

WHAT TO DO WHEN MEETING E.T.?

AUGUST 12, 2019

RADBOD UNIVERSITY

Name: Max Stolk

Student Number: 4150619

Master (Specialization): Political Science (Political Theory)

Radboud University Nijmegen

Supervisor: prof. dr. M.L.J. Wissenburg

Second Reader: dr. B. R. van Leeuwen

Words used: 24352

Reference Method: APA

Picture: NASA, ESA. (2011). NGC 4631. Retrieved on 15th of February 2018 from:
<https://www.spacetelescope.org/images/potw1146a/>

Abstract

In this thesis it is researched whether humanity should adhere to Cixin Liu's Dark Forest Theory when meeting extraterrestrials for the first time. The theory's form is comparable to a social contract theory and was therefore compared to several classical contract theories such as those of Hobbes, Rawls, Kant, and those of Aquinas, Hobbes, Habermas, and Binmore. It became clear that Liu's theory is worthwhile investigating. The Dark Forest theory was, with the use of game theory, deemed internally incoherent but did present challenges that current social contract theories are unable to respond to. Specifically, the challenge of the vastness of space makes communication impossible. The conclusion of this thesis is that the Dark Forest Theory should not be adhered to when meeting extraterrestrial. Additionally, civilizations should not venture into outer space because it can threaten its existence. Future research on extraterrestrials should focus on rationality and game theory's limits and possibilities. Additionally, science fiction's possibilities of researching what constitutes humanity's knowledge and humanity itself should be broadened. Furthermore, science fiction should be investigated as philosophical thought experiments.

Key words: 5

Dark Forest Theory, Extraterrestrials, First Contact, Social Contract Theory, Science Fiction

Abstract	iii
List of Tables	v
List of Figures	vi
1. Introduction.....	1
2. Introduction to Game theory	8
3. The Dark Forest Theory	17
4. The Dark Forest Theory and Social Contract Theory	20
5. The Dark Forest Theory Evaluated.....	33
6. How should civilizations interact?	40
7. Conclusion	50
8. Bibliography.....	54

List of Tables

Table 1: Game with Mixed Motivation.	12
Table 2: Payoff Matrix of Prisoner's Dilemma	15
Table 3: Payoff Matrix of the Dark Forest Theory.....	19
Table 4: Payoff Matrix of Hobbes: Keeping or Reneging on a Contract.	21
Table 5: Alternative Payoff Matrix of Hobbes: Creating the Sovereign.	22
Table 6: Payoff Matrix of Binmore's Social Contract Theory.	24
Table 7: Payoff Matrix of Kant's Social Contract Theory.	27
Table 8: Payoff Matrix of Rawls' Social Contract Theory.	29
Table 9: Payoff Matrix for One Civilization of the Second Axiom.	37
Table 10: Payoff Matrix for Two Civilizations of the Second Axiom.....	38
Table 11: Payoff Matrix of Aquinas' Logic applied in Space.	41
Table 12: Payoff Matrix of Kant's Logic applied in Space.	43
Table 13: Payoff Matrix of Nietzsche's Logic applied to Space, option 1.	45
Table 14: Payoff Matrix of Nietzsche's Logic applied to Space, option 2.....	45

List of Figures

Figure 1: Overview of Game Theory.	15
---	----

1. Introduction

When humanity meets intelligent extraterrestrial life, will they be friendly or hostile? Science fiction has provided us with multiple answers. In *Arrival*, (Shawn, Linde, Levine & Ryder, 2016) extraterrestrials want to help us; in *E.T.*, (Kennedy, Spielberg, 1982) extraterrestrials appear to be neutral toward us and; in *Alien*, (Hill, Carroll, Giler & Scott, 1979) extraterrestrials are hostile toward us. In *Ender's Game* (Ulbrich, Card, Orci, Kurtzman, Pritzker, Chartoff, Hendee, McDough & Hood, 2013), it is unclear what kind of behavior we can expect from the extraterrestrial beings. It appears that in the best-case scenario, extraterrestrials help us advance and in the worst-case scenario, they help us meet our end (Musso, 2012).

Although many are convinced that the debate on the existence of extraterrestrial life is a contemporary one, it is, in fact, not (Crowe, 1997); in actuality, the debate started during the heyday of Greek philosophy. Atomists argued that extraterrestrial life exists, while Plato and Aristotle, among others, put forth that it does not. In the 13th century, Christian thinkers started to write on the topic as well, including Albertus Magnus and Thomas Aquinas. The thinkers wondered if God was able to create multiple worlds, and if so, do all the worlds have a savior like humanity's Christ? Individuals in the 16th century, like Galileo Galilei, Rene Descartes, and Johannes Kepler, mainly motioned to be cautious regarding extraterrestrials. During the Enlightenment, many thinkers wrote on the issue as well, including Immanuel Kant and François-Marie Arouet (Voltaire). Religious writers of that time argued that there must be other worlds. Their beliefs stemmed from the idea that God's powers were too vast to just create earth, it would have been a waste. There, thus, should be other inhabited planets. Other thinkers thought that the existence of extraterrestrials could destroy humanity's view of the universe, while others even thought that extraterrestrials could destroy humanity's planet. The two thoughts are still occupying many minds. The continuing importance of the debate in society and the involvement of renowned philosophers are indicative of the debate's significance and relevance to society.

Despite the centrality of this debate in the academic community, contemporary academic articles on this topic are rare. In the articles in which authors do engage with the topic, the distinction between hostile and hospitable extraterrestrials is used (Deardorff, 1986, Michaud, 2007, pp. 208-209). Within science fiction, one can notice the same method. Authors in both academia and science fiction, first posit the extraterrestrials' stance (are they hostile or hospitable?) and, based on that, then determine the impact the extraterrestrials' discovery might have. The shock within society and society's reaction is the focus here. More specifically, this research on hostile extraterrestrials mainly tries to answer the question whether we should continue searching for signs of life in space. It only briefly touches on the idea of interaction with extraterrestrials. Other scholarly contributions on the topic of extraterrestrials solely focus on the reaction that humanity will have when discovering extraterrestrials; an example of that is the Rio scale (Steven, 2013, IAA SETI Permanent Committee, 2016). The Rio scale is used to quantify the impact of finding extraterrestrials within societies. What should we do before establishing what kind of extraterrestrials are encountered? And how do we establish that? By the time we have investigated and discovered what extraterrestrials we are dealing with, they could have already destroyed us. This thesis will delve into how we should interact with extraterrestrials without knowing what stance they have.

In the next part of the introduction, I outline this thesis' relevance and support science fiction's inclusion in this project. To do so, I first delineate why research on how we should deal

with extraterrestrials is relevant. Second, the relevance of using science fiction is outlined, after which the science fiction books that are this thesis' starting point are considered.

1.1. Relevance of the topic

At first sight, the existence of extraterrestrials is a philosophical question; it remains a theoretical discussion if (and when) humanity will ever meet them. Natural scientists have tried to answer the question of extraterrestrials' existence and have developed the Drake Equation to assist (Maccone, 2010, Burchell, 2006). Formulated by Frank Drake in 1961, the equation includes several factors that are still unknown, making the expected number of civilizations within our galaxy between a couple and 280 million.¹ These numbers give rise to the Fermi paradox, which refers to calculations expecting galaxies full of life, including our own, despite empirical evidence showing us the opposite (Webb, 2015, pp. 21-26).² Humanity's presence in outer space is increasing due to spaceflight's commercialization, possibilities of space mining and, the colonization of celestial bodies like Mars, which increase the puzzlement created by the Fermi Paradox (Young, 2015, Loizou, 2006, Reddy, Nika, Wilkes, 2012, Capova, 2016). Additionally, humanity is broadcasting its existence into outer space with radio waves, meaning that our first encounter might not even be physical, or known to humankind.

There are already many answers to the Fermi paradox; however, most, if not all, cannot be (dis)proven. Stephen Webb (2015) provides us with 75 answers, stating that he did not include all theories he knows for a variety of reasons. Examples of these answers include that civilizations are around us but do not want to be found, humanity is unable to detect civilizations with current technology or, humanity is being monitored by extraterrestrials but, not evolved enough to be privy to their existence (Marino, 2012, Neuvel, 2016). Another answer is that civilizations are engaged in self-exploration via virtual reality, making them difficult to detect. All answers force us to think outside the box and they can hint at an answer to this thesis' main question.

This thesis' social relevance initially lies with the fact that humanity is unprepared for a chance encounter. The few existing documents lack moral and/or philosophical background and lack action-oriented planning. Examples include the UN's (1966) Treaty on Outer Space that only includes a small section on meeting extraterrestrials and the international Academy of Astronautics' (SETI Permanent Study Group of the International Academy of Astronautics, 2010) document, which is too limited and, therefore, not widely accepted. This is an issue because this ethical choice is one that can have deadly implications not only for humanity's planet and civilization but, also for another cosmic civilization. Vegetius' (2002, p. 79) famous quote is very apt here: "*Si vis pacem, para bellum*" [If you want peace, prepare for war]."

More specifically, this thesis' social relevance, which is political, is that its findings can be used to glean insights into the normative aspect of how humans should interact with each other (Walter, 2000). A contemporary example is that of refugees. The discourse surrounding refugees

¹ The factors are: 1) the number of galaxies, 2) the fraction of galaxies with stars, 3) the fraction of stars having planets circling them, 4) the fraction of planets that are able to support life, 5) the fraction of the planets that are able to support life that actually developed life at some point, 6) the fraction of planets that gave rise to intelligent life a.k.a. civilizations, 7) the fraction of those civilizations capable of sending detectable signs of their existence into space, 8) and the last step is the length of time that those signs are being sent into space (Webb, 2015, pp. 21-26). The first three factors are known, the other inputs must be hypothesized.

² Michael Hart's (1975) idea regarding the Fermi Paradox is thought provoking. He showed that a civilization, using ships that can travel at conventional speeds, could colonize the entire galaxy in less than one hundred million years. Seeing the age of our galaxy there has been ample time to do so.

focuses on the distribution of our finite resources (Angrist, Kugler, 2003, Borjas, 1994, Christopher, 2017, Dustmann, Hatton, Preston, 2005,). Examples of this discourse include individuals arguing that refugees are taking their jobs and/or receive a share of their resources. Cosmic civilizations can meet similar questions but in a different setting. Science fiction that show this include *District 9* (Jackson, Cunningham, Blomkamp, 2009). Science fictions might provide interesting solutions to the problems stated in discussions regarding refugees. Additionally, this thesis can add to research on how nations are interacting and how they possibly should interact. It can, for example, provide new insight into neo-colonialist practices around the world (Nkrumah, 1965, Obadina, 2000, Sartre, 1964). These practices tie into how finite resources are and should be distributed, which is at the core of the theory assessed in this thesis.

Conversely, current experiences of humanity of how individuals, groups, and/or states have interacted can influence the manner with which first contact with extraterrestrials is approached. More specifically, the way that states interact might be applied to other civilizations or deter certain actions. Thought experiments regarding extraterrestrials are spaces in which our current solutions to problems of cooperation and coordination can be tested, for example federations, intergovernmental and international institutions, and bilateral negotiations.

The scientific relevance lies within philosophy, specifically within Clark's (2001, pp. 201-202) argument that to truly know what humanity is, humanity needs to be compared with and referred to another species. Levinas investigates the same point but goes one step further. He argues that one can only exist and verify one's existence in relation to the other (Bevan, Werhane, 2011, p. 52). Children illustrate this point perfectly. Children grow when they can relate what other people express to their own needs. This in turn helps them understand both the other and themselves. Our society might be a toddler in that regard. Clark draws on Kant's writings who, in turn, confirms Derrida's suspicion that by studying ourselves we are also (re)creating ourselves. This makes it impossible to study what humanity is because you are changing, constructing and forming it in the process.³

Additionally, most studies that delve into the rationality of humankind, and the way humanity should deal with animals and its environment, use the human as a reference point. For example, when investigating animals and their rights this is an issue. Humanity always anthropomorphizes to a certain degree, humanity projects its ideas of a good life, rationality, and experiences on those of animals. This means that humanity is unable to 'objectively' study the case of animals and its environment as well as its own rationality because it cannot be neutral: it is currently tied to its own point of view. When thinking of humanity throughs science fiction, a researcher can come the closest to a situation in which thinking about and without referring to humanity is possible because the thought experiment includes radical non-human elements. This already points toward the relevance of science fiction. The next part will elaborate more on science fiction's relevance in this thesis.

1.2. Relevance of Science Fiction and the Dark Forest Theory

Many of the theories in Stephen Webb's (2015) book are used in, or derived from, science fiction books on the topic of extraterrestrials. Science fiction is thus used when researching this topic in general as well as in this thesis. The relevance of using science fiction for this topic is set out in this part. First the relation between literature in philosophy is delved into, while the second step

³ This sentence problematizes itself; this sentence does comment on what humanity is, i.e. humanity cannot be studied.

consists of specifically reviewing the relation between science fiction and philosophy. The relevance of the specific science fiction series is investigated in the third step.

Literature and philosophy have been going together for a long time with literary works often being used in philosophy as examples (Kalin, 1977). Although literature is often used, philosophers tend to neglect the philosophizing done in literature despite that literature can, and does, contribute to philosophy in three ways (Nussbaum, 1990, pp. 4-8, 38, 142). First, the writing style influences the content. A result of this, certain truths about humankind and the human life cannot be properly engaged with by philosophical prose. The issue lies with philosophy's argumentative and abstract style. Second, philosophy's focuses on rules and thereby tends to miss the particular and possibly unique details of a concrete situation. Therefore, philosophy misses the chance to instill moral rules in individuals (Kalin, 1977, p. 170). Literature can investigate specific situations, people, and circumstances and is therefore capable of instilling moral rules. Third, while literary works make the individual work to get the moral, ethical, or other insights from the text, philosophical works do not require the same kind of work. This latter point is especially true for analytical philosophy, as the discipline of phenomenology engages with many of these points by incorporating literature, poetry, and music in its analyses.

Science fiction is even better at tackling philosophical issues than other literary genres (Gomel, 2011, pp. 340-341, Freeman, 1968, Michaels, 2000). Science fiction can be best understood as the literary genre of "what if" (Evans, 1988). The following questions are raised: what if there are aliens? What if humanity creates a perfect artificial intelligence? What if cloning or making cyborgs were possible? Science fiction's thought experiments that explore 'what if' questions can furthermore include situations that do not resemble the world created by humanity: ordered, anthropocentric, and governed by humanity's rationality (Deleuze, 2003, p. 71). Science fiction, thus, creates opportunities from which new and unpredictable thoughts, ideas, and concepts can flow. Deleuze (1994, pp. 162-164), the philosopher who wrote on difference, argues that humans and specifically philosophers should create a new way of thinking, asking questions, and approaching the world. Deleuze puts difference in the spotlight because with difference new insights and possibilities can be explored. For example, new concepts of what humanity is and should be, and new knowledge. An example of science fiction that tries to respond to Deleuze's call for difference is Alex Garland's *Annihilation* (Bush, Macdonald, Reich, Rudin & Garland, 2018); in which, an extraterrestrial object lands on earth, but it is unclear if it is a life-form, a building or, a spacecraft. Specifically, during the final scenes Garland has tried to, in his words, "create an alien alien" (CNET, 2018).⁴ The alien in *Annihilation* appears to be unknowable because it lacks any definitive motivation and even its physical form is totally incomprehensible. In general, science fiction can dramatize elements of our reality, estrange our current reality, and highlight important aspects of our ethical and empirical world. Science fiction's cyborgs, clones, aliens, and hybrids provide us with insights on ethics in the realm of the post-human that reflect on our contemporary times.

With these tools, science fiction can invent, and prevent, futures; however, it is not involved in predicting the future. It can create futures that are plausible and desirable for humanity to strive toward (Xingshi, 1997). It can, and I argue it does, provide us with creative and

⁴ The entirety of the quote is rather interesting in this light: "When we deal with aliens, we often make them like us in some way. Maybe they want to eat us, or maybe they want our water, our resources (...). But these are all sort of human concerns and it seems like a legitimate thing to say that an alien might not be like us in any way at all. We are motivated by things and we have agendas, and an alien might not have an agenda, or might not be motivated and so it was an attempt to create an alien alien" (CNET, 2018).

sometimes inspirational material. Science fiction reflects on situations that can become reality quickly in this ever-changing world (Cole, 2017). Accordingly, military personnel are often required to read science fiction because it shows them a glimpse of what future combat might look like.

Science fiction lends itself better to investigating the future than philosophy's thought experiments and other literary genres because it allows for the inclusion of more foreign entities and elements. Expressed differently, it can create a bigger gap between current reality and possible future realities than academic thought experiments and other literary genres. Contrasted with philosophy's thought experiments, science fiction's include humanity and the unique details of certain situations, making them more emotionally charged which makes reflect humanity's reality better. This in turn makes it possible for science fiction to teach us more about humanity. This thesis highlights this possibility, thereby strengthening the scientific relevance by highlighting a tool suited for researchers to know more about humanity. It also ties into the social relevance, seeing that the more humanity knows about itself, the better it can predict its own actions and, perchance, prescribe certain actions for the future.

Liu's (2008a, 2008b) theory, called the Dark Forest Theory (DFT), is chosen as the starting point for this thesis. The science fiction series *Remembrance of Earth's Past* is where Cixin Liu conceived of this theory. This series is chosen for several reasons. First and foremost, it seems internally consistent when reading the series and thus possibly plausible in its adaption and application. It was a new and innovative view on the universe and the civilizations in it, at least to me. It is a theory that combines elements of rational choice theory, mutually assured destruction, social contract thinkers like Hobbes, and game theory, but does not resemble any of them. The DFT is supported by axioms in the books and there is enough information in the books to gather how the theory functions. Additionally, the DFT formulates and solves the Fermi Paradox, which is founded on its envisioned interaction between cosmic civilizations.

Moreover, the series has the possibility of influencing many, and via them, policies, laws, and institutions. This is due to the books being bestsellers around the globe, to which the many posts on the internet forum Reddit testify (Noixius, 2017, Pueojit, 2017, Alexandrawallace69, 2018, -Chinchillax-, 2016). Additionally, the fact that the previous president of the United States, Barack Obama, has read the series as well underscores the possible influence that the books have on policies and laws (Kakutani, 2017, Andersen, 2017). This makes this thesis highly relevant in social terms.

