979.

Usher, Shaun. "I am not afraid of robots. I am afraid of people." Letters of Note, www.lettersofnote.com/2010/09/i-am-not-afraid-of-robots-i-am-afraid.html.

Accessed on 5 December 2017.

Anderson, Nicholas. "'Only We Have Perished': Karel Čapek's R.U.R. and the Catastrophe of Humankind." *Journal of the Fantastic in the Arts* Vol. 25 2014, pp 226-46.

Asimov, Isaac. "Runaround." *Astounding Science Fiction* 29.1 1942, pp 94-103.

- McCauley, Lee. "AI armageddon and the three laws of robotics." *Ethics and Information Technology* 9.2 2007, pp 153-164.
- Williamson, Jack. "With Folded Hands..." *With Folded Hands ... And Searching Mind*. Haffner Press, 2010.
- McCaffery, Larry. "An interview with Jack Williamson." *Science Fiction Studies* 54.18 July 1991.
- Dinello, Daniel. *Technophobia! Science Fiction Visions of Posthuman Technology*. UT Press, 2005
- LaGrandeur, Kevin. *Androids and Intelligent Networks in Early Modern Literature and Culture: Artificial Slaves*. Routledge, 2013.
- Jenkins, Will F. "A Logic Named Joe." *Astounding* 37.1 1946, pp 139-155.
- Forster, Edward Morgan. *The machine stops*. Sheba Blake Publishing, 2017.
- Clark, Andy. *Natural-Born Cyborgs: Minds, Technologies, and the Future of Human Intelligence*. Oxford University Press, 2003.
- Sladek, John T. "The Happy Breed." *Dangerous Visions* edited by Harlan Ellison, Orion Publishing, 2012, pp 318-30.

Tüfekçi, Zeynep. "Machine Intelligence Makes Human Morals More Important."

TED, June 2016,

www.ted.com/talks/zeynep_tufekci_machine_intelligence_makes_human_morals_more_important

Bartneck, Christoph. "From Fiction to Science—A cultural reflection of social robots." *Proceedings of the CHI2004 Workshop on Shaping Human-Robot Interaction*. 2004, pp 1-4.

Rubin, Charles T. "Machine Morality and Human Responsibility", *The New Atlantis*, www.thenewatlantis.com/publications/machine-morality-and-human-responsibility. Accessed on 30 November 2017.

Pepperell, Robert. "The Posthuman Manifesto." *Kritikos*, February 2005, pp 14.

Gottfredson, Linda S. "Mainstream science on intelligence: An editorial with 52 signatories, history, and bibliography." 1997, pp 13-23.

Lem, Stanislaw. "Golem XIV" *Imaginary Magnitude*. A Harvest / HBJ Book, 1985, pp 97- 226.

Dick, Philip K. "The Defenders." 2009. Project Gutenberg, www.gutenberg.org/ebooks/28767

Ellison, Harlan. *I have no mouth and I must scream*. Hachette UK, 2012.

Hubbard, F. Patrick. "Do Androids Dream: Personhood and Intelligent Artifacts." *Temple Law Review* 83 2010, pp 405-74.

Bukatman, Scott. *Terminal identity: The virtual subject in postmodern science fiction*.
Duke University Press, 1993.

Sloman, Aaron. "Why Asimov's Three Laws of Robotics are Unethical." July 2006.
www.cs.bham.ac.uk/research/projects/cogaff/misc/asimov-three-laws.html

Dick, Philip K. "Man, android, and machine." *The shifting realities of Philip K. Dick: Selected literary and philosophical writings*. Vintage, 1995, pp 211-232.

--- "The Defenders." 2009. Project Gutenberg, www.gutenberg.org/ebooks/28767

--- *Do Androids Dream of Electric Sheep?*. Del Rey Books, 1996.

Turing, Alan M. "Computing machinery and intelligence." *Mind* 59.236 (1950): 433-460.

MacDorman, Karl F., et al. "Too Real for Comfort?: Uncanny Responses to Computer Generated Faces." *Computers in Human Behavior* 25.3 (2009): 695-710.

--- "Individual Differences Predict Sensitivity to the Uncanny Valley." *Interaction Studies* 16.2 2015, pp 141-172.

Del Rey, Lester. *Helen O'loy*. Delos Digital, 2014.

De Fren, Allison. "Technofetishism and the Uncanny Desires of ASFR (alt. sex. fetish. robots)." *Science Fiction Studies* 2009, pp 404-440.

Kalioby, Rana el. "We Need Computers with Empathy." *MIT Technology Review*, 20 October 2017, www.technologyreview.com/s/609071/we-need-computers-with-empathy/ Accessed on 4 December 2017.

Larson, Glen A. and Ronald D. Moore, creators. *Battlestar Galactica*. British Sky Broadcasting, David Eick Productions and NBC Universal Television, 2004.