***JAST*, 2022; 57: 27-51**
Submitted: 14.01.2022
Accepted: 05.03.2022
ORCID# 0000-0001-6846-7529

Science Fiction Literature as Thought Experiment: An Ethical Analysis of Michael Crichton's *Prey*

Firuze Güzel

Abstract

Science fiction literature, a product of unlimited imagination, often contains several philosophical issues related to ethics within its narratives. Science fiction works, which sometimes function as a thought experiment, provide examples of how humanity may react in various situations with a controlled scenario. Michael Crichton's novel *Prey* (2002) is one of such works. Crichton's novel, which warns the reader about the probable dangers of irresponsible use of technology with a striking scenario, essentially functions as a thought experiment. This article aims to reveal the relationship and similarity between philosophical thought experiments and science fiction literature. In this context, *Prey* will be analyzed in terms of ethical theories, and how ethics becomes the subject of science fiction literature will be elaborated.

Keywords: Thought Experiment, Science Fiction Literature, Ethics, Ethical Analysis

Bir Düşünce Deneyi Olarak Bilim Kurgu Edebiyatı: Michael Crichton'ın *Prey* Romanının Etik Analizi

Öz

Sınırsız bir hayal gücü ürünü olan bilim kurgu edebiyatı, anlatıları ile çoğu zaman içinde etik ile ilintili birtakım felsefi sorunlar

barındırır. Kimi zaman bir düşünce deneyi olarak da işlev gören bilim kurgu eserleri, kontrollü bir senaryo ile insanlığın çeşitli durumlarda nasıl tepkiler verebileceğine dair örnekler sunar. Michael Crichton'ın *Prey* (2002) adlı romanı da böylesi eserlerden biridir. Teknolojinin sorumsuzca kullanılmasının doğurabileceği tehlikeler hakkında okuyucuyu çarpıcı bir senaryo ile uyaran Crichton'ın romanı, esasen bir düşünce deneyi olarak işlev görmektedir. Eldeki çalışma, felsefi düşünce deneyleri ile bilimkurgu edebiyatı arasındaki ilişki ve benzerliğin ortaya çıkarılmasını amaçlamaktadır. Bu bağlamda, *Prey* romanı etik kuramları temelinde analiz edilerek etiğin nasıl bilimkurgu edebiyatının konusu hâline geldiği incelenecektir.

Anahtar Kelimeler: Düşünce Deneyi, Bilim Kurgu Edebiyatı, Etik, Etik Analizi

> "*They didn't understand what they were doing."*
> *I'm afraid that will be on the tombstone of*
> *the human race.*
>
> -Michael Crichton, *Prey*

Introduction

As the readers and researchers of SF are already familiar, this genre has always been "a literature of ideas" and in fact, "it is the *only* literature of relevant ideas" as Isaac Asimov suggests (307). In this sense, SF literature embodies a philosophical dimension that asks epistemological and ontological questions by depicting an imaginative world. As the title of this article suggests, there seems to be a unique dimension to the SF literature, which is more of an analogy that claims a striking similarity between science fiction and ethical thought experiments. According to this, SF serves as a valuable source for us to imagine the ethical responses we may make, right and wrong actions we may perform in the event of an encounter with the unknown, whether this unknown may be aliens, AIs, robots, cyborgs and organic/inorganic/hybrid existences, or any other technology. It helps us to imagine ourselves in such scenarios, serves as extended thought experiments, and provides us ethical guidance. After all, "SF

may not be the only moral literature left but the best of it is that," says Timothy Dolan in "Science Fiction as Moral Allegory" (2020) (112). In this context, this article firstly explores how such an analogy can be constructed and aims to examine Michael Crichton's novel *Prey* (2002) as an example. In such an examination, it is seen that the powerful narration Crichton presents implores ethics of technology, governments, institutions and individuals by constructing a gray goo[1] scenario of nanotechnology, thus also functioning as a thought experiment for an ethical analysis. Such an examination (a) would point out how significant the relationship between literature and philosophy is, in this particular case it is science fiction literature and ethics, and (b) demonstrate that literary ethical analysis can be a valuable method to conduct a more systematic examination by making direct references to the theories of moral philosophy.