1.3. The Starting Point

Cixin Liu (2008a, 2008b) introduces the DFT in his astounding science fiction series the *Remembrance of Earth's Past*. In the series, Cixin Liu combines a thrilling story about humankind and extraterrestrial life with theories about the universe and the physics that make it whole. Only the first two books are introduced, shortly, because they create enough context to place the DFT in and the theory is thoroughly explained in the second book (Liu, 2008b, pp. 515-521). The books describe how first contact with extraterrestrial life is made and how earth's societies react; one group wants to invite the extraterrestrials to take over the world to get rid of corruption while another group wants to fight the invasion. The extraterrestrials' collaborators are defeated. However, the extraterrestrials gain direct access to all human communication, including the defense plans, via superior technology. The only remaining black box is the human mind. This principle is the backbone of the Wallfacer Project. The project entails that four individuals receive almost unlimited resources to design strategies that will deceive humankind and extraterrestrials

alike. Luo Ji is one of them. He is an astronomer and sociologist that is unambitious and seems to be a most unlikely choice. The extraterrestrials specifically want him dead though. In the end it is indeed Luo Ji who comes up with the plan that saves the world from destruction. The plan is referred to as the Dark Forest Deterrence and leans on the DFT. The DFT expects that civilizations will try to destroy each other to ensure their own survival.

Liu's "theory" is actually a theorem as he uses axioms, i.e. statements, and logical inference (Liu, 2008a, 2008b). However, I will keep referring to it as a theory because he himself uses the term and to avoid confusion. The axioms that Liu employs come down to civilizations wanting to survive while they are growing in a universe with resources remaining the same. Although these axioms seem coherent and logical, theorems are best challenged by questioning the underlying axioms.

The DFT makes use of game theory, a mathematical method of research investigating how social behavior between two or more interdependent actors should, does, and will look like (De Bruin, 2005, Sensat, 1998, Zagare, 2011). Game theory does so based on the preferences of the actors. It is an important field of study and is vital in the research of many other fields, including biology, law, political science, and economics (Watson, 2013, p. 2). However, game theory has not been used to try to answer the question of possible interactions with intelligent extraterrestrial life. I do just that in this thesis. Moreover, game theory is used as a tool in this thesis to model humanity's interaction with extraterrestrials, and to clarify and improve theories trying to explain and/or predict this interaction.

1.4. Research Question

The DFT's view regarding other cosmic civilizations and their interaction is grim and pessimistic about the chances of the survival of humankind. In the second chapter, I delve into the DFT and explain why it posits such a grim view. Internationally renowned individuals such as Stephen Hawking (Greshko, 2018) and Neil deGrasse Tyson ("Neil deGrasse Tyson Thinks", 2016) agree with this grim view. Both individuals think that if there are aliens, they are far more advanced than us, technologically and intellectually. They might see us as ants and destroy us as humans destroy ants. However, does that mean that we should (try to) destroy them the moment we meet them to safeguard ourselves? In this thesis, I try to find an answer to the question how we should interact with intelligent extraterrestrial life. Therefore, the main question for my thesis will be the following:

Should humanity follow the Dark Forest Theory when it interacts with extraterrestrial life?

1.5. Layout

The goal of this research is to investigate if the DFT should be applied by humanity when meeting extraterrestrials. To answer this question I use game theory as the tool to reach three goals. The first and most important element is that by dismantling the DFT into its game theory components, it becomes possible to set out the theory more clearly to create the best understanding of the theory as possible. The second use is that it is a mathematical tool with which it is possible to compare the theories. Third, it allows me to test theories and improve them. In the first chapter, I delve into game theory and its components to be able to apply it.

Next, I introduce Cixin Liu's (2008b) DFT in the third chapter after which I will compare the DFT to social contract theories in chapter four. Seeing the similarities and differences between

the DFT and social contract theories, the latter is used to challenge the first. This is possible because the DFT explains why agents should not work together, while social contract theories set out why it is best for individuals to work together.

There are two strands within social contract theories: contractarianism and contractualism. Thomas Hobbes' (Hobbes, Rogers, 2005) and Ken Binmore's (1994a, 1994b) social contract theories are used to introduce the contractarian strand. Hobbes' theory was the start of contractarianism and seems to resemble the DFT, while Binmore's theory makes explicit use of game theory and is therefore interesting to explore considering game theory being the tool used in this thesis. For contractualism, Immanuel Kant's (1785) and John Rawls' (1971) are chosen. Kant's theory is used because it started this social contract theory strand while Rawls' revitalized social contract thinking in the 20th century.

In chapter five I delve into the axioms underlying Liu's theory. A theorem's most vulnerable elements are its axioms; thus I test their internal coherence. First, I highlight the theory's intuitiveness and similarities with liberal thinking. Multiple authors have investigated this way of thinking, including Ayn Rand, Karl Marx, and Thomas Malthus (Hicks, n.d.). Max Weber (2002) provides us with the most insightful and critical investigation of the liberal and capitalist way of thinking because solely he includes both the economic and sociological elements in his research. His theorizing is, thus, used to reflect on this aspect of the DFT. The second step is to test the DFT's internal coherence, which I do by accepting the first axiom that civilizations' ultimate goal is survival, and testing if the second axiom supports the first axiom.

Seeing that the first axiom was accepted in the fifth chapter to challenge the second axiom, the first axiom is challenged in the sixth chapter. To do so, four philosophers are introduced, these are Thomas Aquinas (2006), Immanuel Kant (1785), Friedrich Nietzsche (Kenny, 2010, pp. 935-939), and Jürgen Habermas (1987). These philosophers are used because they are exemplars for many other theories, making them suitable for this exploratory research because it thereby can reflect on many other theories and focus future research. Additionally, these philosophers are well-known and broadly supported, increasing the likelihood of them being applied in a chance encounter with extraterrestrials, thus increasing this thesis' social relevance. A second goal of this chapter is to improve the DFT, if possible, based on the insights of the four philosophers.

In chapter six, the idea of communication is put forth as well. In Cixin Liu's (2008b) theory, communication is mentioned, but rejected, as an option to change the grim outlook of the DFT. The famous game theorist Schelling (1960) argues that communication can get actors out of a situation that is marred with hostility. Rawls, one of the most famous contemporary social contract theorists, argues the same. The question becomes: can communication provide a solution to the issue of interacting in outer space? This question is investigated in the sixth chapter as well.

Conclusions are drawn in the last chapter and future research paths are explored.

2. Introduction to Game theory

Game theory is in this thesis to model the interaction between humans and extraterrestrials. Thought experiments are devised with this tool because it is the sole method through which this interaction can be rationally approached. This is due to humanity's lack of experience interacting with extraterrestrials and it does not have an inclination on how it will be either. Game theory is used to clarify the theories and arguments presented, and to improve theories, if necessary. Specifically, it is used to create insights into the DFT. Game theory, furthermore, provides a method by which the included theories can be compared. These are the pay-off matrix tables of the theories. In this chapter, game theory's foundation is introduced, enabling the use of game theory to clarify, test, and improve theories in this thesis.

2.1. Start of Game Theory

Since ancient times, game theory has tried to answer how traditions, religious codes and civil codes came about, and how their standards for negotiation, contract, and punishment were established (Watson, 2013, pp. 1-2). Researchers still use game theory to logically approach human behavior. In game theory, actors should be aware of the influence of the other(s) on themselves and *vice versa* (Dixit, Skeath, 1999, p 16). This means that an individual, or in game theory jargon "a player", can and should consider what the other is going to do. In contrast, the act of making decisions refers to situations in which the other's response is not considered (Dixit, Skeath, 1999, p. 16). An actor can be an individual, group, or even country; however, the choice is always presented as the action of a single actor. Group processes are broken down into individual actions and internal decision-making processes by the scholar to make this possible.

The first step toward contemporary game theory was made in the 18th century, by mathematicians formulating strategies for parlor games (Watson, 2013, pp. 1-2). In the same century, Augustin Cournot and Francis Ysidro Edgeworth explored economic models in oligopoly and bargaining problems in an exchange economy. They were the first to formally apply game theory to social behavior of individuals in the society. In 1913 Ernest Zermelo proved the first formal theorem and later Emille Borel provided the academic community with a first concept of strategy.

Game theory, as it is now known, started with Von Neumann (Sensat, 1998, p. 382, Watson, 2013). He started the quest of the modern game theory in 1928 when he asked this question:

n players, $S_1; S_2; \dots; S_n$; play a given game of strategy G : How must one of these players, S_n ; play, in order thereby to achieve a result as favorable as possible? (As quoted in Sensat, 1998, p. 382).

Von Neumann defended that a player needs to opt for the strategy ensuring the maximization of the player's security in the worst-case scenario (Sensat, 1998, pp. 382-383, Watson, 2013, p. 2). Von Neumann, in collaboration with Morgenstern, concludes this based on accepting two structural elements. The first element is that each recommendation is for an individual player, consequentially, that recommendation should maximize that player's security. Second, all possible actions should be considered, even if the opponent does not. This ensures that the player is not taken advantage of by a nonconforming opponent. These structural elements still stand for most games today.

In the beginning, Von Neumann and Morgenstern focused only on two player games; later they realized that the introduction of more players is necessary to explain economics (Dufwenberg, 2011, p. 170). They also realized that there are non-zero-sum games. In their book, *The Theory of Games & Economic Behaviour*, Von Neumann and Morgenstern introduced these ideas and pioneered game theory. Their work led to game theory developing two strands, which were eventually developed by John Nash: non-cooperative game theory and cooperative game theory (Watson, 2013, p. 2). These will be explained in the next part on terminology.

A theorist who used game theory, and influenced political science, was Nicolas de Condorcet (Hamlin, 1996, p. 1334). Condorcet lived from 1743 till 1794 and criticized the revolutionary authorities during the French Revolution from his liberal perspective (List, 2013). Condorcet researched, among other topics, majority voting and found a paradox in possible voting strategies of the constituency. The simplest form of the paradox entails that three players must pick one of three alternatives (Herings, Houba, 2016, p. 142). The players' preferences are such that pairs can be made that favor one alternative. This means that one pair of players' interests align more with the first alternative than the second alternative, another pair's interests align more with the second alternative than the third alternative, and the third pair's interests align more with the third than the first alternative. How the players can ever come to an agreement is still debated, for example by Herings and Houba (2016). What Condorcet showed by this is that individual preferences can be rationale, while the aggregated majority preferences are not. To prove his paradox Condorcet made use of principles of game theory.

Game theory has been used to determine what the best manner is to proceed or, to analyze how to prevent something from happening. An example showcasing game theory's possibilities in a conflictual setting is the Cold War (Hagemann, Kufenko, Raskov, 2016, Belletto, 2009, Weintraub, 2017). After the Enlightenment, humankind started focusing on reason, even more so after World War II (Weintraub, 2017, pp. 150-151). Additionally, rationality was to be captured in (algorithmic) rules that could be applied consistently to a range of specified topics. Game theory tries to do just that. Condorcet was one of the theorists who was part of the movement who did not, explicitly, research the Cold War. Theorists who did research the Cold War using game theory include Thomas Schelling, Oskar Morgenstern, John von Neumann, and Daniel Ellsberg in America, and Nikolai N. Vorob'ev, Leon A. Petrosyan, Olga N. Bondareva, and Elena B. Yanovskaya in the Soviet Union (Erickson, Klein, Daston, Lemov, Sturm & Gordin, 2013, pp. 1-21, Hagemann, Kufenko, Raskov, 2016, pp. 99-100). Government intervention and funding sped up the development of game theory considerably, thereby showing once more that science, specifically game theory, can influence governments and that governments do influence science, in turn highlighting the relevance of this thesis.

2.2. Terminology

In the following part, the structure of strategic games is set out. Relevant to mention is that elements might differ from one class of games to another.

2.2.1. Rationality

When game theorists refer to the idea of rationality, they refer to individual's decision-making process being built on three axioms (Green, 2002, pp. 6-9, Osborne, Rubinstein, 1994, p. 6). The first axiom entails that the possible choices are known to the actor. The second axiom is the axiom of completeness and requires the actor to have a clear preference for an option or to be indifferent. The third axiom says that the preferences are transitive, meaning that if choice B is preferred to choice A, and choice C is preferred to choice B, then choice C should be preferred to choice A. In case of indifference, the actor should be indifferent to all options. The last axiom leads to the actor necessarily making a choice.

Based on the three axioms, theorists assume that the actors will always be able to perfectly calculate their payoffs and will always, without fault, opt for the best strategy (Dixit, Skeath, 1999, pp. 27-29). This does not entail that the players are selfish or short-term focused, and it does not mean that players need to share the system of values they use to assign payoff values to strategies. This is most interesting for this thesis, seeing that it allows for investigative research into extraterrestrials employing different rationales than humans. Furthermore, in most cases, it is unclear what kind of rationale the other actor will apply. One's value system is an information resource; therefore, finding out the other's value system is an important part of strategizing.

Underlying game theory is rational egoism (Shaver, 2017). Its principle is that self-interest is reached by fulfilling an end for which one needs to intend the means (Brunero, 2012, pp. 125-126). Rational egoism is considered an objective account of self-interest meaning that any gain, whether it was desired or not, should be valued. The other category is the preference account; this argues that solely the satisfaction of what one desires should be valued.

Rational egoism itself is the thought that an action is rational when it maximizes one's self-interest, which is the necessary and sufficient criterion (Shaver, 2017). It is supported by the instrumental theory on rationality which is roughly the same as rational egoism. The twist is that the instrumental theory on rationality makes use of the preference account while rational egoism makes use of the objective account.

Psychological and ethical egoism are other conceptions of egoism, but they do not fit game theory (Shaver, 2017). Psychological egoists accept that the "best" choice will not maximize their self-interest but does take care of their welfare (May, 2011). The latter part means that it is at odds with game theory models. Ethical egoism's main thought is that an action must be morally right to be maximize one's self-interest (Shaver, 2017, Nielsen, 1972, Burgess-Jackson, 2012). Considering that morals are not something that is embedded in game theory's structure or axioms, psychological egoism is also at odds with game theory.

2.2.2. Strategies

Strategies refer to the choices available to players (Dixit, Skeath, 1999, pp. 25-26). In games where only one action can be taken, and the players take the action at the same time, strategies of one choice exist. In sequential games, the choices later in the game can be adapted based on the actions of others, meaning that more choices are part of the strategy and that strategies need to consider the evolving circumstances. Strategies are considered complete if they take all the actions

of the other players into account. Strategies would need to be so complete that others can play the game for you. An action within such a strategy is referred to as a move.

2.2.3. Payoffs

Payoffs are considered by game theory to be one of the most important aspects of the decision-making process (Dixit, Skeath, 1999, pp. 26-27). Payoffs are tied to strategies and refer to the benefits, or negative consequences, one will experience when choosing that strategy. This is the foundation of actors' preferences. Payoffs are numerically scaled but can differ in their interpretation. Sometimes the numerical values refer to a ranking of outcomes, from the worst (1) to the best (5); however, it is also possible that they refer to the actual value attached to the outcome. The latter is considered to reflect reality better. Both systems consider the game's entire outcome; that is, the payoffs reflect everything that a player is interested in.

It is possible to have chances included in the payoffs; for example: 50% chance of success when choosing option A with a payoff of 100 and 50% chance of success when choosing option B with a payoff of 150. In this example the expected payoff of option A is (100×0.50) 50 and for option B is (150×0.50) 75; thus, the player would have chosen option B. Expected payoff thus refers to the mathematical or statistical expectation.

Sometimes the payoffs are referred to as utility, but utility in this context should not be linked to utilitarianism (Binmore, 2007, pp. 7-8). Utilitarians, like Bentham and Mill (1998), focus on pleasure and pain, or well-being, of individuals, or conscious beings. To determine the pleasure and/or pain of an action, utilitarians need to know the motivation behind the action. The question "what does this action bring to the actor?" needs to be answered. Utilitarianism is a form of consequentialism because it looks at the consequences of actions. Game theorists do not make assumptions regarding why someone regards a choice as providing more utility or based on what the differences between choices is made. To game theory's advantage, it does not make psychological assumptions; it solely investigates the consistency of the player. Here the rational element of game theory comes into play: consistent behavior is seen as rational behavior which is shown to be mathematically correct. Connected to the payoffs is the equilibrium.

2.2.4. Equilibrium and Evolutionary Games

An equilibrium is an important aspect in game theory because it highlights the optimal decisions for all parties (Dixit, Skeath, 1999, p. 32). There are multiple equilibria, but John Nash's was the first that could pin down the strategy chosen in non-cooperative games based on its distribution of payoffs with the use of only a few axioms (Dufwenberg, 2011, pp. 168-169).⁵

Nash argued that if one were to accept that players are rational and thus try to maximize their average payoff, it is unclear what the other would choose (Binmore, 2007, pp. 11-16). The Nash Equilibrium refers to the cases where both or all players at the same time pick the best option considering the other's choices. In such an equilibrium, a unilateral change results in a loss of payoff for the other, considering the other actor's strategy. In those cases, game theorists have begun circling those results. In these equilibria, the best choice for both actors considering the

⁵ As stated there are multiple forms of equilibria, all of these are alternatives to the Nash Equilibrium but all rest on the same fundamentals. There are for example the roll-back equilibrium, the strict Nash equilibrium, the strong Nash equilibrium, and the correlated equilibrium (Dixit, Skeath, 1999, pp. 192-193, Apt, 2015, pp. 114-116). There are some equilibria, such as the Walrasian Equilibrium or the Rational Expectations Equilibrium, that are merely a manner to apply the Nash Equilibrium to a certain case (Dubey, Geanakoplos, Shubik, 1987). Due to the limited size of this paper I will not delve further into these differences, seeing that it is solely necessary to expound on the original Nash Equilibrium.

possible actions of the other actor is chosen, meaning the individual payoffs might not be the maximum possible (Dixit, Skeath, 1999, p. 30). This is perfectly illustrated by the prisoner's dilemma.

An example of the Nash Equilibrium can be found in Table 1 (Binmore, 2007, p. 11-16). It is the strategy matrix for the following game: two automobile drivers are in a narrow street which is too narrow for them to pass each other without one of them slowing down. When both slow down both lose more speed than necessary and time is lost. No damage is done to the cars though. In the case in which neither slows down, the cars will collide and both will sustain damage. When only one slows down, they will not hit each other, not lose time, and gain the most from this interaction. Table 1 is the numerical representation of this game; the number to the right represents the payoff for driver one, and the number to the left represents the payoff for driver two.

Table 1: Game with Mixed Motivation.

	Driver Two		
	(1, 2)	Slow Down	Sustain Speed
Driver One	Slow Down	3, 3	0, 4
	Sustain Speed	4, 0	-1, -1

The Nash Equilibrium's relevance for game theorists is twofold (Binmore, 2007, pp. 15-16). The Equilibrium "short circuits the infinite regression by observing that any other strategy profile will eventually be destabilized when the players start thinking about what the other players are thinking" (Binmore, 2007, pp. 15-16). This means that instead of thinking of all the possibilities that the two actors can take in reaction to the other's previous action *ad infinitum*, it solely focuses on the best choice of the actor regardless of the other actor's actions.