An Analogy of Science Fiction Literature and Thought Experiments

To begin with, a thought experiment can be defined as a hypothetical situation in which the possible consequences of a principle or a theory to be tested through thinking and imagination. First established by Hans Christian Ørsted in 1812 as *Gedankenexperiment*, the English version of the term we use today appeared in Ernst Mach's translated paper "On Thought Experiments" (1897). The purposes of this kind of experiment vary; it may be used for entertainment, education, exploration, theory selection and implementation and so on. However, thought experiments hold a very significant place in the field of philosophy, particularly in ethics. By taking Philippa Foot's famous "Trolley Problem" as her example, Frances Myrna Kamm explains the use of these in ethics as such:

> "Trolley problems" are not supposed to describe actual ethical problems or to be solved with a "right" choice. Rather, they are thought-experiments where choice is artificially constrained to a small finite number of distinct one-off options and where the agent has perfect knowledge. These problems are used as a theoretical tool to investigate ethical intuitions and theories—especially the difference between actively doing vs. allowing something to happen, intended vs. tolerated consequences, and consequentialist vs. other normative approaches. (qtd. in Müller)

As understood from the quotation, such experiments help us to inquire into the possibilities of actions and their outcomes, not for the aim of finding a right action but for examination of ethical theories within variant scenarios. Therefore, the power of a thought experiment lies in its imaginative nature, in the "what if" questions it presents. When the aforementioned analogy is considered, it is observed that the relationship between thought experiments and literature becomes very functional for providing an understanding of moral discussions.

Thought experiments are conducted in controlled environments, they are designed for testing specific phenomena or principles. They have a narrative structure and they often need to be interpreted. It is in these aspects that they share a similarity with fiction; while reading works of fiction, like thought experiments, the reader is required to participate cognitively in the imaginative hypothetical scenarios. In her "Fiction as Thought Experiment" (2014), Catherine Z. Elgin suggests that thought experiments are imaginative exercises, which put forward the question of what would happen if certain conditions were to occur. They are conducted with a suspension of disbelief that the conditions may not exist in reality, or may be inconsistent with reality or even cannot be obtained (231). This similarity is an important starting point for considering fictions as thought experiments. As she states:

> If an austere thought experiment can afford epistemic access to a range of properties, and can do so in a context that is not tightly beholden to a particular theory, there seems to be no reason to deny that a more extensive thought experiment can do the same. This opens the way to construing works of literary action as extended, elaborate thought experiments. They afford epistemic access to aspects of the world that are normally inaccessible—in particular, to the normative, psychological and metaphysical aspects that philosophical thought experiments concern. (232)

When the role of fiction in understanding these aspects of human thought is deliberated, as explained by Elgin above, it would not be completely wrong to say that whatever the thought experiments are to philosophy and ethics, science fiction holds a similar role for literature and ethics. Why science fiction holds a separate role is also of significance because compared to other genres, SF is "the" genre that allows the imagination to create hypothetical what if scenarios in the infinite numbers of

ways. Hence, philosophy, ethics in particular, and science fiction share many common themes and discussions that inquire into hypothetical imaginative scenarios and how humanity would or should act in the face of the unknown. "Intriguingly, if you read science fiction writers like Stanislaw Lem, Isaac Asimov, Arthur C. Clark [sic], and Robert Sawyer, you are already aware that some of the best science fiction tales are in fact long versions of philosophical thought experiments" (2) says Susan Schneider in "Thought Experiments: Science Fiction as a Window into Philosophical Puzzles" (2016). She presents the shared themes between philosophy and science fiction literature, and emphasizes the convergence point of science fiction and science fact: "some of the most lavish science fiction thought experiments are no longer merely fictions – we see glimpses of them on the technological horizon" (5).

As the technology advances, the themes of science fiction literature also change, keeping up with the pace of our world. Timothy Dolan explains that "a key point regarding the moral dimensions of science fiction is that it often rises to the level of classic literature because it exposes the contradictions of prevailing norms in its investigation of the moral ambiguities that arise from the socio-technological nexus" (111). Within this nexus, there arise many themes, which can be examined to understand the ethical dimension of SF literature and as stated by Harald A. Wiltsche, "If written in a scientifically responsible manner, science fiction has thus the potential to increase our understanding of what it means for us, as human beings, to live in the kind of world our increasingly arcane scientific theories purport to describe" (18). A similar argument is also made by Russell Blackford in *Science Fiction and The Moral Imagination: Visions, Minds, Ethics* (2017). Standing as one of the most important sources in the study of ethical/moral dimension of SF, Blackford's propositions gain more importance. He states:

> When its tropes are used more seriously, however, it often explores the social and psychological effects—and hence the moral significance—of scientific and technological innovations. With its greatly extended narrative possibilities, science fiction can illuminate the social impact of change, propose blueprints for a better future, or implicitly criticize any naive optimism about where the human species is headed. . . . Recurring themes in science fiction include the design

> and functioning of future societies, terraforming and cosmic engineering, reshaping ourselves with technology, and questions about our treatment of non-human persons (perhaps extraterrestrial aliens or advanced artificial intellects). Science fiction writers have employed the genre's tropes to engage with a wide variety of moral questions. ("Introduction" 14)

In this respect, SF has proven itself a vast area for ethical discussions, not just for the technological but also for the individual, social, institutional, political and many other levels.