The second reason for the Nash Equilibrium's relevance is its link to evolutionary processes (Binmore, 2007, pp. 16-17). The evolutionary interpretation of the Nash Equilibrium refers to the weaker forms of the game's results dying out due to natural selection. The weaker forms are not played, and, eventually, will not be an option. Examples of the Nash Equilibriums are found in biology. The Nash Equilibrium refers to genes' or actions' options that 'fit' best; survival of the fittest takes care of the rest. The strategies become part of a species; they are hardwired to use a certain strategy in certain cases (Dixit, Skeath, 1999, pp. 12-13). The same happens in society. Education, monetary incentives, and other's approval will 'program' individuals to portray certain behavior. This evolutionary interpretation of game theory increases its explanatory power by showing that individuals do not even need to be aware of game theory, much less understand it, to apply the principles. The same is true regarding the effectivity of adjustments: they get rid of the equilibriums that cannot be characterized as a Nash equilibrium.

2.2.5. Rules

The game consists of four rules: 1) there is an actors' list, 2) there are available strategies for all players, 3) actors know their respective payoff for every combination of strategies, and 4) all players try to maximize their payoff (Dixit, Skeath, 1999, p. 29). This does not mean that in real life these rules are met; in some cases, it is unclear if there is another player, what strategies are possible, and what each player's own value system is. Players can use this to their advantage.

2.3. Classifying games

Games can be separated into many categories based on their features and the contexts in which they arise (Dixit, Skeath, 1999, pp. 18). These have a great influence on how one can think of the game and its solutions. In this part, the categories will be introduced and, when relevant, they will be reviewed more in-depth.

2.3.1. Sequential or Simultaneous moves

The first distinction to be made is that between sequential and simultaneous moves (Dixit, Skeath, 1999, pp. 18-19). The difference lies with the actions of the actors being executed at the same time, like in traffic, or actions being executed at different times, such as when playing chess. This influences how actors need to think about their actions. In a sequential game one can wait to think of their next move till after the other has moved. Sequential games can inform us when it is advantageous for an actor to move first. In a game where actors make moves simultaneously, they need to think about what the other is doing right now. This makes it possible for actor A to try to decipher what actor B is going to do, while actor B is doing the same regarding A, and they will both consider that the other is doing this.

2.3.2. Interests

There are multiple classifications regarding the compatibility of the interests of the actors (Dixit, Skeath, 1999, pp. 19-20). When the interests are in complete opposition, the game is a zero-sum game. A zero-sum game is a game in which the gain of payoff of one of the players will be compensated by the equal loss of payoff of another player. There are games in which the interests of the actors are not in complete competition, for example some trade and economic transactions can benefit all actors involved. It is even possible in a zero-sum game in which more than two actors participate that two or more actors will collaborate. Alliances and coalitions are the focus of this field of game theory.

2.3.3. Repeated or Singular Game

Some games are played ones while other are repeated with the same or different players (Dixit, Skeath, 1999, pp. 20-21). In games where actors only meet once, it is possible for the players to not be aware of the capabilities of the other and have less information as a result. Games with multiple encounters make it possible for the actors to get to know each other and to build a reputation. In cases where different players partake in different rounds of the same game, players do not know the characteristics of the players or have information on them in general. Repeated encounters with the same player(s) make it possible to introduce mutually beneficial strategies or to punish wrongdoings in certain games. Examples of these strategies are tit-for-tat and eye-for-an-eye. On a more general level: one-time games might be zero-sum games, but the same game played more often can become mutually beneficial. This means that a game does not have to be characterized by only conflict or only cooperation, but rather a mix of both.

2.3.4. Information

Information on other's or the situation can influence one's strategy significantly as, for example, some strategies seem possible with or without information (Dixit, Skeath, 1999, pp. 21-22). Both players can have the same information, which can be all the available information or just parts of it. Card games such as poker, bridge, or klaverjassen are examples of players just knowing certain information. It is also possible that one player has more information than the other. Breakthroughs in a certain field, e.g. technology, are perfect examples because these instances also show that the sharing of information needs to be done strategically. Actors want to release

information that works to their benefit and hide information that might have negative consequences for them. The question then arises: do actors trust in the information that you share? The devices and strategies surrounding the sharing and receiving of information can become as complex as the game itself.

2.3.5. Rules

One of the more important distinctions for this thesis is of games with fixed or flexible rules (Dixit, Skeath, 1999, pp. 22-23). When playing card and board games, or a sport, the rules are set; however, in most aspects of life, the players make their own rules. When the latter is the case, then the most important moment is when the rules are made and or changed. Dixit and Skeath (1999, p. 22) refer to this as the pregame. Another element that comes into play when talking about rules is changing them and acting within them. This clears the path for the introduction of threats and promises, for these depend on the rules. They especially depend on the moments before or in the game when the rules are established.

2.3.6. Cooperative versus non-cooperative games

The last distinction made within game theory is between those games in which agreements to cooperate are enforceable and those games in which that is not possible (Dixit, Skeath, 1999, pp. 23-24). This happens mostly in games which are not classified as zero-sum games but require cooperation. Most games are characterized by the latter option. Games like this can be used to approach negotiations. After an agreement has been made and the spoils have been distributed, it is still unclear if the participants will uphold their end of the bargain. How the players hold each other accountable depends on the case. In many cases, the agreed upon actions cannot be executed immediately and/or are not directly observable. This makes the enforceability of agreements difficult.

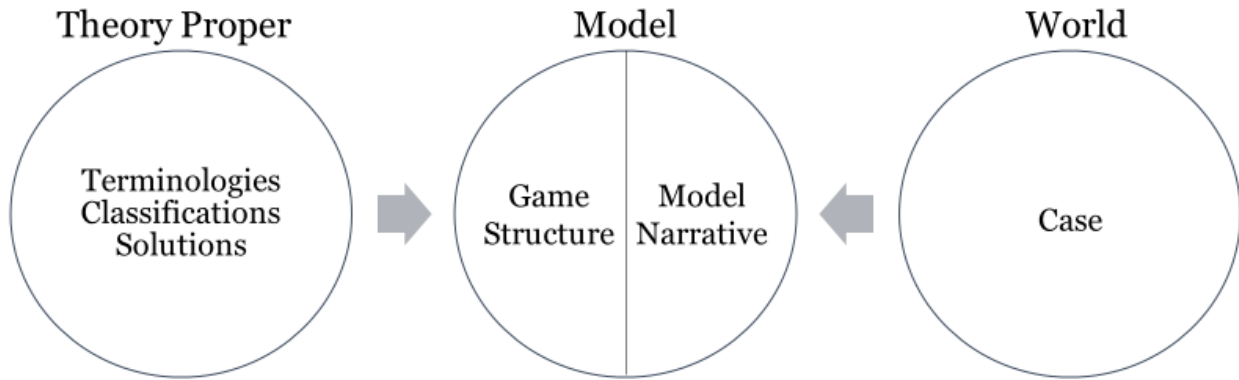
Games in which agreements can be enforced are referred to as cooperative games and games in which enforcement is difficult are referred to as non-cooperative games (Dixit, Skeath, 1999, p. 24). The difference is that in cooperative games, strategies are collectively chosen, while in non-cooperative games it is done by individuals. The distinction does have a significant impact on how the games are played and modelled. In non-cooperative games, the actions made by the players are treated as individual actions (Watson, 2013, p. 3, De Bruin, 2005, Zagare, 2011, p. 12). The actions by players are made individually and are consequently not tied to the actions of other players. The game is seen as the interaction between players which consists mostly of offers, counteroffers, and other gestures individuals can make in a game. Cooperation will last only as long as the interests of the different players align.

Cooperative games step away from the individualistic approach to game theory and look at the solution to the game as a joint decision made by all partaking players (Dixit, Skeath, 1999, p. 24, Watson, 2013, p. 4). This requires different models with different underlying concepts compared to non-cooperative game theory. Due to the focus being on joint decision-making, it is often used to analyze the process of establishing a contractual relationship. Contractual relations include situations in which the contracts are formal, written and explicit, as well as informal, such as agreeing to go to dinner at a restaurant.

2.4. Conclusion

It has become clear that game theory consists of many different elements that interact, and that the researcher needs to consider when applying them to his/her case. When applying game theory to a case, the situation can be summarized in the following scheme:

Figure 1: Overview of Game Theory.



First, the theory proper refers to the all the elements outlined in this chapter: the terminology, classification, and solutions to the game (Grüne-Yanoff, n.d.). The third element is introduced by expounding on the equilibrium. The model narrative is an interpretation of a case that the game theorist wishes to investigate, which informs the game theorist what game structure he/she needs to apply with the elements stemming from the theory proper. An example is the famous prisoner's dilemma. In the dilemma, two criminals are apprehended, but it is unclear who perpetrated and who abetted the crime. Both get the option to tell on the other and get released immediately while the other serves the entire time of ten years; if both choose this, then they both get a reduced sentence of five years. If neither does, both get released after a year. If the criminals only care for themselves, then they will choose the option that gets them the least prison time, which is the model narrative. The game structure is that it is a singular, non-cooperative game with simultaneous moves. The model is then as set out in Table 2, with the numbers indicating the lengths of the sentences in years:

Table 2: Payoff Matrix of Prisoner's Dilemma

		Criminal Two	
		(1, 2) Remain Silent	Tell
Criminal One	Remain Silent	1, 1	10, 0
	Tell	0, 10	5, 5

Game theory's number of elements is vast; thereby complicating finding the correct classification of the case. The correct interpretation of cases is subject to constant research and questioning within the academic world; it is not always clear what kind of classification fits the situation in the real-world best. We should, therefore, keep in mind the difficulties of selecting the correct classification when applying game theory to the DFT. This way, we can identify the theory's underlying rationality and determine if we can, and should, translate science fiction to academia.

For this thesis it is important to remember why game theory is included, which was to test, compare, and improve on the theories included in this thesis. Game theory can test the theories by approaching the theories as cases and determining if they are consistent and applicable. The comparing of theories can be made easier by game theory as well, by including the table that can be formulated by the interaction of the case and game theory. These tables can be more easily compared, also because it is a more visual presentation of the theories.

3. The Dark Forest Theory

In this chapter, I outline the DFT as explained by Cixin Liu (2008b, pp. 515-521) and dismantle the DFT to its game theory components. Liu sets out the theory the best in the trilogy *Remembrance of Earth's Past's* second book, aptly called the Dark Forest Theory. Liu builds the theory on two axioms. According to the first axiom, every civilization's central aim is to survive. The second axiom refers to the continuous growth of civilizations while matter in the universe remains constant. Due to civilizations expanding in numerical and geographical terms, they will need more resources over time. How this leads to fighting I will explain below, but first there are two more elements that should be introduced. These are Liu's idea that distances in space are so vast that communication between two civilizations can take a while. The other element is that civilizations can experience a technological leap. A technological leap refers to relatively short periods of time in which a civilization experiences tremendous technological growth.

In the case that two cosmic civilizations meet, there are three stances that they can adopt, according to Cixin Liu (2008b, pp. 515-521); these are: benevolent, malicious, or to keep quiet. "Benevolence means not taking the initiative to attack and eradicate the other civilizations. Malice is the opposite" (Liu, 2008b, p. 517). Keeping quiet is only shortly introduced and dealt with quickly in the book. Keeping quiet means that the civilization that finds the other civilization does not start communication to become friends, nor attacks to eradicate the other. Benevolence regarding the other civilization is dangerous: civilization A does not know if civilization B has malicious intentions. Maybe civilization B wants A's resources or is xenophobic; whatever the reason is, civilization A and B are not sure that the other is benevolent. The ever-growing civilizations will eventually meet each other in the finite matter of this universe, making it impossible to ignore the other eternally. Due to the first axiom, civilizations will never want to risk their survival, thereby eliminating this option. Keeping quiet is also dangerous. In the case of civilization A finding civilization B, B might know it was spotted, possibly leading to civilization B wanting to eliminate civilization A. Or civilization B experiences a technological leap and is then able to find civilization A and maybe even threaten its existence. The third option of assuming a malicious stance is the only option that is viable. This is strengthened by the *chain of suspicion*. The chain looks like the following: if civilization A suspects that civilization B might want to eliminate civilization A, then civilization A is motivated to eliminate civilization B. This means that civilization B is in a position that it should assume that civilization A might want to eliminate civilization B as well. This repeats itself *ad infinitum*. This leads to distrust and fear; thus civilization B and A are, due to their need to survive, logically ordered to eliminate the other civilization.

To situate the different elements of the DFT in game theory I start by setting out its structure and rules. By doing so it becomes clear how the DFT functions. Succeeding, I present the payoff matrix of the theory.

First, there are only two players in the game: civilization A and civilization B. Do keep in mind that the DFT does not apply to every civilization in the universe, but only to cosmic civilizations, i.e. civilizations that can venture into outer space. Additionally, the DFT assumes that both cosmic civilizations have, or can obtain with a technological leap, the means to destroy another civilization. For if this were not the case, one of the civilizations would not be threatened with destruction, limiting the cases to which the DFT is applicable. It is of course possible to have more players in the scenario; however, the DFT does not make any suggestions on how such a

scenario will develop. The scenario will still suffer from the same distrust, and thus all civilizations should try to destroy the other two civilizations.

Second, the DFT puts forth that the game between the civilizations should be characterized as a zero-sum game. The fact that only one of the civilizations in the end will be safe by destroying the other civilization makes it so that the gain, i.e. survival, is compensated by the loss of the other civilization, i.e. destruction. One could even argue that the destruction of the other civilization leads to the possibility of mining its planet(s) and learning from its technology and way of life unobstructed. In such a case, the game is still a zero-sum game because the civilization destroying the other would gain an equivalent to what the other loses. The game is non-cooperative as well, as the acts of the players are seen individually, creating a setting in which the enforcement of an agreement is difficult, if not impossible. This distinguishes this game from a cooperative game.

Third, the choice to want to become friends with, leave alone, or destroy the other civilization is made by both civilizations simultaneously the first time they meet. However, the DFT accepts that one can change positions, but only from befriend and leave alone to destroy. It is impossible to befriend or leave alone a civilization that you have destroyed. This conclusion must be reached seeing that in the DFT, a civilization is either capable of destroying a civilization or becoming capable by means of a technological leap. The DFT is best represented in a simultaneous game because the civilizations will not wait and abide their time to let the other civilization decide.

Fourth, regarding rationality and common knowledge, I believe that there is no additional information to give than the introduction to the DFT as presented in the previous chapter. This comes down to the axioms of wanting to survive and civilizations continuously growing in a universe that consists of finite matter. Two other elements of common knowledge were that civilizations can experience a technological leap and that communication between cosmic civilizations takes time. The axioms and the two added nodes of information are assumed to be discovered by every civilization that can rationally approach problems and has the capability of going into space.

Fifth, Cixin Liu's (2008b) DFT accepts that the cosmic civilizations will have only three strategies to choose from: to be benevolent, malicious, or to keep quiet. However, it is unclear if it is better for a society to keep quiet or benevolently approach the other. In both cases, the likelihood that the civilization is going to be destroyed is present, making benevolence and keeping quiet worse options than maliciously approaching the other in an attempt to destroy it. Another option that becomes possible if both civilizations opt to destroy the other is that of a mutual destruction. This can even be something that can save both civilizations for a duration of time, by one of the civilizations threatening Mutually Assured Destruction (MAD). In the book, this is an equilibrium found as well. However, implied throughout the books is that civilizations can differ in power. This can be the case when they meet or after when one of the civilizations experience a technological leap. If one civilization is stronger, it can possibly negate MAD or sees a chance during a stalemate and grasps it. In the MAD situation, neither civilization has ensured its survival. In that sense, a MAD situation is like a befriend situation, as survival is impossible to guarantee in both situations.

To conclude, two things have become clear from investigating the DFT. The first is how to translate the DFT to game theory terminology and a pay-off matrix. Cixin Liu's DFT is a game played once in which decisions are made simultaneously. The pay-off of the options is unclear, but it can be assumed that befriend and fleeing have the same pay-off, seeing that neither eliminates the chance of getting destroyed. The most preferred action is to destroy the other. In

the case of both civilizations trying to destroy the other, it is unclear who is going to succeed (first), meaning that there are three possible outcomes. These are: both civilizations are destroyed, civilization A is destroyed or, civilization B is destroyed. Concluding, the DFT's pay-off matrix can be found in Table 3. The number one stands for the option that has the highest pay-off and is thus most desirable; two stands for the less desirable options.

Table 3: Payoff Matrix of the Dark Forest Theory.

	Civilization B			
	(A, B)	Befriend	Leave Alone	Destroy
	Befriend	2, 2	2, 2	2, 1
	Leave Alone	2, 2	2, 2	2, 1
	Destroy	1, 2	1, 2	1, 2, 2-1, or 2, 2

The second insight that is distilled from DFT's analysis and its translation to game theory is that on the surface it resembles social contract theory. However, instead of rationally indicating why individuals should cooperate, the DFT indicates why, rationally, individuals should destroy each other. This contradiction is investigated in the next chapter.

4. The Dark Forest Theory and Social Contract Theory

4.1. What is Social Contract Theory?

In this chapter, Liu's (2008a, 2008b) DFT is compared to social contract theories to critically engage with the former. I compare them because of the commonalities between the DFT and social contract theories (D'Agostino, Guas, Thrasher, 2017). Both the DFT and social contract theories present a case why two or more individuals should or should not cooperate based on logic. Liu's theory shows why it is the best choice to not cooperate, but to destroy the other. Social contract theories the opposite; it shows why players should cooperate. By comparing these theories, it allows me to assert that Liu's theory is a new theory and thus interesting to be researched. Additionally, it shows how new elements are introduced by the DFT, i.e. the fact that it is taking place in outer space, with which social contract theories have not been in contact. These elements make it interesting to delve into the axioms underlying the DFT, which can be used to ascertain how different and coherent the theory is.

To compare the DFT to social contract theories, I start by setting forth what a social contract theory is and then engaging with its two main strands. Two theorists are discussed per strand, which will in the end compared to Liu's theory. By comparing the theories, other ways that individuals can, and perhaps should, interact are shown. This, in turn, provides the foundation to examine the axioms underlying the theories in chapter five.

Social contract theories' goal is to show how governmental authority can be rationally justified and how a society can be started (D'Agostino, Guas, Thrasher, 2017, Gauthier, 1986). By focusing on a rational justification, social contract theories distance themselves from other moral and political philosophy. The two main strands of social contract theory are contractarianism and contractualism (Cudd, Eftekhari, 2017). The first strain of social contract theory presented is contractarianism and the second contractualism.

For each strain two thinkers are presented, followed by a short conclusion before showing how Cixin Liu's (2008a, 2008b) DFT is different from social contract theories. For contractarianism, the two thinkers presented are Thomas Hobbes (Hobbes, Rogers, 2005) and Ken Binmore (1994a, 1994b). Hobbes' theory is chosen because Hobbes' theory seems very similar to the DFT and it is the theory that founded the contractarianism strain of social contract theory. Binmore's theory is especially interesting because the theorist makes explicit use of game theory to formulate, develop, and show the strength of his social contract theory. Thus, both Hobbes' and Binmore's theories are easily connected to the main themes of this thesis. For contractualism, Immanuel Kant (1785) and John Rawls (1971) are chosen. Kant is one of the founding individuals of the tradition, while Rawls is the theorist who has rekindled the tradition in the 20th century (Cudd, Eftekhari, 2017). Both theorists and their theories are therefore well-known and have influenced many other contractualist theories, making the theories a great starting point for this exploratory thesis. Seeing that contractualism stems from a Kantian thinking, the part on contractualism starts with his theory.

All the thinkers' social contract theories are expressed in pay-off matrixes to make the comparison easier. The pay-off matrixes might not align perfectly with the intentions of the theories' authors, but the aim is not to reflect the theories completely. Rather, the goal of the matrixes is to simplify the comparison of social contract theories on two elements and their interaction, i.e. incentives and outcomes.


4.2. Contractarianism

4.2.1. Thomas Hobbes

Thomas Hobbes' theory starts with the construction of the state of nature in which everyone is at war with everyone because there is no government overseeing and enforcing rules (Gauthier, 1988, pp. 71-72, Friend, n.d., Ryan, 2012, pp. 424-436). In it, nobody profits because everyone is fighting for survival, resources, and honor. Hobbes assumes that the actors are rational egoists and roughly equal in strength (Gauthier, 1988, p. 71, Friend, n.d., Hampton, 1986). This entails that actors only act, engage, or enter into a contract to strengthen their own position. This means also that one should not follow a contract even when signed. This is due to the pay-off matrix in the state of nature, as reflected by Table 4 Part I (Hampton, 1986, pp. 62-63). One refers to the strategy with the highest pay-off for the actor, two to the second highest, three to an even lower pay-off, and four to the strategy with the lowest pay-off. Even though the highest pay-off is when no actor attacks, both actors are forced to attack to ensure survival. This is because the risk of not attacking is too high and does not secure survival. Hobbes thus argues that if one does not attack, it is rational for the other to attack.⁶ A similar argument is presented in the well-renowned prisoner's dilemma.

Hobbes argues that one should adhere to the contract if one wants to survive and others are willing to adhere to it as well (Hampton, 1986, pp. 63-79). The underlying reasoning is that if an actor does not adhere to a contract he agreed on, he distances himself from the others and creates distrust. This makes it impossible for the dissenter to enter a confederation by which a sovereign is established. A sovereign is the best chance to ensure security by enforcing rules and legal sanctions (Hampton, 1986, p. 133). Considering this line of argumentation, not following a contract, threatens, or at least lowers, the individual's chance of survival. Hampton shows, with game theory, how Hobbes' pay-off structure changes with the introduction of the sovereign. This is reflected in Table 4, in which part I refers to pre-sovereign situation, and Part II refers to the situation which includes the sovereign. The number one refers to the strategy with the highest pay-off for the actor, four to the lowest pay-off.

Table 4: Payoff Matrix of Hobbes: Keeping or Reneging on a Contract.

I:	Individual B				II:	Individual B		
Individual A	(A, B)	Keep	Reneg		Individual A	(A, B)	Keep	Reneg
	Keep	2, 2	4, 1			Keep	1, 1	2,3
	Reneg	1, 4	3, 3			Reneg	3,2	3,3

Yet another interpretation of Hobbes' theory in game theory terms can be found in Table 5. If individuals do not enter into a social contract and continue to stay within the war of all against

⁶ Hampton (1986, pp. 61, 68-79) discusses this interpretation of Hobbes, but contends that there are multiple interpretations of Hobbes possible that contradict. For a more in-depth analysis of Hobbes' theory one should read Hampton's (1986) book "Hobbes and the Social Contract Theory".

all, it becomes possible that no one or only one will survive. Because humans need someone of the other sex to procreate, it is possible that, if the state of nature is not ended, this is made impossible. This would mean humanity ceases to exist. The translation to a pay-off structure of this interpretation of Hobbes can be found in Table 5. The number one refers to the strategy with the highest pay-off for the actor, two to the lowest pay-off. If there is no one survives to enjoy the pay-off.

Table 5: Alternative Payoff Matrix of Hobbes: Creating the Sovereign.

II:	Individual B		
Individual A	(A, B)	Keep	Reneg
	Keep	1, 1	-, -
	Reneg	-, -	-, -

4.2.2. Ken Binmore

Ken Binmore's (1994a, 1994b) social contract theory relies on evolutionary game theory, as biology and economics have done as well (Luetge, 2015, pp. 183-184, Snugden, 2002, pp. 230-231). Binmore (2014) argues that human morality is the result of evolution in human behavior. Evolution is the result of human interactions. These interactions are considered coordination problems, meaning that there are Nash equilibria to be found which determine the standards of society. Underlying the theory is the concept of emphatic preferences, which states that individuals can adapt their behavior more appropriately when others' behavior is predictable. This is also the reason why individuals enter into Binmore's social contract; the social contract makes the standards enforceable and thus one can easier predict others' behavior (Luetge, 2016, pp. 351-352). This drive to be able to predict stems from the fact that the individuals want to achieve/gain something and that it is harder, or even impossible, to achieve alone. Individuals in Binmore's theory thus adhere to social contract for the same reason as in Hobbes': for their self-interest.

Binmore's social contract theory relies on the construction of two games: the Game of Life and the Game of Morals (Luetge, 2015, pp. 183-184, Snugden, 2002, pp. 230-231). The first is built on empirical terminology: it is the game that reflects the actual framework within which the interaction takes place. The Game of Life is subject to change but cannot be changed by humans. It is a repeated game, in between which the individuals can opt to be play the Game of Morals (Binmore, 1994b, pp. 422-424, Luetge, 2015, pp. 183-184, Snugden, 2002, pp. 231-232). This is a fictional game, for it takes place in the individuals themselves and in the game the individuals do not know their position in the interaction but retain their emphatic preferences. Within this game, the individual will choose the option that maximizes his/her pay-off, like in all game theory related thought experiments. All actors will then act on one option within the Game of Life and if the individuals find that they have acted in correspondence with each other, i.e. have found an equilibrium, this consensus reached is considered fair by Binmore. This consensus then becomes the social contract according to Binmore. In the case that the actors do not find an equilibrium, the social contract found is not fair and the Games of Morals will be invoked again. With every iteration the actors involved will make a social contract that is considered fairer than the previous one.

Individuals in Binmore's (Luetge, 2015, Snugden, 2002, pp. 228-229) theory thus adhere to the social contract for the same reason as in Hobbes' theory: for their own self-interest. The actors want to be able to predict the others' behavior to be able to adapt their own behavior to what is deemed acceptable and expected. This drive to be able to predict stems from the fact that the individuals want to achieve/gain something and that it is harder, or even impossible, to achieve alone. However, in the cooperation with others, the actor wants to know what he/she can expect and should consider.

Binmore's theory explains how by iterating the game, the weaker equilibria are phased out of existence (Luetge, 2015, Snugden, 2002, pp. 231-232). This means that certain equilibria will not even be entertained after a while. This means that the best way of acting changes constantly. It is therefore of no added value to attempt to capture the norms established according to the theory in a pay-off matrix. What can be captured in a pay-off matrix is the choice between engaging with an other. This is because it is only by interacting with others that it becomes possible to achieve certain goals and possibly influence the changing of the equilibrium. Thus, the pay-off matrix reflecting the choice to participate in the game according to Binmore's theory is captured in Table 6. One refers to the strategy with the highest pay-off for the actor, and two refers to the lowest pay-off.

Table 6: Payoff Matrix of Binmore's Social Contract Theory.

	Individual B		
	(A, B)	Participate	Block
	Participate	1, 1	1, 2
	Block	2, 1	2, 2

4.2.3. Contractarianism's Basis

It has become clear that the basis of contractarianism is mutual self-interest, thus contractarians assume that individuals are egoists. Maximization of one's interests is the goal in the bargaining with others. This also means that contractarians claim that legitimate authority (of a government) is based on the will of those that are governed by it. Additionally, contractarians argue that morality is found by individuals rationally selecting strategies that maximize their payoff. Before comparing these assumptions with Liu's (2008a, 2008b) theory, contractualism is investigated by introducing Kant (1785) and Rawls (1971).

4.3. Contractualism

4.3.1. Immanuel Kant

Immanuel Kant's goal was to create a situation in which individuals' freedom and their lifestyles are safeguarded (Rauscher, 2016). The latter is what makes Kant a contractualist. Kant argues that individuals, as rational beings, have a duty to engage in a social contract to safeguard their freedom. One of the two social contract theories of Kant focuses on property.^{7 8}

Kant approaches the social contract via property by arguing that individuals' private property should be safeguarded (Rauscher, 2016).⁹ The idea being that one needs private property to pursue one's ends. Striving to reach one's ends consists of one legitimately using their objects to their own liking without being obstructed by others. State authority is necessary for both the existence of private property¹⁰ and the exercise of freedom by all. The state will then act in accordance to the will of the people; however, this will is only a fiction. There is no empirical, historical act that created the sovereign. The will of the people should be interpreted as if, and only as if, all the people went to vote in line with the policies enacted.¹¹ In other words, as if the people have actively consented to the policies. This is a limitation of the sovereign's power.

The social contract is a rational justification of state power, and in Kant's case based upon the idea that the contract reflects reason, and, as rational beings, humans can only be part of the erected society (Rauscher, 2016). There is no choice in the matter. This is reflected in the game theory perspective on Kant's social contract theory because there is no other option than to work together. The reflection of Kant's social contract theory can be found in Table 7. One refers to the strategy with the highest pay-off for the actor, and two to strategy with the lowest pay-off.

⁷ It is important to note that Kant did not write a single book or chapter explicitly on the social contract, but "his" social contract can be found throughout his readings.

⁸ Kant also has another social contract in which the philosopher argues that the sovereign can only implement laws that reflect the will of the people. Seeing that this social contract does not touch on the manner individuals should interact, it has not been included in this thesis.

⁹ In Kant, property can be possessed by individuals in two manners: physical, or sensible, and 'intelligible' (Williams, 1992, pp. 389-390). The first manner simply refers to the individual holding the object. In the second manner it means that the individual owns the object and that he or she does need to have the object on them to possess it. Both manners need to apply to the object to be able to claim the object as my property.

¹⁰ Kant argues that property is noumenal, meaning that one cannot find the fact of property in empirics (Williams, 1992, p. 391). One cannot ascertain what is mine and what is someone else's via observation, even systematic observation. Williams (1992, p. 391) summarizes this point perfectly: "Kant perhaps senses here that property is not an object, but an institution which depends for its functioning on the observance of a certain system of rules." This means that a government is not only necessary to safeguard private property, it is necessary to enable its existence.

¹¹ Do note that in Kant's second formulation of his imperative lies the foundation of contractualism: 'act in such a way that you treat humanity, whether in your own person or in the person of another, always at the same time as an end and never simply as a means' (Kant, 1785, p. 429). Thus: one should always respect the other.

Table 7: Payoff Matrix of Kant's Social Contract Theory.

	Individual B		
	(A, B)	Consent	Dissent
	Consent	1, 1	-, -
	Dissent	-, -	-, -

4.3.1.1. John Rawls

John Rawls (1971, pp. 10-19, 19-24, 109-112, 118-123) attempts to outline a manner via which the social institutions in a liberal society can be justly set up. Examples of the social institutions include the political constitution, the legal system and the economy. These are referred to as basic institutions as they distribute social and primary goods. It is important to recognize that Rawls assumes that the society is pluralistic and has enough resources. However, there is still moderate scarcity. Rawls reasoning here is that if there is abundance, one can do whatever one wants without harming others, negating the usefulness of justice. When there is only scarcity, abiding justice would mean that one would starve, also making the norms of justice useless. Several elements are excluded because Rawls is creating an ideal theory; examples are criminal justice and other countries.

Rawls (1971, pp. 17-19, 266-267) assumes that the citizens are reasonable and rational. Reasonableness entails that one will abide by the agreed upon terms, even if that goes against their own interests. However, one should only be willing to do so when others are as well. Rawls refers to this as a sense of justice. Rational means that what one deems to be valuable in life is pursued and can change. Rawls refers to these assumptions as the conception of the good.

These abstract qualities of citizens are used in the original position (Rawls, 1971, pp. 10-19, 109-112, 118-123; Ryan, 2012, pp. 424-436). The original position is an instrument with which individuals can contemplate justice. Due to the imaginary quality of the original position, Rawls' theory is a thought experiment. In the position, imaginary representatives take place, representing individuals as free, meaning not subject to nature's deterministic laws, and equal. Those imagined representatives debate justice as the foundational principles of society. Within the original position, individuals are placed behind the *veil of ignorance*. This *veil* prohibits arbitrary facts from influencing the agreement on the principles which will guide society. Knowledge regarding race, class, gender, age, natural endowments, the political system, class structure and many more elements are withheld from the representatives. The representatives are unable to represent the specific situation of the citizen, ensuring that the representatives agree to principles best for all citizens and thus fair to all real citizens. Rawls argues that this constitutes a fair process which in turn can only lead to a fair outcome. Through this thought experiment individuals can come to their own conclusion that Rawls' two principles of justice should be followed.

In Rawls' (1971) theory, individuals will engage with others and create the social contract because he argues that it is the rational thing to do. Entering the social contract is rational because it is the most effective means with which individuals can achieve their goals, i.e. the discussion in which the social contract is created provides individuals the possibility of influencing those principles. It also guarantees that the position of an individual in a society is safeguarded to not be worse than a certain level and will most likely improve. However, in Rawls' theory, the option of not entering into the social contract is never entertained. Therefore, in Rawls' theory, refusing to enter the contract is not an option. The pay-off structure of Rawls' theory becomes how Table 8 portrays it. One refers to the strategy with the highest and only pay-off for the actors.

Table 8: Payoff Matrix of Rawls' Social Contract Theory.

	Individual B		
	(A, B)	Enter	Refuse
	Enter	1, 1	-
	Refuse	-	-

One can see that Hobbes', Kant's, and Rawls' pay-off matrixes are the same; however, it is necessary for this thesis to delve into the argumentation that leads to these pay-off matrixes. Only then is it possible to compare the theories.

4.3.2. Contractualism's Basis

It seems that both Kant and Rawls have one major overlap: both seek principles and moral rules that everyone can support. These are to be grounded in a social contract which is, according to both philosophers, a fiction, i.e. a thought experiment. Rawls' theory is even more abstract than Kant's though because of the sheer impossibility of enacting it. Additionally, contractualists argue that one should not treat a human being as a means, but always as an end in themselves. This in turn means that one should treat others in accordance with the moral rules that everyone can support.

Thus, the biggest difference between contractualism and contractarianism is that, in the latter, a state, is yet to be formed, while in the former, that is not necessarily the case. According to contractarianism, the reason for forming a state or adhering to rules is that individuals try to maximize their own interests. Contrarily, contractualists first assume the rationality and autonomy of agents and, based on that, argue that one can only pursue one's interests while leaving space for others to do the same. The question becomes which is more like Liu's theory? This is investigated in the next part.

4.4. Comparing Liu's Theory to Social Contract Theories

In this part, I compare the social contract theories of Hobbes (Hobbes, Rogers, 2005), Binmore (1994a, 1994b), Kant (1785) and Rawls (1971) to Liu's theory (2008a, 2008b). It becomes clear that Liu's theory is different from all social contract theories but is closest to Hobbes'. Kant's and Rawls' theory are not as grim as Liu's, while Binmore's is very different because of its use of evolutionary game theory. Hobbes' theory is, therefore, also investigated separately from the other three social contract theories, specifically the axioms underlying the theory.

Four out of the five theories focus on resources; however, Hobbes' (Hobbes, Rogers, 2005) and Liu's (2008a, 2008b) theories focus on the necessity of resources to survive. Kant adds that private resources should be protected to ensure the fulfillment of one's life as well. Conversely, Rawls argues that a social contract is necessary solely to safeguard resources and societal positions because of their worth and not their use. According to Liu, this is naïve, because it does not consider the threat of death: even if your resources are safeguarded, you can still die. Binmore's (1994a, 1994b) theory is radically different from the other four theories because it is not built upon the challenges posed by others and property. Binmore argues solely that the social contract is the way everybody interacts because it is in everybody's interest to do so.

All social contract theories, including Liu's (2008a, 2008b), use rational and methodological actors to a certain degree, meaning that they assume the interests and the goal of actors is based on the actors themselves. In Hobbes' state of nature, the interests of two or more actors are most easily conflicting. The actors, however, are never sure if their interests do or do not conflict, resulting in distrust. This might incentivize actors in Hobbes' state of nature to execute a preemptive strike, an option that is also highlighted in the DFT (Baliga, Sjöström, 2010). In game theory terminology: attacking is the dominant strategy of actors. Liu even explains his theory via his character Luo Ji in a very Hobbesian way:

"The universe is a dark forest. Every civilization is an armed hunter stalking through the trees like a ghost gently pushing aside branches that block the path and trying to tread without sound. Even breathing is done with care. The hunter has to be careful, because everywhere in the forest are stealthy hunters like him. If he finds other life – another hunter, an angel or a demon, a delicate infant or a tottering old man, a fairy or a demigod – there's only one thing he can do: open fire and eliminate them. In this forest, hell is other people. An eternal threat that any life that exposes its own existence will be swiftly wiped out. This is the picture of cosmic civilization. It's the explanation for the Fermi Paradox. (Liu, 2008b, p. 521).

In none of the theories is there an authority regulating the decision-making units' actions, allowing them to act as they prefer. For Hobbes' (Ryan, 2012, pp. 424-436) and Liu's (2008a, 2008b) theories, this entails that killing without consequences is possible. Their theories deviate on the question of whether the introduction of an authority would be enough to deter killing. Hobbes argues that an authority can do so via regulations that ensure that not killing the other is in one's own best interest. This creates a tenuous agreement, an agreement by which actors commit to not killing each other if the others commit to doing the same. Liu's reasoning is that this is not assurance enough; there is too much risk involved for both actors. The reason is that both actors' main goal is to ensure their survival and even the slightest hint at a threat is taken seriously and destroyed, if possible.

This difference between Hobbes' and Cixin Liu's theory resonates in the different options that their actors have (Ryan, 2012, pp. 424-436, Liu, 2008a, Liu, 2008b). Although in both theories, actors fight for their survival, they are presented with different options. In Hobbes' case, actors have three options: fight until the death, become the sovereign, or be subjected to a sovereign (Hobbes, Rogers, 2005, p. 138). Actors in Liu's theory can be benevolent, malicious, or keep quiet. In Hobbes' theory, two of the three options lead to survival, while in Liu's theory, only one does for only one involved actor. The difference stems from Hobbes' laws of nature (Ryan, 2012, p. 430-440). The pay-off structure of Hobbes is different than the one of Liu's DFT because the goal of survival of humankind, and consequently of both individuals, is safeguarded by the laws. In Hobbes' theory, the actors need each other to safeguard the future of humankind or their species' survival. In Liu's theory, this is not the case. Cosmic civilizations can, in principle, continue to exist without the help from outside forces. Therefore, the situation as displayed in Table 4 does not arise in Liu's theory and neither does the change in pay-off structure presented in Table 3. Thus, Liu's theory resembles Hobbes regarding the grim outlook in the state of nature, but it seems there are elements in Hobbes that makes it possible to escape this situation that Liu's theory does not possess.