Ethics of Science Fiction: *Prey*

This overall discussion brings us to the practical use of the aforementioned analogy: the ethical analysis of *Prey*. Beginning with the emergence of the genre, American science fiction literature has involved moral themes, and presented the ethical dilemmas of the period it was written in. "Though fictional narratives engage with ethical questions, they seldom comment directly on ethical systems such as Kantian deontology and utilitarianism. There are, however, some stories that go close, inviting us to judge the character of individuals or societies that seem to embody philosophical stances" (Blackford, "Engaging" 75). This statement is also valid for SF literature, as it does not directly address ethical theories, yet, the implications of these can be found within works of science fiction. We can include *Prey* among such works, which examines the ethical implications of nanotechnology as well as artificial intelligence, agent-based computing, emergence and complexity and host-parasite coevolution. As stated by Michael Crichton in the novel's introduction, titled "Artificial Evolution in the Twenty-first Century,": "it is always possible that we will not establish controls. Or that someone will manage to create artificial, self-reproducing organisms far sooner than anyone expected. If so, it is difficult to anticipate what the consequences might be. That is the subject of the present novel" (xv).

In attempt to attract attention to this subject, as it can be seen above, Crichton begins the ethical discussion right at the beginning of the novel. In the introduction, Crichton further proposes that every living organism changes at every instant in response to other living organisms, which means that all human actions have uncertain,

unpredictable effects. He criticizes how humanity interacts with nature in an "obstinate egotism" and calls for caution, which according to him we failed to possess in the past (x-xi). He states, "sometime in the twenty-first century, our self-deluded recklessness will collide with our growing technological power. One area where this will occur is in the meeting point of nanotechnology, biotechnology, and computer technology" (xi). As a writer who uses recent technological developments in his books, Crichton again warns us about the potential dangers of nanotechnology, and proposes that we should make international regulations on the use of such nanorobots. Written in 2002, in the article titled "Could Tiny Machines Rule the World?" Crichton proposes as follows:

> We know these machines are coming. We know we will have to control them when they do. It is not too early to plan how we will treat them, what we will allow in the way of research and what we will forbid. Historically, human beings have a poor record of addressing the hazards of new technologies as they arrive. We generally pass laws after the accidents occur. But in the case of self-reproducing machines, we simply can't wait.

What the author proposes here is a very relevant issue for the field of ethics, as he essentially makes two suggestions: firstly, we should discuss the ethical use of such technology and secondly we should create ethical machines. Even though we may not control such machines all the time, we should take preemptive measures. "All the potential uses and risks, and all the unknowns surrounding nanotechnology, seem to call for reflection and potentially for regulation, even at this early stage - both to avert disaster and to avoid uninformed panic" ("The Dust" 142). In this context, the book, in the broadest sense, functions as a thought experiment as well as a warning for humanity to act responsibly and carefully while dealing with nanotechnology, and to create ethical machines/codes so that the machines would inherently have ethical norms of not harming others. This gray goo scenario makes the claim that the nanotechnological particles can get intelligent, evolve very fast and destroy humankind. Therefore, the responsibility lies in the work of the scientists along with the corporations that fund and allow such inventions to come out in the first place.

In this context, Crichton clearly disapproves the pragmatist

outlook of the companies in the book as Jack, the protagonist, makes a criticism of this issue at the end of the novel:

> I didn't understand how they could have embarked on this plan without recognizing the consequences. Like everything else I'd seen at Xymos[2], it was jerry-built, half-baked, concocted in a hurry to solve present problems and never a thought to the future. That might be typical corporate thinking when you were under the gun, but with technologies like these it was dangerous as hell. (502-503)

Through Jack's words, Crichton warns corporations to adopt a more consequentialist, utilitarian ethical understanding in their use of technologies, which is also echoed by an idea repeated throughout the novel: "things never turn out the way you think they will" (2). The unpredictable nature of artificial intelligence and technology in general may produce undesired consequences and the ethical use/creation of such technologies should always come first. However, as the author presents this thought experiment, what happens when such an ethical outlook is not adopted can be observed. As Jack states: "I was feeling angry about what had happened in the desert. A chain of bad decisions, errors and fuckups extending over weeks and months. It seemed as if everyone at Xymos was doing short-term solutions, patch-and-fix, quick and dirty. No one was paying attention to the long-term consequences" (261-62).