The biggest difference between the social contract theories discussed here and Liu's theory (2008a, 2008b) lies with the comparative strength of the actors. In Hobbes' (Hobbes, Rogers, 2005), Binmore's (1994a, 1994b), Kant's (1785), and Rawls' (1971) theory the actors are roughly similar in strength, while in Liu's theory this is not necessarily the case. Actors can be very different in power. This makes it possible, first, for another actor to kill you and all the others. Second, it might be possible for yourself to eliminate all threats to your existence and thus ensuring your survival on the long term. If civilizations have such power, it is indeed logical for actors to enter into conflict. Conversely, if these situations are not possible in the other theories, then cooperation ensures one's survival in the long term.

5. The Dark Forest Theory Evaluated

In the previous chapter, social contract theories have been introduced to challenge the DFT and it has become clear that there are similarities and differences (Liu, 2008a, 2008b). In this chapter, I continue to determine what parts of the DFT withstand critical evaluation, specifically the DFT's underlying axioms. Determining if humanity should follow the DFT when encountering extraterrestrials should be internally consistent and be able to withstand critical evaluation. To disprove a theorem such as the DFT, one needs, in most cases, only to present a counterexample. Additionally, Liu does not explain how these axioms came to be, or how the reasoning leading to the axioms goes, nor does he explain the interaction between said axioms.¹² There are no facts, no grand theories, or rationale presented to substantiate the axioms. In this chapter I, thus, continue to determine what parts of the DFT can withstand critical evaluation. I specifically analyze the DFT's underlying axioms.

Although not necessarily incorrect, the axioms appear to be the frailest element of Liu's theory. To test these I first reiterate the axioms of and as introduced by Cixin Liu (2008, pp. 515-516), followed by a brief recapitulation of why these axioms lead to civilizations destroying each other. The axioms are:

1. Survival is the most important goal of every civilization;
2. Every civilization will continue to expand and grow, but resources in the universe are limited.

Based on these two axioms, Cixin Liu argues that civilizations will meet and must compete. While civilizations grow in population and the necessary resources to sustain themselves, civilizations will without a doubt meet at a point in time and they will fight. The reason is that when civilizations meet and they are both looking for resources that are scarce, in the economic and physical sense, they will fight over the resources to secure their chance of survival. Many academics like Ayn Rand, Karl Marx, Thomas Malthus, and Garrett Hardin argued along similar lines to explain human conflict (Hicks, n.d.). Liu goes beyond their reasoning and argues that the only manner to ensure survival is to destroy the other civilization. In cases of collaboration and ignorance there is always a chance that the other civilization determines that you are a threat and consequently tries to destroy you. The safest option, that is, the option that brings your chances of survival closest to hundred percent as possible, is to destroy the other civilization first. Or so Liu presumes.

¹² The two sentences leading up to the only concrete mentioning of the axioms are: "Right. So now we're going to set out two axioms for cosmic civilization." (Liu, 2008b, p. 515).

5.1. Intuitive and Max Weber

In this part, I argue that the theory is intuitive and is familiar, mainly because it is in congruence with the current capitalist and liberal way of thinking. Many authors have written on these ways of thinking, Karl Marx, Ayn Rand, Thomas Piketty, and Alexis de Tocqueville, and Max Weber (2002) are renowned examples (Hicks, n.d.). In his writing, Weber critically reflects on the capitalist and liberal way of thinking, not only focusing on the economic aspect of the system, but also on the sociological. By including the sociological aspect, his writing becomes more suitable for reflecting on this thesis' topic. Specifically, sociology studies the problem of how individuals function in a society which can be likened to how civilizations should interact.

The DFT, at its core, argues that the individuals, terrestrials and extraterrestrials alike, will fight over scarce resources when individuals need them to survive. So far, the DFT's propositions make sense, as does the argumentation: it stays in line with social contract theories which rely on the intuitive idea of agreement (Freeman, 2007, p. 17). Freeman argues that within the liberal framework, we are familiar with consent and individual benefit. Together they form the foundation of agreements. The DFT seems to include an element of agreement as well: the moment that civilizations meet there is an agreement among us and among the other party that we should try to destroy each other. Additionally, social contract theories and the DFT reflect how humans intuitively approach situations of scarcity. Individuals in such a scenario think of what it needs first, and what they need to safeguard their life and way of life. The underlying intuitive idea is that an object cannot be owned by two individuals or groups with competing interests.

This line of thinking is supported by the idea of a scarcity mindset (Huijsmans et al., 2019). It suggests that individuals in a situation in which scarcity is the norm pay more attention to the scarce resources, tending to forget all unrelated aspects, like what might happen after they have attained the resource. Huijsmans' research has shown that this mindset even changes neural processes underlying decision making, thus implying that physiologically, it changes individuals, exacerbating their scarcity mindset. Additionally, Liu's theory is not utopian, meaning there is no wishful thinking present. It is dark and grim, making it more in line with how we think than with theories which are utopian. This is reflected by the theory of negative bias, which points out that individuals are more likely to remember negative material in cognitive processes (Reed, Carstensen, 2012).

Next to being intuitively convincing, examples in which the DFT's rationale is applied are abundant. Thus humanity is familiar with applying a rationale that is focused on securing resources and while doing so making others powerless and thus making retaliation impossible. Examples range from enslavement, subjugating, resocialization, massacring, and attempts thereof. These examples are present in humanity's entire history: the enslavement of many individuals from and colonization of the African continent (Fomin, Ndobegang, 2006, pp. 633-634), the attempted massacre of native Americans and the resocialization that followed (Churchill, 2003, Wolfe, 2006), the subjugation of the Middle East by Western countries (Coskun, 2012, pp. 255-257), and the subjugation of large swaths of land and civilizations in Asia and the Middle East by Genghis Khan ("Genghis Khan", 2009). A contemporary example of individuals acting in a line with Liu's theory is that of immigrants and individuals' actions within the capitalist system. Within the current debates regarding immigrants and immigration, many arguments resemble the rationale of the DFT's theory as well (Brimelow, 1995, Schneider, 2008, pp. 62-63). The argument that strangers are stealing jobs and resources, indicate a focus on problematizing the sharing of resources in the immigration discussions.

Max Weber's (2002, Calhoun, Gerteis, Moody, Pfaff, Schmidt, Virk, 2002, p. 165) sociological work provides an explanation for these actions. Weber's work is a critical assessment of how our world works, specifically, it is a sociology of the capitalist system. Weber positions the individual within the capitalist system, which he argues is a Western cultural phenomenon. 'Rationalization' is the concept that Weber introduced to indicate the idealization of purposive-rational action in all areas possible (Habermas, 1987, pp. 81-94). The most important example of this is the capitalist system, but Weber's scope includes the areas of scientific and technological development. The essence of the rationalization is that "we can in principle *control everything by means of calculations* (Weber, 2004, pp. 13, italics in original). To make this possible, the creation of the "person of vocation" is necessary, as well as a secularized and disenchanted world (Weber, 2004, pp. 30-31). This means that in Weber's theory, the individuals are focused on efficiency, effectiveness, and the continuance of the system they find themselves in. He argues that the individuals within such a system are slaves of that system. The same can be applied to whole planets; when civilizations go into space and meet another civilization, they will most likely react as a singular unit.

Because of our current disposition toward the use of liberal forms of governance and government in combination with the capitalist system, I expect that humanity will adhere to the logic set out by Weber (2002, Calhoun, Gerteis, Moody, Pfaff, Schmidt, Virk, 2002, p. 165). Capitalism and liberalism both accept a certain type of violence to that of Darwin's survival of the fittest: violence in the form of competition. There is a certain cold-heartedness about his theory that I believe will apply to humanity's interaction with extraterrestrials. And like in Weber's theory, humanity might be a slave to the system of capitalism and does not have a choice in the matter.

This means that, if extraterrestrials meet humanity now, individuals will ensure the continuation of the capitalist system by aggressively engaging with the extraterrestrials. The reason is humanity's distrust toward, and fear of, the very existence of extraterrestrials because they are not included in humanity's calculations and can possibly form a threat to humanity's goal of ensuring the continuation of its system. Concluding, the DFT seems to be intuitively appealing, to be empirically supported by Weber (2002) and examples, making the DFT's intrinsically plausible.

5.2. Internal Coherence

Based on the arguments provided in the previous part, it is likely that humanity will follow the DFT or create outcomes similar to the DFT's. However, this does not necessarily mean that humanity should follow the DFT. In this chapter, I investigate first the is/ought fallacy, thereby assessing the first axiom (Liu, 2008b, pp. 515-516). The second step consists of assessing the internal coherence of Cixin Liu's (2008b, p. 516) theory. I do so by investigating if the second axiom can be upheld when the first axiom is accepted.

First, although there are counterexamples to the DFT going back 236,000 to 335,000 years, all examples lean on the is-ought fallacy (Stoddard, 2017). Hume (1896, pp. 244-245) first articulated this fallacy as reasoning what one should, or *ought*, to do based on what currently *is* (Black, 1964). Hume found there to be a difference between positive/descriptive and prescriptive/normative statements. The argument being that there is no clear way to go from a descriptive to a normative statement. The underlying argument is what one should do, which is a normative statement linked to ethical questions, relies on logic and that is what a descriptive statement lacks. That we have attacked others in the past or that we are fighting for resources does not mean that we should continue this behavior in the future.

To prevent humanity from making an is-ought fallacy, it is necessary to investigate Liu's axioms to provide an answer to the question of whether humanity should follow the DFT or not. Liu's first axiom entails that every civilization's ultimate aim is to survive; however, does the second axiom guarantee this? Liu's second axiom states that civilizations grow and that resources in the universe are finite (Liu, 2008b, p. 16). Max Weber (2002) has shown that this reasoning is in line with our current mode of production and form of government, i.e. capitalism and liberalism. This does not mean that every other civilization needs to adhere to this mode of production and thus not every civilization necessarily reflects this theory by growing and expanding. The second axiom reformulated according to these insights is:

2'. Civilizations can expand and grow in a system with finite resources.

This is not an actionable axiom though; it does not prescribe action(s). Additionally, it is problematized by combining it with the first axiom: it can be argued that growing and expanding in a system with finite resources is dangerous. The reason is that civilizations increase the risk of being detected by going into space to collect resources and it consequently increases the chance of fighting with other civilizations. Additionally, because civilizations are growing and expanding, there is more reason to fight, seeing that both civilizations have competing interests: both want the resources. Based on this reasoning, civilizations should limit their resource use to limit the necessity to look for resources in outer space and thus risk their survival. Therefore, the second axiom should be turned into:

2''. Civilizations should not (want to) expand and grow, because the resources in the universe are limited.

This version of the axiom could explain why we have not met any cosmic civilizations yet: they remain close to their "home" planet. Multiple answers to the Fermi paradox point in the same direction (Webb, 2015, pp. 111-113). Darwin's theory sets out a similar scenario with elements that limit a civilization's growth, such as food, predation, and diseases. One might argue against this reformulation by stating that civilizations can also go into the universe to secure the resources that they need to survive. Civilizations are, however, never sure that they are the strongest or able to withstand possible attacks from encountered and/or threatened civilizations.

One can argue that there is value to different life forms and that all life forms are just trying to survive, thus arguing that civilizations should safeguard other civilizations and be safeguarded.¹³ Several thinkers from Africa go one step further by arguing that the world is a living thing, that consists of all humans, animals, plants and all other things present in the world (Kodjo-Grandvaux, 2013). Every individual part is and needs to take care of the whole and to do so needs to interact consciously with the other elements. In many philosophies from Africa, this means that one should only kill and eat what is necessary. It is also necessary to keep the balance and live in harmony with the environment because all actions influence the whole and thus also humanity's food sources. Individual civilizations are thus responsible to not only care for themselves, but also for its planets, its co-inhabitants, and the whole of the universe. This means that one should not grow because it will threaten the balance. These insights give rise to a third reformulation of the second axiom:

2'''. Civilizations should not want to grow and expand, because it might threaten the balance of the universe which in turn might threaten their own and other civilizations' survival.

However, civilizations can be prone to explore and conquer other civilizations and their resources, which is the same critique that was formulated in opposition of the second reformulation. A similar negation can be applied to this critique: civilizations cannot guarantee their own survival if they go along such a path. Thus, an isolationist policy, instead of exploring outer space, is better once more.¹⁴¹⁵ These two adaptations of the second axiom, the second and third adaptation, result in the same pay-off structure, which is expressed in Table 9. The number one refers to the strategy with the highest pay-off for the actor, two to the strategy that has the lowest pay-off for that actor.

Table 9: Payoff Matrix for One Civilization of the Second Axiom.

Planetary Isolationist Policy	Explore Outer Space
1	2

If the adapted versions of the second axiom are expressed in a pay-off structure for two civilizations, it looks like the one presented in Table 10.

¹³ Several contemporary movements and groups argue along similar lines for humanity on earth. They advocate sustainable development, assigning rights to non-humans, and circular economy to reach this (Stahel, 2016, von Weizsäcker, 1994, Bocken, Olivetti, Cullen, Potting, Lifset, 2017, Regan, 1983, Singer, 1995, Degrazia, 1999, Sunstein, Nussbaum, 2004, Leib, 2011).

¹⁴ There are instances during which humanity has already employed such isolationist policies. Most notably by Chinese dynasties (Webb, 2015, p. 111, Thornton, 2000, p. 2, Mueller, Wang, Liu & Cui, 2016, p. 77).

¹⁵ The "Voluntary Human Extinction Movement" (n.d.) goes one step further and argues that it might be time for humanity to die off instead of harming its environment.

Table 10: Payoff Matrix for Two Civilizations of the Second Axiom.

	Civilization B		
	(A, B)	Planetary Isolationist Policy	Explore Outer Space
	Planetary Isolationist Policy	1, 1	2, 2
	Explore Outer Space	2, 2	2, 2

5.3. Conclusion

In this chapter, I have investigated the internal coherence of the DFT by accepting the first axiom for now. I reformulated the second axiom to ensure the reaching of the first axiom in three different ways. It became clear that to safeguard their existence, civilizations should limit their expansion. It is thus not possible to uphold the second axiom as formulated by Liu (2008b). Thus, I conclude that the theory's coherency is lacking.

What should be acknowledged though is that when one accepts either reformulation of the second axiom and humanity does meet another civilization that is actually venturing into outer space, one should think of that other civilization as dangerous. The issue lies with the fact that if the civilizations behave morally, they would have not encountered humanity, or any other civilization for that matter, meaning that the exploring civilizations are either a- or immoral. Thus when accepting this reasoning, civilizations should try to destroy civilizations that venture close to their home planet. Although the result is in line with the DFT, i.e. the destruction of the other civilization, the reasoning that led to the result is not the same.

Thus far the second axiom has been investigated; however, one can also question the first axiom: is survival indeed the most important goal of every civilization? Even on earth this is a tough question to answer. Darwinism, one of the most accepted theories about how species survive, does not necessitate species wanting to survive (Lennox, 2015, Darwin, 1861, pp. 60-119). It argues solely that the genes, and accompanying phenotypes, that are most compatible with the environment will be passed on. Not the species themselves, nor the individual gene carriers. This is referred to as 'natural selection' by Darwin himself, in both cases enforcing the idea that it is not the species most wishing to survive that will survive.¹⁶ There are even examples of humans sacrificing themselves to save another individual. Several philosophers, like Jacques Derrida, Emmanuel Levinas, and Jan Patocka argue that self-sacrifice is the highest ethical gesture one can make (Milbank, 1999). One can even go one step further; suicide is present in our societies and there are even cases of animals doing the same (Preti, 2007). It is thus possible to argue that survival as the ultimate goal is neither something that is, nor should be, inherent in every creature.

The question becomes twofold: what should be one's first goal and how should one go about interacting with the Other? The latter is a question that is reflected by many theorists, who might provide an improvement to Liu's theory.

¹⁶ Alfred Russel Wallace refers to this as 'survival of the fittest', however, it seems that Darwin did not accept this kind of formulation of his theory (Lennox, 2015).

6. How should civilizations interact?

In the previous chapter, we have seen that it is impossible to uphold both the DFT's first and second axiom. To remedy that, multiple reformulations of the second axiom were investigated. These reformulations relied on accepting the first axiom, so the next task is to investigate the first axiom. That the DFT seems to be partly supported by an is-ought fallacy strengthens the relevance of such an inquiry. The goal of the task is to increase the DFT's internal coherence and foundation. I, therefore, search for improvements for Liu's (2008a, 2008b) theory by investigating how other thinkers have argued that entities should go about interacting with each other. The other in these thinkers' theories is not as radical as extraterrestrials, because there are most probably less commonalities between a human and other species than between humans. The theories are, however, still applicable to extraterrestrials because the theories either note how every organism in the universe should behave or how one should behave regardless of the other's nature. In this chapter I thus present divergent theories on how individuals should interact, which may provide an improvement of Liu's theory. The theories are translated in axioms resembling Liu's, as well as in a pay-off matrix, enabling me to test the theories and reflect on its potential considering Liu's DFT.

The theorists investigated are those of Thomas Aquinas (2006), Immanuel Kant (1785), Friedrich Nietzsche (Kenny, 2010, pp. 935-939), and Jürgen Habermas (1987, pp. 92-94). Aquinas argues that the goal is to follow the virtues of God to go to heaven in the afterlife, while Kant argues that the end of one's action should be acceptable to everyone and that humans should be seen as ends, rather than a means. Under Nietzsche's theory, an individual's goal should be to maximize power. Habermas argues that individuals should communicate to find a morality that works for both parties. These theorists have been chosen considering that this thesis is exploratory; their theories are best known and exemplar of many other theories. Because these theories are well-known, it is likely that they will be applied the quickest and thus should be tested first. Additionally, because they are so well-known, they form the groundwork of many other similar theories; by considering these theories many other theories usefulness is tested. This enlarges the relevance of this thesis as well as focuses future research.

The theories are presented chronologically: Aquinas', Kant's, Nietzsche's, and Habermas'.

6.1. Thomas Aquinas

Although Aquinas (2006, qq. 2-3, 61-62, Kenny, 2010, pp. 312-313, Hibbs, 1988, pp. 61-62, Floyd, n.d., Jenkins, 1997, pp. 136-138) was a Christian thinker, he deems his ethics and rules also applicable to non-Christians. Aquinas argues that God has created an eternal law: all things are created to be able to meet their end and have the capacity and the want to do good. Actors need reason to distill the laws of nature from the eternal law and the law is for that reason objective and universal. This is the law that is part of us and will guide us to flourish as humans and can be translated into the sentence: one should do good and avoid doing of evil. To do good, one needs to follow the virtues set out by God, which include the theological virtues consisting of faith, hope, and charity, and the cardinal virtues consisting of prudence, justice, fortitude (or courage), and temperance. Applying these virtues to the first contact case as presented by Cixin Liu (2008b) would mean that the cosmic civilizations would do good. This means that the first axiom is changed into:

- 1a. Ensuring the act of doing good and avoiding evil, by applying the seven virtues, is the most important goal of every civilization.

Aquinas argues that part of being good is to not be overly fearful, especially because he argues that there are good deeds to be done that are more important than personal safety (Floyd, n.d.). The chain of suspicion is a construct that is very fearful and should therefore not be used to base decisions on. Prudence should be interpreted as doing the act that is most responsive to the situation at hand and aimed at reaching your goal. With temperance, Aquinas referred to not indulging in bodily pleasure too much to refine the way that we enjoy the bodily pleasures. Courage is needed to combat our excessive fears. Justice is the sole virtue that concerns the relation between us and the other, meaning we should provide the other with what they deserve. One should also consider the general precepts like do not defraud and do not commit adultery. Based on these virtues, the civilizations should seek contact, because it is only possible after establishing contact what the other is due. However, one should do that with prudence and without too much fear for the unknown. Destroying the other is therefore only an option after having established contact and will not be considered an option by Aquinas. This means that the pay-off structure for Aquinas' theory applied to outer space would look as portrayed in Table 11. One refers to the strategy with the highest pay-off for the actor, and two refers to the lowest.

Table 11: Payoff Matrix of Aquinas' Logic applied in Space.

	Civilization B		
	(A, B)	Contact	Leave Alone
	Contact	1, 1	2, 2
	Leave Alone	2, 2	2, 2

An aspect to consider regarding the foundation of Aquinas' theory is that it is built on the Bible. This is a book that is believed by many to not have any value as a theoretical or ethical foundation. Additionally, the book consists of stories and is in most cases neither coherent nor are its precepts supported by logic. Humans, thus, should not follow the course of actions prescribed by Aquinas.

6.2. Immanuel Kant

Immanuel Kant's ethics oppose theories that make happiness individuals' ultimate purpose, such as Aquinas' (Kenny, 2010, pp. 698-700). The solution is the formulation of a theory that applies to everyone by that is characterized by reason. Kant goes as far as stating that all creatures that are 'rational' make and are subject to universal laws. Intelligent extraterrestrials would most likely have a rationale; otherwise a civilization would be difficult to set up, and would thus should adhere to the theory (Kant, 1785, p. 438).

Kant's (1785, Kenny, 2010, pp. 698-700) central notion is the will. When one has a will; to act would mean to act according to some principles. Kant refers to principle as maxims which are always subjective and are always aimed at reaching an end or goal. Two categories exist; material and formal maxims. The first entails that one acts to satisfy one's desire(s), while the second should always be lived up to, regardless of what one desires. The formal maxim can also be referred to as categorical imperatives; this is the category to which moral laws belong to because they should be upheld always. If one does not, then one is acting immorally.

Kant (Kenny, 2010, pp. 689-700) argues that one can choose to let desire determine one's behavior, meaning that individuals are free creatures and can thus make their own choices. This, in turn, makes it also possible for rational creatures to choose their own maxims. When choosing maxims, one does need to assess the morality of the chosen maxim. One can do so by asking the question if everyone should follow the maxim or 'act only according to that maxim whereby you can at the same time will that it should become a universal law' (Kant, 1785, p. 421). Kant reformulates the imperative to reflect that humans have value in themselves, whereby humans can only be ends-in-themselves. The second formulation is thus: 'act in such a way that you treat humanity, whether in your own person or in the person of another, always at the same time as an end and never simply as a means' (Kant, 1785, p. 429). The axiom reflecting this line of thinking is:

- 1b. Civilizations should act according to that maxim that it thinks should become a universal law, taking into consideration that rational beings are considered an end and not a means.

Kant's ethics diametrically oppose Liu's, making it impossible that the latter is amended based on the first. This difference stems from Kant's rule making the destruction of other civilizations impossible. Additionally, being left alone is not a rule that all civilizations would always want to adhere to; for it might be the case that a civilization needs help from another civilization. The maxim that can become a universal law would most likely advise the civilizations to work together. To put this in a pay-off structure means that destroying the other results in the lowest pay-off, leaving the civilization alone is not the best and not the worst option, and working together has the highest pay-off. This means that the pay-off structure would look like as portrayed in Table 12. One refers to the strategy with the highest pay-off for the actor, two to the second highest, and three to the strategy with the lowest pay-off.

Table 12: Payoff Matrix of Kant's Logic applied in Space.

	Civilization B			
	(A, B)	Befriend	Leave Alone	Destroy
	Befriend	1, 1	2, 2	3, 3
	Leave Alone	2, 2	2, 2	3, 3
	Destroy	3, 3	3, 3	3, 3

The issue with Kant's ethics for this thesis is that they lack content, as also stated by Hegel (1991, §135). They do not point to a singular answer; multiple are possible. This is also the case for this thesis.

6.3. Friedrich Nietzsche

Friedrich Nietzsche's theory is built on the belief that life is the supreme value and to safeguard this, one needs power (Kenny, 2010, pp. 935-939). This will to power is present in every living thing and refers to using power to reach the limits of one's maximum power. Every living thing's goal is to intensify its life to feel alive which includes extraterrestrials, if or when found, as well. This principle is referred to as the Will to Power.

What power should be defined as according to Nietzsche is still debated. Some argue that it is power to overcome obstacles (Reginster, 2006, Katsafanas, 2013) or a yearning for growth and similar terms like strength and expansion (Schacht, 1983, Hussain, 2011).¹⁷ Both these are internal conceptions for both are character traits. The power that one should have can be used to create oneself, to be the best that they can be (Leiter, 2015). What the best precisely means is not clear because it depends on the individual. However, the individual that has reached that stage is referred to as *Übermensch* and it can be said that it is "the ultimate affirmation of the will to live" (Kenny, 2010, p. 937). The stage of the *Übermensch* is constantly evolving and to reach the next iteration of oneself, one needs power. Individuals therefore will continuously search for power. Nietzsche was a proponent of war, he said that it enlightens individuals regarding freedom. All individuals can be liberated via war and thus reach their pinnacle of power. Nietzsche describes himself as an immoralist and thinks that by not imposing a moral system on humanity it can flourish. Based on Nietzsche's theory the first axiom should be changed to:

- 1d. Ensuring the continuous expansion of power is the most important goal of every civilization, by any means necessary.

This axiom is grimmer than Liu's, for Nietzsche argues that a civilization needs to garner power to try reach its full potential (Kenny, 2010, pp. 935-939). This need to garner power is a never-ending process. If two civilizations meet, one can assume that Nietzsche would require that the other civilization is destroyed because it could be in the way of garnering power. It is also possible that the first civilization needs the other civilization to fulfill its potential. A reflection of both interpretations can be found in Table 13 and Table 14. In the last option, reflected in Table 14, leaving alone the other civilization also means no help from that civilization to help its full potential. However, there is still the possibility that cooperation can happen in the future. The number one refers to the strategy with the highest pay-off for the actor, two to the second highest, and three to the strategy with the lowest pay-off.

¹⁷ There are many more interpretations of Nietzsche's Will to Power doctrine, such as it being his version of foundational metaphysics (Jaspers, 1965, Nehamas, 1985, pp. 74-105, Soll, 2015, Anderson, 2012, Clark, 2000). These interpretations do not hold relevance for this thesis and will therefore not be introduced, nor mentioned in the main body of text.

Table 13: Payoff Matrix of Nietzsche's Logic applied to Space, option 1.

	Civilization B			
	(A, B)	Befriend	Leave Alone	Destroy
	Befriend	2, 2	2, 2	3, 1
	Leave Alone	2, 2	2, 2	3, 1
	Destroy	1, 3	1, 3	3, 1 or 1, 3
Civilization A				

Table 14: Payoff Matrix of Nietzsche's Logic applied to Space, option 2.

	Civilization B			
	(A, B)	Befriend	Leave Alone	Destroy
	Befriend	1, 1	2, 2	3, 3
	Leave Alone	2, 2	2, 2	3, 3
	Destroy	3, 3	3, 3	3, 3 or 3, 3
Civilization A				

It is impossible to adhere to Nietzsche's ethics seeing that there is no solid base underlying the theory, solely the unsupported assumption that all creatures aim to grow in power. Not only adhering to the theory is impossible, also amending Liu's theory based on the theory is impossible because of this reason.

6.4. Jürgen Habermas

Jürgen Habermas (1987, pp. 92-94) builds on Weber's theory. Habermas argues that humanity should transcend capitalist and liberal way of interacting and get rid of our slave mentality. Habermas argues that rationalization as set out by Weber is not enough to capture humanity's norms, specifically that humans and their interaction can be captured by calculations. Instead, the philosopher argues, one should communicate with the other to form social norms together. Habermas tries to steer humanity in a direction in which interaction is more important than the technical rules whereby humanity can "define reciprocal expectations about behavior" (Habermas, 1987, p. 92). This form of communication can liberate humanity from rationalization and calculations.

Habermas' communication form is a rationality that in which practical communication is central (Bohman, 2014, Dillon, 2014, pp. 212-216). It is practical because it maps what a knowledgeable social actor should do when interacting with actors. The actors need to be knowledgeable to be able to speak and understand speech acts, which are necessary to come to a common understanding. The difference with past modes of communication is that the goal is not strategic but aimed at sustained communication. In this setting a speech act is accepted by the receiver when the message fits within what they deem acceptable, only then a consensus can be reached. Concluding, according to Habermas' communicative rationality theory the first axiom should be:

1d. Ensuring that all actors involved communicate and try to find successful speech acts on which reciprocal expectations about behavior can be defined.

This axiom is radically different from the axioms constructed based on the other theories, and even antagonistic to the DFT's. The theory does away with strategic communication and makes it impossible to put in a pay-off structure based on game theory that is founded on strategic communication. What is clear is that the civilizations are required to communicate with each other. What the results of that communication are going to be is based on the consensus found. We can only guess what that consensus will entail at this moment in time. This solution is, however, problematized in the next part.

What becomes clear is that one needs to accept that a civilization handles a certain reasoning, or ethics, to ascertain if the DFT holds up. Conversely, Liu (2008a, 2008b) does not provide any substantiation which one should be upheld. The question then becomes, what reasoning and/or ethics does humanity adhere to and should it? And do we also adhere to them in the case of meeting another cosmic civilization? The movie *Arrival* (Shawn, Linde, Levine & Ryder, 2016) shows us that humanity itself can be divided regarding the best option vis-à-vis the interaction with extraterrestrials. The question that humanity as a whole needs to ask itself: what is the goal of humanity and its civilization? Existentialists have been trying to answer that question on an individual basis, but it should also be asked on a civilizational, and thus global, level. The answer to the question differs person to person and the whole of humanity is different than the sum of its population, and the singular individual undergoes ethical harm when doing so (Large, 2015, pp. 23-24).

6.5. Communication

Within all the theories discussed, communication is central. Rawls (1971, p. 16) shows this by explaining that social contract theories translate the problem of justification to a problem of deliberation. Communication is to be understood as the exchange of information which can occur through mediums like speaking and writing, which in turn can be transmitted over long distances via electronic devices. Communication then makes it possible to communicate our situation, our knowledge, and our intentions. In social contract theories it is held to be the manner via which one can overcome a situation that is sub-optimal (Hobbes, Rogers, 2005, Locke, 1970, Rawls, 1971). First, one needs to communicate one's needs and conditions that must be incorporated in the contract, for example survival or the safeguarding of one's private belongings, and second, the contract is itself a medium to communicate one's intentions. Some theorists go even further than the social contract theories, like Habermas (1996), as well as Fishkin (2011) and Dryzek (2000).

Liu (2008b), however, argues that communication's capabilities to overcome the coordination problems central in social contract theories are hindered in outer space. He argues so because the distances between the communicating parties are too vast. This would make all the included social contract theories just discussed impossible. I therefore first delve into how according to Schelling (1960, pp. 100-101, Dixit, 2006), a coordination game is changed by communication in this section. Second, I explain how Liu (2008b, p. 519) argues that communication does not provide an escape of the coordination problem. This is followed by a third, brief, consideration of the possibilities of communication in space.

Schelling (1960, pp. 100-101) argues that communication will make it so that the 'game' falls apart and concedes to bargaining in the case of a mix motive game. He argues that mix motive games are a form of a coordination issue with communication challenges. Coordination issues can then be solved via the introduction of the possibility of contact if "players can concert with certainty, without difficulty, and without cost" (Schelling, 1960, pp. 100-101). There are several requirements that a game should fulfill for communication to be effective. These are 1) individuals are inhibited from their preferred action and are unable to communicate perfectly, and 2) actions should matter, they can either be symbolic or they can change the significantly by changing the situation.¹⁸

Liu (2008b, p. 519) acknowledges the possibility of communication but argues that communication is not relevant for interactions in space. His argumentation rests on the chain of suspicion that is relevant in space but not on earth. Liu argues in his book that the chain of suspicion:

"[i]s something that you don't see on earth. Humanity's shared species, cultural similarities, interconnected ecosystem, and close distances means that, in this environment, the chain of suspicion will only extend a level or two before it's resolved through communication" (Liu, 2008b, p. 519)

These conditions are already difficult, but possible according to Liu, to meet on earth, but are impossible to meet in space. The vastness causes communication to take a long time, thereby allowing the chain of suspicion to become too long. This can cause the better pay-off to shift from not doing anything to destroying the other to ensure one's survival. Without communication,

¹⁸ Schelling (1960) introduces the mutual motive game, the mixed motive game, and the pure motive game as categories to indicate when a game theory is a coordination problem. If one wishes to know more about this, I kindly refer to Schipper 2001 article.

distrust grows till one determines it is necessary to take the initiative and attack, because this banishes uncertainty. This decision needs to be made relatively quickly because the civilizations are unsure what the result will be of the communication and it is the civilization that has its ships closer to the other can act quicker and is thus in an advantage position. In Schelling's words, the distance is big enough to become a cost and brings back uncertainty. Concluding, communication cannot end the chain of suspicion in outer space. This also entails that Habermas' theory does not hold, seeing that it relies on communication's ability to overcome coordination problems, which it loses in outer space.

Considering Liu's assumption that it is possible to destroy civilizations in space means there is less of a possibility to communicate as well. What happens, however, if this is not the case? What if the civilizations cannot easily eradicate the other civilization? Even though planets might be destroyed easily, ships can hide, and new planets inhabited. This could mean a prolonged conflict due to the vast distances. Communication could then be established even if it just resembles an armistice. If civilizations conclude that clashing is more expensive then they will leave the other alone, for an uncertain period. However, in such a situation a technological leap can shatter the armistice, just like a crack in the defense of other civilization can.

Different languages and reference points problematizes communication even more. Even the most basic question of what kind of building blocks extraterrestrials are constructed from remains unanswered (Morris, 2003, Dessy, n.d., Ossola, 2017). Humankind and all life we now know is built out of from carbon, but extraterrestrials might have a foundation build from silicon, or liquid ammonia, or another material.¹⁹ This means that they might have a completely different form of communication. Additionally, Quine (1960, pp. 26-79) shows that one should be familiar with the whole of the other's society, meaning its ideology, its members, its environment, to be able to correctly translate messages. Spatial and temporal contexts matter, as well as traditions, and individual and collective identities. The critiques of spatial and temporal context are especially valid when looking at the case that is researched in this thesis. This is impossible to become familiar with during first contact. The difficulties of these elements are brilliantly shown in the science fiction movie *Arrival* (Shawn, Linde, Levine & Ryder, 2016). This means that one should either be familiar with the other's rationality to understand their language, or already know their language from which one can discern their rationality. Seeing that humanity and extraterrestrials have information on neither it seems impossible to communicate with extraterrestrials.

Second, the theories require the other to be rational, what would happen if the other is not rational though? Would communication be possible and if so, how will it communicate? Even if it can communicate and make itself understood, is it possible to reach a rational conclusion if one of us is not rational? These questions still rely on the dichotomy between rational/irrational that has been constructed by humanity and uses humanity as its reference point. It is, however, impossible to determine if extraterrestrials employ a rationality and if it is similar to humanity's. In this paper these questions are too big to entertain though.

¹⁹ The influence of such a difference is interesting to investigate seeing that it could entail that the conflict over resources is limited, seeing that the extraterrestrials and humanity would be interested in different resources.

6.6. Conclusion

In this chapter, different theories that outline what one should do when meeting another have been discussed. I first investigated this by delving into the theories of Aquinas, Kant, Nietzsche, and Habermas. Their theories were outlined and then translated into an axiom that would fit in the DFT. There were two goals: first to challenge the first axiom and second to find a theory based on which an axiom could be formulated that could withstand critical philosophical inquiry. I found that solely Habermas' theory seems to stand strong. Both Aquinas' and Nietzsche's theory lack a strong foundation. Aquinas' theory is based on a book that not all individuals believe in, while Nietzsche's theory is not based on facts, but on his belief that power is necessary. He does not even give a definition of power. Kant's theory is diametrically opposed to Liu's and provides no answer to our question because of its lack of content. Since it was impossible to determine what to do at that point, humanity should follow Habermas' (1987, pp. 92-94, Bohman, 2014) theory because it solely prescribes to communicate.

The problem with all, however, turned out to be their reliance on communication. In outer space communication is problematized by the vast distances between the civilizations. Not only establishing communication is a problem in space, it is also problematized by the lack of similarities in spatial and temporal contexts. Quine (1960, pp. 26-79) put forth why: to properly understand the other one, needs to be knowledgeable of the other's entire world, literally and figuratively, to make precise translations. In outer space this type of information is more difficult to garner than on earth. This means that civilizations cannot communicate their intentions. Humanity's focus on rationality is also complicated because of this, seeing that language is the manner with which one expresses the way that one thinks, i.e. one's rationality. These insights problematized Habermas' theory most, particularly because his reliance on the civilizations communicating their moral argumentation. However, moral argumentation depends on the specific context of the individuals involved (Stryker, 2000, pp. 215-223). It is therefore impossible to rely on any of the theories included here and the reformulations of the first axiom.

7. Conclusion

This thesis set out to research whether humanity's civilization should interact with other cosmic civilizations based on Cixin Liu's (2008b) Dark Forest theory. While there is a chance that humanity meets extraterrestrials, indicated by the Fermi paradox, humanity might use insights from this thesis to model future interactions between humans. To investigate, test, and to improve the DFT game theory has been introduced. By comparing the DFT to other social contract theories, I found that it was indeed new approach, after which I tested the internal coherence and the foundation by delving into the axioms underlying the theory. Afterward other social contract theories were explored to possibly improve the DFT.