The human factor in this disaster is also taken into consideration and why Julia, Ricky and Xymos company released the swarm to the environment as well as supporting its development, even though it was gaining autonomy, is also explained within the perspective of those involved by providing their personal justifications. While the company does not want to lose the funding and seeks profit, Ricky's motives are explained by Mae to Jack: "'I think Ricky sees Xymos as his last big chance to score. He's been here five years. If this doesn't work out, he'll be too senior to start over at a new company. He's got a wife and baby; he can't gamble another five years, waiting to see if the next company clicks. So he's really trying to make this happen, really driving himself'" (254). While Ricky seems to be influenced by the stress of his financial situation and career path, he performs actions that harm others, and this is presented as ethically wrong. Even though he does not directly mean harm, the actual consequences of

his actions are different, and therefore he is criticized by Jack. This criticism makes the reader consider if it is ethical or not to perform actions that in the end harm others but provide a personal gain. The same idea is reinforced when Julia explains her motives:

> You see, the thing is, I just wanted to save the company, Jack. That's all. The camera failed and we couldn't fix it, we lost our contract, and the company was falling apart. I've never lost a company before. I never had one shot out from underneath me, and I didn't want Xymos to be the first. I was invested, I had a stake, and I guess I had my pride. I wanted to save it. I know I didn't use good judgment. I was desperate. It's nobody else's fault. They all wanted to stop it. I pushed them to go on. It was... it was my crusade. (417)

While she seems to accept that she made bad decisions, she tries to justify her actions, yet again; the reader is invited to question if this justification is an acceptable one. Jack's inner thoughts are presented to the reader for ethical guidance for that matter: "I hated that she would start this when I was exhausted, when I had just gone through an ordeal that nearly got me killed and that was, ultimately, all her doing. I hated that she dismissed her involvement as 'bad judgment' when it was considerably worse than that" (417).

This ethical problem of personal gain in business and in personal life is not only limited to Xymos company, the story of how Jack is fired is also presented as another example. At MediaTronics company, Jack was running a program division that aims to create distributed parallel processing or agent-based programs modelled after biological processes inside a computer. After one of his codes is stolen from the company, Jack also becomes responsible for the security and increases the surveillance of the workers. This is how he initially learns that his boss, Don Gross, is having an affair with a woman working for the same company, and he has given her a company car. From Jack's reaction, it is assured that, he is a virtuous character who tries to do the right thing in all aspects of his life. The classical virtue theories in ethics consider virtue essential to well-being, and a golden mean should be found in actions. As suggested by Aristotle in *Nicomachean Ethics*, people have two types of virtues, theoretical virtues and virtues of character (1103a). In this division, while theoretical virtues point out to the virtues controlled by reason such as judgement and wisdom,

virtues of character point out to virtues controlled by desires such as courage, temperance, justice and generosity. It is seen that Jack possesses both types of virtues; he acts with his reason rather than his desires and demonstrates virtues of character on many occasions. As a first example of his virtuous character, Jack tries to do the right action and gives Gross a choice: "I went to him and said that based on emails relating to Jean in accounting, it appeared that someone unknown was having an affair with her, and that she might be getting perks she wasn't entitled to. I said I didn't know who the person was, but if they kept using email, I'd soon find out" (11). This instance also reflects the contemporary understanding of virtue which mainly focuses on motives and intentions as suggested by Elizabeth Anscombe in "Modern Moral Philosophy" (1958) (9). Jack follows his reason and intends well in his actions and this invites the reader to trust the moral rightness of the actions he will perform in the future.

Rosalind Hursthouse also suggests that the right action is what a virtuous agent would do because such an agent would have action guiding and action assessment senses helping him/her to find a fine tune, the Aristotelian golden mean (16). In this sense, Jack's action represents a right course of action, yet his motives come from a duty-based, Kantian perspective. He believes that he has duties against other people and particularly he should not harm them. Therefore, he intervenes and feels a moral responsibility to correct Gross' actions. While he does not directly condemn and reveal the moral wrongness of Gross' actions, he chooses to warn him indirectly. However, the abuse of power and sources of the company, as well as the affair and betrayal, are only the beginning of the unethical actions of Gross. Later, Jack discovers that Gross is actually selling company software to foreign distributers and taking fees in return; as it turns out this is indeed how Jack's code was "stolen" (12). The ethical dilemma Jack finds himself in once more becomes whether to inform authorities about this situation or not. As a virtuous character, Jack again feels a responsibility to correct this situation, his feelings in the matter are clear: "This was clearly illegal, and I couldn't overlook it" (12). The emphasis on the illegality of this action shows that today most ethical conducts are also regulated with laws, and laws regard a utilitarian understanding of ethics, which tries to maximize the benefit of all parties influenced by the actions. As a violation of this utilitarian principle, the actions of Gross are considered ethically wrong by Jack and therefore should be reported.

However, when Jack consults his lawyer, Gary, about the right course of action in such a situation, the advice he gives is interesting:

> "As your attorney, my advice is that if you are aware of any illegal activity you have a duty to report it. But as your friend, my advice is to keep your mouth shut and get out of there fast."
>
> "Seems kind of cowardly. I think I have to notify the investors." Gary sighed. He put his hand on my shoulder. "Jack," he said, "the investors can look out for themselves." (12)

The author here presents four ethical invitations to consider. The first one is the corruption of state and justice. In such a clear action against the law, a lawyer can give such an advice, knowing that the corrupted system would turn against Jack when he tries to do the right thing. The second issue here is that the investors are also corrupt, and seek personal gain, violating the trust between them and others. The author here invites the reader to think about how these unethical actions are against the utilitarian or even duty-based notions of ethics and the corporations; state and system of justice are corrupted; they do not function as they should and they violate the terms of social and legal contracts. The third invitation is the type of ethical duty Jack faces. In "The History of Utilitarianism," the author explains two basic types of duties that form the demandingness problem of utilitarianism: required and supererogatory ethical conducts. According to the notion of required ethical conduct, people have an ethical duty to perform some actions whereas supererogatory ethical conduct points out the actions that are beyond their duties and they cannot be blamed for not performing them (Driver). In this context, Jack's lawyer considers this as a supererogatory conduct as this duty which, according to him, is beyond Jack's duties; he would not be blamed if he chooses not to interfere and the right thing for Jack to do would be seeking out his own interest. This reveals the last invitation, which is Jack's response to Gary. His reaction to his lawyer's advice emphasizes his virtuous character and demonstrates that he considers this as a required ethical conduct on his behalf. He considers quitting from his job as a cowardly action and therefore informs one of the board members about this offense. However, Jack's lawyer proves right and the board member is in this conspiracy as well. Jack is immediately fired "for gross negligence and misconduct" (12) and he becomes a "marked

man" labeled as "Troublemaker. Not cooperative. Belligerent. Hot-headed. Not a team player . . . involved in some kind of shady dealings" (13). The ethical invitation offered by the author is reinforced by this damaging of Jack's reputation, and the critique of the society, which punishes the effort of right ethical action, presents itself.

Apart from these ethical issues on the individual, social, institutional and technological levels, there lies the most important matter of ethics that needs to be discussed yet again in the technological level. Crichton's gray goo brings out the ethical discussion of artificial intelligence as objects and subjects. The artificial intelligence of the nanoparticles in the novel is enough for them to create emergent, complex behaviours and consequently an artificial life. The main ethical issue is based on their use as objects. The initial aim of producing such a technology was actually a military one; it was designed as a defense project (170). However, presented as a medical imaging technology, the nanoparticles were mere objects of surveillance and real-time imaging. A fail in their design resulted in malfunctioning in high winds, and when the scientist could not find a solution to the problem, they released the nanoparticles to the environment, letting them solve the problem instead. With this aim, they rewrote their code by adding a genetic algorithm and provided them with solar power and memory. When the swarms were released, their biological part enabled them to evolve, reproduce, and learn how to self-optimize.