What can be concluded is that the DFT does not provide us with a coherent answer to the question of how we should interact with extraterrestrials. It does, however, provide us with challenges to current social contract theories that they are unable to answer. The cause are the difficulties faced when trying to communicate in outer space, specifically the vast distances. This means that most social contract theories do not hold in outer space. Habermas' theory, as the most promising theory, was therefore also discounted. What I was able to discern as well is that humanity will rely on an is/ought fallacy when interacting with extraterrestrials, meaning that humanity will apply what they have experienced to what will happen without questioning its ethical viability. This means that DFT's outcome will become reality; however, humanity will not follow the logic set out by the theory.

Additionally, it was concluded that if a civilization is indeed keen on surviving, it should not explore outer space. The reasoning was that exploring will increase its presence and thus the likelihood of discovery and war, while also possibly harming other civilizations. This means that if a civilization does explore, it is either a-moral or immoral according to this rationale. Such a civilization might consist of locusts or be a civilization that does not think about morality at all. These options in turn mean that if humanity meets such a civilization, it should be wary of them and should defend itself to ensure its survival. Such a reaction will be purely out of self-defense and not a confirmation of the DFT. Additionally, this entails that humanity's search for extraterrestrial intelligence (SETI) should be ceased. The broadcasting of our existence threatens our civilization and therefore should be deemed immoral considering the presented rationale.

Applying this thesis' results to earthly interactions means applying the DFT and Habermas' theory to, for example, the refugee crisis, neocolonialism, and international relations. First, the DFT is not applicable to earth because the theory's incoherence. However, it would be interesting to research how many earthly interactions can be explained based on the DFT's way of thinking, even with its incoherence. I already gave some examples of probable cases, but an in-depth case analysis might prove insightful. Many possible cases do become difficult due to history that has led to those situations. An example is the case of refugees fleeing to Western countries; for who ventured into whose space first? Who should be considered the dreaded extraterrestrial? Who is im- or a-moral? Second, Habermas' theory has been applied to earthly interactions often, specifically to neoliberal way of thinking. Most issues of Habermas' theory are related to power-distributions which are caused by Habermas' focus on communicative action. His theory, for example, denies disempowered groups and countries to use their strongest political tool, i.e. demanding that their interests are also considered.²⁰ Neither is his theory actionable and thereby

²⁰ If one wishes to know more about the critique on Habermas' communication theory, I advise them to read Purcell's (2009) article. In the article, Purcell sets out several critiques in a clear and concise manner.

difficult to apply. What can be concluded is that earthly interactions should not be guided by the DFT for the same reasons that they should not be applied to outer space interactions. Furthermore, Habermas' theory should not be applied because it is not actionable and too sensitive to power-distribution.

The most obvious direction for future research is investigating other social contract theories and theories that investigate how individuals should interact. The goal of such research would be to make axioms amended by me more coherent and reliable and/or to identify more criteria that should be applied to such axioms. When doing so researchers, need to take the issue of communication seriously and investigate possible ways to deal with the difficulties outlined by the Dark Forest theory. This already points to other interesting research paths, for example the inclusion of philosophy of language which specializes in communication, which in turn necessitates the inclusion of philosophy of mind to investigate the relationships between body and mind.

An element that should be researched in the future is how we have constructed the other and ourselves. As was indicated in the introduction, science fiction can help philosophers by estranging or dramatizing societal themes, or create probable future ones. It can help us think about how we have constructed ourselves by posing interesting philosophical and moral dilemmas. On the one hand, the meeting of aliens will necessitate the reconstruction of ourselves and consequently history as well. It is impossible to say definitively what human identities are constructed; what we can say is that memory and how we think of ourselves is important in the process of creating, constructing, finding, or accepting one. In history as it is predominantly studied, we are the center of the descriptions of the passing of time. What if humanity's position changes? Our memories will be tainted by the fact that it is missing two glaring aspects: the inclusion of another intelligent lifeform and the acknowledgement that we are not at the center of history. Humanity will need to create new ideas of what it is and along the way rationality will be challenged by the creation of new ideas that imply new forms of rationality. This is a question I could not and will not answer in this thesis, but one that I am very much looking forward to reading about and possibly doing research on. The quest for new rationalities' importance can be found in Weber's statement that we are but slaves of the capitalist system. An example is that, if you ask individuals what they would do when they do not need to work as much as they do now, many are afraid because they distill their meaning in life from their work (Piper, 1957).²¹ Meeting aliens might help us deconstruct this view and reconstruct a new one. Considering Clark's (2001) idea, this reconstruction becomes even more important. He argues that to be able to truly know humanity, it needs another sentient species to be compared with. Meaning that we need aliens to truly know ourselves to create a true construct of ourselves.

Conversely, science fiction can help us think about how we have constructed the other. This will become necessary once we meet intelligent extraterrestrial life. Not only because we believed we were the only intelligent lifeform in space, but also because the other can influence humanity's thinking about itself. Levinas investigates this thoroughly, arguing that the relation between actors is what makes the self-exist, which is dissimilar to Lacan's (1977) intersubjectivity theory (Bevan, Werhane, 2011, p. 52). In the latter the other is needed to become aware of one's self or constitute it, while in Levinas' theory the other is needed because the self would otherwise not exist at all.

²¹ Goethe (2012) was also astounded by this mindset. This becomes most clear from this quote: "Most people work most of the time in order to live, and the little freedom they have left over frightens them so, that they will do anything to get rid of it." (von Goethe, 2012, May 17th).

This is a like Clark's argument, as both point future research toward investigating and possibly creating new ways to think about humanity and its interaction with others. Science fiction can do so better than other forms of literature because it is less obstructed by what humanity is experiencing now, it can experiment more and thus test humanity's ideas and concepts more critically. Regarding philosophy, and science in general, science fiction places back the human back in the center of its thought experiments, making it more attuned to the ways humanity has formed and referred to itself and the way the other implicates humanity. Thus further research should consider science fiction as philosophical thought experiments and not as literature that can be read philosophically.

The tool used in this thesis, i.e. game theory, suffers from a similar downside in that it focusses solely on the one making the choices. This stems from the fact that it does not necessitate the opening of the others' black box. It does not assume if the other is rational or not; it requires the actor to consider all possibilities of actions regardless of the other being rational or not. The sole actor for whom it makes assumptions is the actor for which the game is modelled. This means that game theory cannot help us learn more about the other amongst which are extraterrestrials. Game theory can therefore also not fully answer the question regarding the interaction between humanity and extraterrestrials. Humanity does not base its action on just one set of rationality and moral norms and values, as is illustrated by including theorists like Nietzsche, Kant, and Aquinas.

Furthermore, there are scenarios that are different from the ones described and discussed in this thesis; it is for example possible that a plant covers the entire planet or that the extraterrestrials are under the control of a hive-mind. It is unclear in what category we should register both, are they rational or not, is it one organism or multiple? What becomes clear when asking these questions is that all the previous lines of reasoning assume that the other civilization wields some sort of a logic, a rationality. For example: they attack me and thus I need to defend myself to survive. Currently, that would mean that we argue that the other civilization is rational. If it did not defend itself, we would say it is not acting in its own interest, it will be destroyed, and is to be considered irrational. However, what if extraterrestrials and their civilization are a box on its own: it is not rational, or irrational, or somewhere in between. It has found a separate box that it should be placed in. I would not know how to refer to such a pattern of thoughts seeing that they fall outside our current established reference system for these patterns. Game theory as presented in this thesis will be a self-fulfilling prophecy in these scenarios because it does not open the black box of the other, i.e. it is likely that game theory's predictions have a causal effect on the predicted outcome. This means that one cannot infer predictions based on game theory and thus that game theory is perchance not the best tool to investigate how civilizations should interact.

In the previous paragraph the topic of rationality was only briefly touched on in relation to game theory by investigating not opening the others' black box. Delving deeper in humanity's definition of rationality means that by typifying a creature as rational or irrational, humanity's dichotomy of rational/irrational is applied. This dichotomy is problematic when interacting with extraterrestrials, seeing that they might rely on a different rationality than humanity does or not rely on rationality at all. For example, extraterrestrials can just react on instinct or that it solely reacts based on chemical changes in its body. When humanity then argues that the extraterrestrials are not rational, it becomes impossible to gauge what the extraterrestrials' interests are and what they will do. Thus it becomes impossible to determine what humanity should do. However, it would be very interesting to research interactions with non-rational, irrational, or other-rational life forms that can have an impact on the whole of humanity. Answers

to such research questions would reflect back on the ideas that humanity has about itself and how it constructs itself. This ties into future research paths that investigate language and how communication between civilizations can and/or should go. This comes about because language is most often the mode of communication used to relay the kind of rationality one uses and one's language is in turn influenced by one's rationality. Further research into this topic should actively engage with these questions regarding rationality, language, and the Other. This means that further research can change our vision of ourselves and the Other, which is not only interesting, but exciting as well!

8. Bibliography

- Chinchillax-. (2016, September 18). *Death's End [Discussion Thread and Final Thoughts] [Massive Spoilers]* [Online Forum]. Retrieved on May 30, 2018 from: https://www.reddit.com/r/threebodyproblem/comments/54tofo/deaths_end_discussion_thread_and_final_thoughts/.
- Alexandrawallace69. (2018, April 14). (Spoiler) What I really liked about Dark Forest [Online Forum]. Retrieved on May 30, 2018 from: https://www.reddit.com/r/threebodyproblem/comments/8c453v/spoiler_what_i_really_liked_about_dark_forest/.
- Andersen, R. (2017). *What Happens If China Makes First Contact?* Retrieved on June 12, 2019 from: <https://www.theatlantic.com/magazine/archive/2017/12/what-happens-if-china-makes-first-contact/544131/>.
- Anderson, R. L. (2012). The Will to Power in Science and Philosophy. In H. Heit's, G. Abel's, and M. Brusotti's (Eds.) *Wissenschaftsphilosophie: Hintergrunde, Wirkungen and Aktualitat*. Berlin: De Gruyter.
- Angrist, J. D., Kugler, A. D. (2003). Productive or counterproductive? Labour market institutions and the effect of immigration on EU natives. *Economic Journal*, 113, pp. 302–337.
- Apt, K. R. (2015). *Strategic Games*. Retrieved on May 4, 2018 from: <https://homepages.cwi.nl/~apt/stra13/stra13.pdf>.
- Aquinas, T. (2006). *Summa Theologica, Part I-II [Pars Prima Secundae]*. (Fathers of the English Dominican Province, Trans.). New York: Benziger Brothers. Retrieved on June 5, 2018 from: <https://www.gutenberg.org/files/17897/17897.txt>.
- Baliga, S., Sjöström, T. (2010). The Hobbesian Trap. In M. Garfinkel's and S. Skaperdas' (Eds.) *Oxford Handbook of the Economics of Peace and Conflict*. Oxford: Oxford University Press.
- Belletto, S. (2009). The Game Theory Narrative and the Myth of the National Security State. *American Quarterly*, 61, 2, pp. 333-357.
- Bevan, D., Werhane, P. (2011). Stakeholder Theory. In M. Painter-Morland's & R. Ten Bos' *Business Ethics and Continental Philosophy*. Cambridge: Cambridge University Press.
- Binmore, K. G. (1994a). *Game theory and Social Contract, Vol. 1: Playing Fair*. Cambridge, MA: MIT Press.
- Binmore, K. G. (1994b). *Game Theory and Social Contract, Vol. 2: Just Playing*. Cambridge, MA: MIT Press.
- Binmore, K. G. (2007). *Game Theory: A Very Short Introduction*. Oxford: Oxford University Press.
- Binmore, K. G. (2014). Bargaining and Fairness. *Proceedings of the National Academy of Sciences of the United States of America*, 111, pp. 10785-10788.
- Black, M. (1964). The Gap Between “Is” and “Should”. *The Philosophical Review*, 73, 2, pp. 165-181.

- Jackson, P. Cunningham, C. (Producers) & Blomkamp, N. (Director). (2009). *District 9* [Motion Picture]. USA: TriStar Pictures.
- Bocken, N. M. P., Olivetti, E. A., Cullen, J. M., Potting, J., Lifset, R. (Eds.). (2017). Exploring the Circular Economy [Special Issue]. *Journal of Industrial Ecology*, 21, 3, pp. 471-795.
- Bohman, J. (2014). Jürgen Habermas. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*. Retrieved on August 14, 2018 from: <https://plato.stanford.edu/entries/habermas/>.
- Borjas, G. T. (1994). The Economics of Immigration. *Journal of Economic Literature*, 32, 4, pp. 1667-1717.
- Brimelow, P. (1995). *Alien Nation: Common Sense about America's Immigration Disaster*. New York, NY, Random House.
- Brunero, J. (2012). Instrumental Rationality, Symmetry and Scope. *Philosophical Studies*, 157, 1, pp. 125-140.
- Burchell, M. J. (2006). W(h)ither the drake equation? *International Journal of Astrobiology*, 5, 3: 243–250.
- Burgess-Jackson, K. (2012). Taking Egoism Seriously. *Ethic Theory and Moral Practice*, 16, pp. 529-542.
- Calhoun, C., Gerteis, J., Moody, J., Pfaff, S., Schmidt, K., Virk, I. (Eds.). (2002). *Classical Sociological Theory*. Malden, MA: Blackwell Publishers.
- Capova, K. A. (2016). The New Space Age in the making: Emergence of exo-mining, exo-burials and exo-marketing. *International Journal of Astrobiology*, 15, 4: 307-310.
- Christopher, E. M. (2017). The Geopolitics of Immigrant Labour: A Climate of Fear. In B. Christiansen & F. Kasarci's (Eds.) *Corporate Espionage, Geopolitics, and Diplomacy Issues in International Business (210-231)*. Hershey, PA: IGI Global.
- Churchill, W. (2003). An American Holocaust? The Structure of Denial. *Socialism and Democracy*, 17, 1, pp. 25-75.
- Clark, D. L. (2001). Kant's Aliens: The Anthropology and Its Others. *The New Centennial Review*, 1, 2, pp. 201-289.
- Clark, M. (2000). Nietzsche's Doctrine of the Will to Power: Neither Ontological nor Biological. *International Studies in Philosophy*, 32, 3, pp. 119–135.
- CNET. (2018). 'Annihilation' director Alex Garland chats with CNET about the upcoming film [Video File]. Retrieved on 10 July, 2019 from: <https://www.youtube.com/watch?v=nYhT5Ey42gg>.
- Cole, A. (2017). Science Fiction and the Military Reader. *The RUSI Journal*, 162, 6, pp. 60-64.
- Coskun, B. B. (2012). Colonialism and Mandates. In A. L. Stanton's (Ed.) *Cultural Sociology of the Middle East, Asia, and Africa: An Encyclopedia* (Vol. 1, pp. 255-257). Thousand Oaks, CA, Sage Publications, Inc. Retrieved on 30 August, 2018 from:

https://books.google.nl/books?id=GtCL2OYsH6wC&printsec=frontcover&source=gbs_ViewAPI&redir_esc=y#v=onepage&q&f=false.

- Crowe, M. J. (1997). A history of the extraterrestrial life debate. *Zygon*, 32, 2: 147-162.
- Cudd, A. Eftekhari, S. (2017). Contractarianism. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*. Retrieved on May 18, 2018 from: <https://plato.stanford.edu/entries/contractarianism/>.
- D'Agostino, F., Guas, G., Thrasher, J. (2017). Contemporary Approaches to the Social Contract. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*. Retrieved on May 18, 2018 from: <https://plato.stanford.edu/entries/consequentialism/>.
- Darwin, C. (1861). *The Origin of Species*. New York, NY: D. Appleton and Company.
- De Bruin, B. (2005). Game Theory in Philosophy. *Topoi*, 24: 197-208.
- Deardorff, J., W. (1986). Possible Extraterrestrial Strategy for Earth. *Quarterly Journal of the Royal Astronomical Society*, 27, pp. 94-101.
- Degrazia, D. (1999). Animal Ethics Around the Turn of the Twenty-First Century. *Journal of Agricultural and Environmental Ethics*, 11, pp. 111-129.
- Deleuze, G. (1994). *Difference and Repetition* (P. Patton, Trans.). New York City, New York: Columbia University Press.
- Deleuze, G. (2003). *Francis Bacon: The Logic of Sensation* (D. W. Smith, Trans.). Minneapolis, MN: University of Minnesota Press.
- Dessy, R. (n.d.). Could silicon be the basis for alien life forms, just as carbon is on Earth? *Scientific American*. Retrieved on May 30, 2018 from: <https://www.scientificamerican.com/article/could-silicon-be-the-basi/>.
- Dillon, M. (2014). *Introduction to sociological theory: theorists, concepts, and their applicability to the twenty-first century*. Chichester, UK: John Wiley & Sons, Ltd.
- Dixit, A. (2006). Thomas Schelling's Contributions to Game Theory. *Scandinavian Journal of Economics*, 108, 2, pp. 213-229.
- Dixit, A., Skeath, S., (1999). *Games of Strategy*. New York, NY: W. W. Norton.
- Dryzek, J. S. (2000). *Deliberative Democracy and Beyond: Liberals, Critics, Contestations*. Oxford: Oxford University Press.
- Dubey, P., Geanakoplos, J., Shubik, M. (1987). The Revelation of Information in Strategic Market Games: A Critique of Rational Expectations Equilibrium. *Journal of Mathematical Economics*, 16, pp. 105-137.
- Dufwenberg, M. (2011). Game Theory. *Wiley Interdisciplinary Reviews: Cognitive Science*, 2, 2: 167-173.
- Dustmann, C., Hatton, T., Preston I. (2005). The Labour Market Effects of Immigration. *The Economic Journal*, 115, pp. 297-299.