Henceforth, their use as objects causes several ethical problems. The first ethical problem presented within the context is the irresponsible action in using and designing new technologies, as it would not benefit humanity but the corporation. The author here invites readers to consider two layers of ethical discussion; the first one, as discussed above, is finding the right thing to do is while producing such technologies, and how the dangerous aspects of those can be eliminated, who is responsible when an accident occurs, and what the limit of intelligence that can be given to these creations is. The second ethical issue is the problem of autonomy. When these swarms of nanoparticles learn self-optimization, they gain a sense of autonomy, which necessitates the inquiry whether they can still be treated as objects or would their moral status change into ethical objects. Crichton addresses to this firstly by asking the essential question of whether programs with artificial intelligence (AI) can ever be self-aware; and the common idea is that they cannot. However, he continues to discuss further:

> But there's a more fundamental version of the question, a philosophical question about whether any machine can understand its own workings. Some people say that's impossible, too. The machine can't know itself for the same reason you can't bite your own teeth. And it certainly seems to be impossible: the human brain is the most complicated structure in the known universe, but brains still know very little about themselves. . . . But lately these philosophical questions have taken on new importance because there has been rapid progress in reproducing certain brain functions. Not the entire brain, just certain functions. (107-108)

What he states here, that the brain does not have much information about itself, is a very thought-provoking idea. Such a suggestion connotes that machines, even though they are not self-aware, can be intelligent beings if they have just enough functions, such as human brains. Crichton continues his discussion: "What was important about the programs was that the machines literally learned. They got better at their jobs with experience. Which is more than some human beings can claim" (108). If intelligence is a criterion of assessing human qualities, then how could a person evaluate artificial intelligence, which at times may be smarter than humans? The key answer to the question lies in the concept of autonomy and control. The differentiation between weak AI and strong AI here reveals itself, and this newly gained autonomy of the swarms and their ability of biological reproduction and evolving raise their chances of demonstrating features of strong AI. Crichton states:

> The eighties were a good time for English professors who believed that computers would never match human intelligence. But distributed networks of agents offered an entirely new approach. And the programming philosophy was new, too. . . . The program defined the behavior of individual agents at the lowest structural level. But the behavior of the system as a whole was not defined. Instead, the behavior of the system emerged, the result of hundreds of small interactions occurring at a lower level. (93)

As it can be seen, the author explains how the concept of artificial intelligence was perceived as an unlikely endeavor once. Today, however, the programs can show emergent behaviours as a result of

interaction, therefore making this once unlikely endeavor a possibility. This possibility necessitates a change in the ethical principles about AI as subject, as well. As the systems are not programmed, they could come up with surprising behaviours.

The nanoparticles in the novel demonstrate such surprising behaviours that demand a consideration upon their moral status. They exhibit intelligent behaviours, but surprisingly, the predator code they include causes them firstly to hunt small animals, and then human beings. On the verge of singularity[3], the swarms begin gaining a sense of consciousness. To explain how physical existence and consciousness of humans occur, Crichton makes a comparison between the swarms and the human beings, claiming that a human being is actually a swarm first in physical terms, then in terms of consciousness. Crichton explains this similarity: "If you could enlarge the human body, blow it up to a vast size, you would see that it was literally nothing but a swirling mass of cells and atoms, clustered together into smaller swirls of cells and atoms. . . . The control of our behavior is not located in our brains. It's all over our bodies" (362-63). Hence, this physical resemblance is also responsible for how similar the minds of a swarm presented in the novel and a human being work. Human beings also have a "swarm intelligence" in which there is not one single control unit but the brain takes signals from all organs. For that reason, human brains process many things that escape our immediate attention, consequently building up the subconscious. Crichton continues and gives an example of avoidance. According to this, the advantage of human beings is their unawareness of the obstacles they need to deal with until they lose a necessary organ or a sense. Therefore, according to Crichton, human beings' sense of consciousness or control is a mere illusion, and therefore such a self-consciousness and self-control can be gained by machines as well, as long as they have the necessary set of skills to create this illusion. What brings the swarms of *Prey* to the brink of singularity, as Jack states in the novel, is the mentioned self-consciousness and self-control: "and for all we knew, this damned swarm had some sort of rudimentary sense of itself as an entity. Or, if it didn't, it might very soon start to" (364).

Jack comes to this conclusion as he observes that the swarms learn by interaction, they have memory (a rather limited one), they are capable to hide, nest, adapt to new situations, reproduce, and hunt. At last, they begin imitating physical features of their preys, which

is an emergent behavior. This one, as presented as the "wild type" by Julia in the novel (467), tries to create a simulation of humanity. The most critical point occurs when the other one, the "benign type," evolves and merges with humans, thereby initiating a host-parasite coevolution and creating a human host for itself. To reach their aim of hunting all living things, due to the predator behavior in their code, they begin manipulating infected humans; they violate the cognitive functions of the host by dehumanizing them and thus create a new form of existence: a hybrid human-nanorobot. They also enhance the physical features of the host bodies: the hosts become more attractive and muscled. Nevertheless, while the physical aspects of the hosts are developing, their mental and psychological states deteriorate; they become more stressed than usual, less tolerant, angry and edgy. When these changes occur, their ethical status as objects also changes, and the orthogonality thesis[4] manifests itself, which poses an existential risk from a superintelligence that threatens humanity.