- Erickson, P., Klein, J., L., Daston, L., Lemov, R., Sturm, T., Gordin, M. D. (2013). *How Reason Almost Lost Its Mind: The Strange Career of Cold War Rationality*. Chicago, IL, The University of Chicago Press.
- Evans, C. (1988). *Writing Science Fiction*. London: A. & C. Black.
- Fishkin, J. S. (2011). Deliberative Democracy and Constitutions. *Social Philosophy & Policy*, 28, 1, pp. 242-261.
- Floyd, S. (n.d.). Thomas Aquinas: Moral Philosophy. In J. Fieser & B. Dowden (Eds.), *Internet Encyclopedia of Philosophy*. Retrieved on June 5, 2018 from: <https://www.iep.utm.edu/aq-moral/#>.
- Fomin, E. S. D., Ndobegang, M. M. (2006). African Slavery Artifacts and European Colonialism: The Cameroon Grassfields from 1600 to 1950. *European Legacy*, 11, 6, pp. 633-646.
- Freeman, M. J. (1968). Science Fiction: Its Use in Liberal Studies. *The Vocational Aspect of Secondary and Further Education*, 20, 46, pp. 159-161.
- Freeman, S. (2007). *Justice and the Social Contract*. Oxford: Oxford University Press.
- Friend, C. (n.d.). Social Contract Theory. In J. Fieser & B. Dowden (Eds.), *Internet Encyclopedia of Philosophy*. Retrieved on May 21, 2018 from: <https://www.iep.utm.edu/soc-cont/>.
- Bush, E., Macdonald, A., Reich, A., Rudin, S. (Producers) & Garland, A. (Director). (2018). *Annihilation* [Motion Picture]. United States: Paramount Pictures.
- Gauthier, D. (1986). *Morals by Agreement*. Oxford: Clarendon Press.
- Gauthier, D. (1988). Hobbes's Social Contract. *Nous*, 22, 1, pp. 71-82.
- Genghis Khan. (2009). Retrieved on 30 August, 2018 from: <https://www.history.com/topics/genghis-khan>.
- von Goethe, J. W. (2012). *The Sorrows of Young Werther* (D. Constantine, Trans.). Oxford: Oxford University Press.
- Gomel, E. (2011). Science (Fiction) and Posthuman Ethics: Redefining the Human. *The European Legacy*, 16, 3, pp. 339-354.
- Green, S. (2002). *Rational Choice Theory: An Overview*. Waco, TX: Baylor University. Retrieved on January 7, 2018 from: https://business.baylor.edu/steve_green/green1.doc.
- Greshko, M. (2018). Stephen Hawking's Most Provocative Moments, From Evil Aliens to Black Hole Wagers. *National Geographic*. Retrieved on August 11, 2019 from: <https://www.nationalgeographic.com/news/2018/03/stephen-hawking-controversial-physics-black-holes-bets-science/>.
- Grüne-Yanoff, T. (n.d.). Game Theory. In J. Fieser & B. Dowden (Eds.), *Internet Encyclopedia of Philosophy*. Retrieved on May 21, 2018 from: <https://www.iep.utm.edu/soc-cont/>.
- Habermas, J. (1987). *Toward a Rational Society*. (J. J. Shapiro, Trans.). Cambridge and Oxford: Polity Press.

- Habermas, J. (1996). *Between Facts and Norms: Contributions to a Discourse Theory of Law and Democracy*. (W. Rehg, Trans.). Cambridge, MA: MIT Press.
- Hagemann, H., Kufenko, V., Raskov, D. (2016). Game Theory Modeling for the Cold War on both Sides of the Iron Curtain. *History of the Human Sciences*, 29, 4-5, pp. 99-124.
- Hamlin, A. (1996). [Review of the book *Condorcet: Foundations of Social Choice and Political Theory*. By M. J. A. N. C. Condorcet, I. Mclean (Ed., Trans.), F. Hewitt (Ed., Trans.)]. *Journal of Economic Literature*, 34, 3, pp. 1334-1335.
- Hampton, J. (1986). *Hobbes and the Social Contract Tradition*. Cambridge: Cambridge University Press.
- Hart, M. H. (1975). An Explanation of the Absence of Extraterrestrials on Earth. *Quarterly Journal of the Royal Astronomical Society*, 16, pp. 128-135.
- Hegel, G. W. F. (1991). *Elements of the Philosophy of Right* (H. B. Nisbet, Trans.). Cambridge: Cambridge University Press.
- Herings, P. J.-J., Houba, H. (2016). The Condorcet Paradox Revisited. *Social Choice and Welfare*, 47, pp. 141-186.
- Hibbs, T. (1988). Against a Cartesian Reading of “Intellectus” in Aquinas. *The Modern Schoolman*, 66, pp. 55-69.
- Hicks, S. R. C. (n.d.). Ayn Rand. In J. Fieser & B. Dowden (Eds.), *Internet Encyclopedia of Philosophy*. Retrieved on June 1, 2018 from: <https://www.iep.utm.edu/rand/>.
- Hill, W., Carroll, G., Giler, D. (Producers), & Scott, R. (Director). (1979). *Alien*. [Motion picture]. USA: 20th Century Fox.
- Hobbes, T., Rogers, G. A. J. (2005). *Leviathan*. (Schuhmann, K., Ed.) London: Continuum.
- Huijsmans, I., Ma, I., Micheli, L., Civai, C., Stallen, M., Sanfrey, A. G. (2019). A scarcity mindset alters neural processing underlying consumer decision making. *Proceedings of the National Academy of Sciences*, 116, 24, pp. 11699-11704.
- Hume, D. (1989). *A Treatise of Human Nature* (Ed. L.A. Selby Bigge). Oxford: Clarendon Press.
- Hussain, N. J. Z. (2011). The Role of Life in the Genealogy. In S. May's (Ed.) *Nietzsche's On the Genealogy of Morality*. Cambridge, MA: Cambridge University Press.
- IAA SETI Permanent Committee. (2016). *The Rio Scale*. Retrieved on July 13, 2018 from: <http://www.setileague.org/iaaseti/rioscale.htm>.
- Jaspers, K. (1965). *Nietzsche: An Introduction to the Understanding of his Philosophical Activity*. (C. Wallraff's and F. Schmitz's, Trans.). Baltimore, MD: Johns Hopkins University Press.
- Jenkins, J. I. (1997). *Knowledge and Faith in Thomas Aquinas*. Cambridge, UK: Cambridge University Press.
- Katsafanas, P. (2013). Nietzsche's Philosophical Psychology. In K. Gemes' and J. Richardson's (Eds.) *The Oxford Handbook of Nietzsche*. Oxford: Oxford University Press.

- Kakutani, M. (2017). *Transcript: President Obama on What Books Mean to Him*. Retrieved on June 12, 2019 from: <https://www.nytimes.com/2017/01/16/books/transcript-president-obama-on-what-books-mean-to-him.html>.
- Kalin, J. (1977). Philosophy Needs Literature: John Barth and Moral Nihilism. *Philosophy and Literature*, 1, 2, pp. 170-182.
- Kant, I. (1785). *Grounding for the Metaphysics of Morals*. (J.W. Ellington, Trans.). Indianapolis, Ind.: Hackett Publishing Company.
- Kenny, A. (2010). *A New History of Western Philosophy: In Four Parts*. New York, NY: Oxford University Press Inc.
- Kennedy, K. (Producer), & Spielberg, S. (Director). (1982). *E.T.: The Extra-Terrestrial* [Motion Picture]. USA: Universal Pictures.
- Kodjo-Grandvaux, S. (2013). *Philosophies Africaines*. Paris: Présence Africaine.
- Lacan, J. (1977). *Écrits: A Selection*. (A. Sheridan, Trans.). London: Travistock/Routledge.
- Large, W. (2015). *Levinas' 'Totality and Infinity': A Reader's Guide*. London: Bloomsbury Publishing Plc.
- Leib, L. H. (2011). *Human Rights and the Environment: Philosophical, Theoretical and Legal Perspectives*. Leiden: Martinus Nijhoff.
- Leiter, B. (2015). Nietzsche's Moral and Political Philosophy. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*. Retrieved on July 19, 2018 from: <https://plato.stanford.edu/entries/nietzsche-moral-political/#2>.
- Lennox, J. (2015). Darwinism. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*. Retrieved on May 21, 2018 from: <https://plato.stanford.edu/entries/darwinism/>.
- List, C. (2013). Social Choice Theory. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*. Retrieved on May 21, 2018 from: <https://plato.stanford.edu/entries/social-choice/>.
- Liu, C. (2008a). *The Three Body Problem*. New York City, NY: Tor Books.
- Liu, C. (2008b). *The Dark Forests*. New York City, NY: Tor Books.
- Liu, C. (2016). *Death's End*. New York City, NY: Tor Books.
- Locke, J. (1970). *Two Treatise of Government*. Cambridge: Cambridge University Press.
- Loizou, J. (2006). Turning space tourism into commercial reality. *Space Policy*, 22: 289-290.
- Luetge, C. (2015). *Order Ethics or Moral Surplus: What Holds a Society Together?* London, Lexington Books.
- Luetge, C. (2016). Order Ethics and the Problem of Social Glue. *University of St. Thomas Law Journal*, 12, 2, pp. 339-359.
- Marino, M. (2012). *Daimones*. Retrieved on July 30, 2018 from: <https://books.noisetrade.com/massimomarinno/daimones>.

- Maccone, C. (2010). The Statistical Drake Equation. *Acta Astronautica*, 67: 1366-1383.
- May, J. (2011). Egoism, Empathy, and Self-Other Merging. *The Southern Journal of Philosophy*, 49, pp. 25-39.
- Michaud, M. A., G. (2007). *Contact with alien civilizations: Our Hopes and Fears about Encountering Extraterrestrials*. New York City, NY: Copernicus Books.
- Michaels, W. B. (2000). Political Science Fictions. *New Literary Studies*, 31, 4, pp. 649-664.
- Milbank, J. (1999). The Ethics of Self-Sacrifice. *First Things*, 91, pp. 33-38.
- Mill, J. S. (1998). *Utilitarianism* (R. Crisp, Ed.). Oxford: Oxford University Press.
- Morris, C. S. (2003). *Life's Solution: Inevitable Humans in a Lonely Universe*. Cambridge: Cambridge University Press.
- Mueller, D. R, Wang, G. X., Liu, G., Ciu, C. C. (2016). Consumer Xenocentrism in China: An Exploratory Study. *Asia Pacific Journal of Marketing and Logistics*, 28, 1, pp. 73-91.
- Musso, P. (2012). The Problem of the Active SETI: An Overview. *Acta Astronautica*, 78: 43-54.
- Nehamas, A. (1985). *Nietzsche: Life as Literature*. Cambridge, MA: Harvard University Press.
- Neil deGrasse Tyson Thinks Humans Might Be Too Stupid for Aliens. (2016). *Curiosity*. Retrieved on March 25, 2018 from: <https://curiosity.com/topics/neil-degrasse-tyson-thinks-humans-might-be-too-stupid-for-aliens-curiosity/>.
- Neuvel, S. (2016). *Sleeping Giants*. New York, NY: Penguin Random House LLC.
- Nielsen, K. (1972). Ethical Egoism and Rational Action. *The Journal of Philosophy*, 69, 20, pp. 698-700.
- Nkrumah, K. (1965). *Neo-Colonialism: The Highest Stage of Imperialism*. London: Heinemann.
- Noixius. (2017, November 19). *The Dark Forest theory of the Universe* [Online Forum]. Retrieved on May 30, 2018 from: https://www.reddit.com/r/threebodyproblem/comments/7e38fx/the_dark_forest_theory_of_the_universe/.
- Nussbaum, M. (1990). *Love's Knowledge: Essays on Philosophy and Literature*. Oxford, UK: Oxford University Press.
- Obadina, T. (2000). The Myth of Neo-colonialism. *Africa Economic Analysis*. Retrieved on July 17, 2018 from; <http://www.afbis.com/analysis/neo-colonialism.html>.
- Osborne, M., Rubinstein, A. (1994). *A course in Game Theory*. Cambridge, MA: MIT Press.
- Ossola, A. (2017, November 30). A Key Evolutionary Step May Mean Intelligent Alien Life Doesn't Exist in the Universe. *Futurism*. Retrieved on May 30, 2018 from: <https://futurism.com/evolutionary-biology-intelligent-alien-life/>.
- Piper, O. A. (1957). The Meaning of Work. *Theology Today*, 14, 2, pp. 174-194.
- Preti, A. (2007). Suicide Among Animals: A Review of Evidence. *Psychological Reports*, 101, pp. 831-848.

- Pueojit. (2017, October 10). *Is the dark forest theory fundamental to the universe or just a consequence to galactic war? [Series Spoilers]* [Online Forum]. Retrieved on May 30, 2018 from: https://www.reddit.com/r/threebodyproblem/comments/75esab/is_the_dark_forest_theory_fundamental_to_the/.
- Purcell, M. (2009). Resisting Neoliberalization: Communicative Planning or Counter-Hegemonic Movements. *Planning Theory*, 8, 2, pp. 140-165.
- Quine, W. V. O. (1960). *Word and Object*. Cambridge, MA: MIT Press.
- Rawls J. (1971). *A Theory of Justice*. Cambridge, MA: The Belknap Press of Harvard University Press.
- Reddy, V. M., Nica, M., Wilkes, K. (2012). Space tourism: Research recommendations for the future of the industry and perspectives of potential participants. *Tourism Management*, 33: 1093-1102.
- Regan, T. (1983). *The Case for Animal Rights*. London: Routledge & Kegan Paul.
- Ryan, A. (2012). *On Politics: A History of Political Thought from Herodotes to the Present*. London: Penguin Books.
- Rauscher, F. (2016). Kant's Social and Political Philosophy. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*. Retrieved on May 26, 2018 from: <https://plato.stanford.edu/entries/kant-social-political/>.
- Reed, A. E., Carstensen, L. L. (2012). The Theory Behind the Age-Related Positivity Effect. *Frontiers in Psychology*, 3, pp. 1-9.
- Reginster, B. (2006). *The Affirmation of Life: Nietzsche on Overcoming Nihilism*. Cambridge, MA: Harvard University Press.
- Sartre, J. P. (1964). *Colonialism and Neocolonialism* (S. Brewer, A. Haddour, T. McWilliams Trans.). Paris: Routledge.
- Schelling, T. C. (1960). *The Strategy of Conflict*. Cambridge, MA: Harvard University Press.
- Schipper, B. C. (2001). *Pure vs. Mixed Motive Games: On the Perception of Payoff-Orders*. Unpublished manuscript, Department of Economics, University of Bonn, Germany.
- Schacht, R. (1983). *Nietzsche*. London: Routledge.
- Schneider, S. L. (2008). Anti-Immigrants Attitudes in Europe: Outgroup Size and Perceived Ethnic Threat. *European Sociological Review*, 24, 1, pp. 53-67.
- Sensat, J. (1998). Game Theory and Rational Decision. *Erkenntnis*, 47, pp. 379-410.
- SETI Permanent Study Group of the International Academy of Astronautics. (2010). *Declaration of Principles Concerning the Conduct of the Search for Extraterrestrial Intelligence*. Prague, Czech Republic.
- Shaver, R. (2017). Egoism. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*. Retrieved on March 28, 2018 from: <https://plato.stanford.edu/entries/consequentialism/>.

- Shawn, L., Linde, D., Levine, D., Ryder, A. (Producers) & Villeneuve, D. (Director). (2016). *Arrival* [Motion Picture]. USA: Paramount Pictures.
- Singer, P. (1995). *Animal Liberation*. London: Pimlico.
- Snugden, R. (2002). Ken Binmore's Evolutionary Social Theory. *The Economic Journal*, 111, pp. 213-243.
- Soll, I. (2015). Nietzsche Disempowered: Reading the Will to Power out of Nietzsche's Philosophy. *Journal of Nietzsche Studies*, 46, 3, pp. 425-450.
- Stahel, W. R. (2016). The Circular Economy. *Nature*, 531, 7595, pp. 435-438. Retrieved on June 11, 2018 from: <http://www.nature.com/ruidm.oclc.org/news/the-circular-economy-1.19594>.
- Steven, D., J. (2013). The Societal Impact of Extraterrestrial Life: The Relevance of History and the Social Sciences. In Vakoch, D., A. (Ed.) *Astrobiology, History, and Society: Life Beyond Earth and the Impact of Discovery*.
- Stoddard, E. (2017, May 9). Early humans co-existed in Africa with human-like species 300,000 years ago. *Reuters*. Retrieved on May 31, 2018 from: <https://www.reuters.com/article/us-science-fossils-hominins/early-humans-co-existed-in-africa-with-human-like-species-300000-years-ago-idUSKBN1850QH>.
- Stryker, S. D. (2000). Communicative Action in History. *European Journal of Social Theory*, 3, 2, pp. 215-234.
- Sustein, C. R., Nussbaum, M. C. (2004). *Animal Rights: Current Debates and New Directions*. Oxford: Oxford University Press.
- Thornton, W. H. (2000). Analyzing East/West Power Politics in Comparative Cultural Studies. *CLCWEG: Comparative Literature and Culture*, 2, 3, pp. 1-12.
- Ulbrich, E., Card, O., S., Orci, R., Kurtzman, A., Pritzker, G., Chartoff, R., Hendee, L., McDough, L. (Producers) & Hood, G. (Director). (2013). *Ender's Game*. [Motion Picture]. USA: Summit Entertainment.
- UN. (1966). 2222 (XXI) *Treaty on Principles governing the activities of States in the exploration and use of Outer Space, including the Moon and other celestial bodies*.
- Vegetius. (2002). *Het Romeinse Leger: Handboek voor de Generaal*. [Epitoma rei militaris]. (F. Meijer, Trans.). Amsterdam: Athenaeum – Polak & Van Gennepe.
- Voluntary Human Extinction Movement. (n.d.). Retrieved on June 5, 2018 from: <http://vhemt.org/>.
- Walter, B., M. (2000). Political Science Fictions. *New Literary History*, 31, 4, pp. 649-664.
- Watson, J. (2013). *Strategy: An Introduction to Game Theory*. New York: Norton.
- Webb, S. (2015). *If the Universe Is Teeming with Aliens . . . WHERE IS EVERYBODY? Seventy-Five Solutions to the Fermi Paradox and the Problem of Extraterrestrial Life*. Heidelberg, NY: Springer.

- Weber, M. (2002). The Protestant Ethic and the Spirit of Capitalism. In C. Calhoun's, J. Gerteis', J. Moody's, S. Pfaff's, K. Schmidt's & I. Virk's (Eds.) *Classical Sociological Theory*. Malden, MA: Blackwell Publishers.
- Weber, M. (2004). Science as Vocation. In D. Owens, T. B. Strong's (Eds.) & R. Livingstone (Trans.) *The Vocation Lectures*. Indianapolis, Ind.: Hackett Publishing Company, Inc.
- Weintraub, E. R. (2017). Game Theory and Cold War Rationality: A Review Essay. *Journal of Economic Literature*, 55, 1, pp. 148-161.
- von Weizsäcker, E. U. (1994). Sustainable Economy. *The Science of Total Environment*, 143, pp. 149-156.
- Williams, H. (1992). Kant's Concept of Property. In R. F. Chadwick's (Ed.) *Immanuel Kant: Critical Assessments Volume III*. London: Routledge.
- Wolfe, P. (2006). Settler colonialism and the elimination of the native. *Journal of Genocide Research*, 8, 4, pp. 387-409.
- Xingshi, L. (1997). SF, Do Its Bit For Protection Of Our Old Earth's Environment. '97 *Beijing International Conference on Science Fiction: Essays*, pp 92-93.
- Young, A. (2015). *The Twenty-First Century Commercial Space Imperative*. Heidelberg, NY: Springer.
- Zagare, F. C. (2011). *Game Theory: Concepts and Applications*. Newbury Park, CA: Sage Publications, Inc.