Within this respect, the author once more invites the reader to dwell upon the moral status of the AI, and the moment this status is determined, the actions of the protagonist and the ethical sanctions of these actions can be understood better. In the novel, when Jack first gains the knowledge of the runaway swarm, he directly considers it as an object and underestimates its intelligence. His initial impression of the swarm was that it could be easily killed. With this conversation however, he understands the features of this swarm more evidently:

> My head throbbed. I was seeing all the implications, now, and they weren't good.
>
> "So," I said, "what you're telling me is this swarm reproduces, is self-sustaining, learns from experience, has collective intelligence, and can innovate to solve problems."
>
> "Yes."
>
> "Which means for all practical purposes, it's alive."
>
> "Yes." David nodded. "At least, it behaves as if it is alive. Functionally it's alive, Jack." (245)

As understood from the quotation, this swarm, in technical terms, is alive as it has many capabilities and evolves in every hour. However,

Jack's idea about killing the swarm does not change after discovering its abilities, he says: "We've got to kill these things cold stone dead. We have to wipe them off the face of the planet. And we have to do it right now" (253). At the end of the novel, swarms are stopped by Jack and Mea as they pose a threat both to themselves and to humanity. The utilitarian principles necessitate such elimination of the threat, as these machines are no longer beneficial but harmful to all humankind. These machines cannot be accepted as fully ethical agents because they do not meet the necessary benchmarks. As stated by Martha Nussbaum and H. Peter Kahn, et al., there are several criteria for determining the worth of life and human qualities, and these swarms seem to meet some of these but not all. For instance, according to Nussbaum's criteria of worth of life (76-78), the swarms are alive, they have bodily health, they manage to create and keep bodily integrity, they have senses, imagination and thought (to some degree), however, they do not have emotions, practical reason, affiliations, their relation to other species is built upon dominating/hunting them, they do not know how to enjoy their life and they cannot control their environment. According to the criteria Kahn, et al. present to assess the qualities of a robot (366-381), it is seen that the swarms have autonomy, they can imitate others but these are the only things that they can do. Among other qualities that they do not have, the most significant ones are the notions of intrinsic moral value and moral accountability. In this context, as the novel suggests, they are not fully ethical agents, they do not possess a worth of life, and therefore killing them to protect humanity is not considered ethically wrong.

Another important ethical dilemma can be observed when it comes to harming others. As mentioned above, the swarms do not have moral accountability, and their killings of the team members in the desert are not questioned in the novel on the presumption that it is ethically wrong. However, killing other people is considered immoral in all major strands of ethical philosophy, and the fact that Jack kills his wife Julia and his friend Ricky at the end of the novel is another issue of ethics to be discussed. Jack explains his relationship to Ricky early in the novel, describing him as "cheerful and appealing" so that everyone would forgive his mistakes at work. He says: "At least, I always did [forgave], when he worked for me. I had become quite fond of him, and thought of him almost as a younger brother" (38). At this point, Crichton's thought experiment dwells upon the conditions

of having the urge to kill another person could be justified and given the circumstances provided in the novel, if Jack could be forgiven for such ethically questionable action. At this point, four important ethical theories should be mentioned. The first one is the aforementioned autonomy and worth of life concepts; Julia no longer possesses the necessary autonomy and other traits of being human as she is infected by the nanoparticles. Losing her set of cognitive skills, she functions as a host body controlled by this technology and therefore, as suggested in the book, can no longer have moral responsibility. After their confrontation, Jack realizes that the creature before him is no longer his wife: "In a way it made everything easier. Because I understood I wasn't dealing with Julia anymore. I didn't have to worry about what might happen to her. I just had to worry about Mae—assuming she was still alive—and me" (472). After this realization, Jack does not consider Julia as a fully ethical agent anymore; she clearly poses a threat to Jack, Mea, her children, and all humanity. Therefore, in a utilitarian manner, the killing of Julia, and Ricky for that matter, is justified within the ethical understanding of the novel. This situation is also an example for utilitarianism, as Jack, in an altruistic manner, makes a great sacrifice by killing his wife and friend; his motive to do the right action forces him to make such a selfless deed.

The second ethical theory, which applies to this dilemma, is the euthanasia theories. While some philosophers support voluntary euthanasia on the grounds of self-determination and helping others in distress, others support only involuntary euthanasia in rare cases[5]. In this context, Julia's case can be categorized as involuntary euthanasia because she does not hold a moral status anymore, her worth of life is questionable; she is in no position to make rational decisions, as she does not have self-awareness. The swarm infects Julia and Ricky to the degree that they form a symbiotic relationship and the humanity aspect of this relationship, both physically and mentally, is almost vanished. Jack witnesses how further the swarm takes over Julia in two instances. The first instance happens in the MRI room of the laboratory in the desert when he uses the MRI machine to clear the swarms away from her:

> And then in a sudden rush Julia literally disintegrated before my eyes. The skin of her swollen face and body blew away from her in streams of particles, like sand blown off a sand dune. . . . And when it was finished, what was left behind—

> what I still held in my arms—was a pale and cadaverous form. Julia's eyes were sunk deep in her cheeks. Her mouth was thin and cracked, her skin translucent. Her hair was colorless, brittle. Her collarbones protruded from her bony neck. She looked like she was dying of cancer. (469-70)

After this first encounter with the real Julia within the swarm, Jack's first impulse is to save her by making her drink a vial that would destroy the nanoparticles in her body. However, Julia claims that it is too late for such a rescue and very soon, when the machine loses its power, the nanoparticles assemble back to recreate the infected Julia. The second incident happens at the end of the novel, which indeed convinces Jack that his wife is beyond rescue.

> Julia came swirling up through the air toward me, spiraling like a corkscrew—and grabbed the ladder right alongside me. Except she wasn't Julia, she was the swarm, and for a moment the swarm was disorganized enough that I could see right through her in places; I could see the swirling particles that composed her. I looked down and saw the real Julia, deathly pale, standing and looking up at me, her face a skull. (477)

When Peter Singer's defense on the involuntary euthanasia stated in *Practical Ethics* (1979) is considered, the validity of Jack's decision is reinforced, as Singer suggests that killing someone without his or her consent could only be regarded as euthanasia on the grounds that the motive of such a killing is to prevent the unimaginable suffering of this person (158). Jack's decision of killing Julia and Ricky, therefore, is more than mere murder. As discussed before, these people are not fully ethical subjects anymore and the real people, who are hardly alive beneath the swarms, are in suffering. Hence, their death can be regarded within this perspective.

The third theory, which constitutes the concepts of rights mentioned by Philippa Foot in her "The Problem of Abortion and the Doctrine of Double Effect" (1967), applies to this situation as well. Negative rights, i.e. right to live, to be free, freedom of speech, religion, freedom from violence and slavery and rights of property, suggest the rights that a person cannot be prevented from or interfered with. Positive rights include the things that people are free to do, such as helping someone in distress. Non-interference is the right to not performing an

action, such as seeing two people fight and not to be involved (27). In this regard, negative rights of Julia no longer have validity because she loses her moral autonomy, and therefore these rights can be violated by Jack's positive rights to minimize the damage. Killing Julia (and Ricky for that matter) aims for this damage minimization by choosing the lesser evil. As exemplified by Foot's "Trolley Problem[6]," this action becomes acceptable and justified within the context of this novel.

The final theory that applies to this situation is Foot's doctrine of double effect. In the above-mentioned article, she explains this doctrine, and in basic terms, she suggests that people might perform morally good actions, even though this action may produce morally wrong side effects (20). Foot also distinguishes the moral difference between doing and allowing and she includes the doctrine of double effect to consider morally difficult cases: "The words 'double effect' refer to the two effects that an action may produce: the one aimed at, and the one foreseen but in no way desired. By 'the doctrine of the double effect' I mean the thesis that it is sometimes permissible to bring about by oblique intention what one may not directly intend" (20). In this sense, Jack's intention of killing is directed at the dangerous swarms, and killing Ricky and Julia is an undesired side effect. As Jack's intention is not directly murdering the human form of these people, the doctrine applies and Jack's action is once more justified.

So far, the moral rightness or wrongness of harming others and if such actions could be justified or not have been discussed, which leave one final comparison to be mentioned. Jack, who can take the life of infected people on the grounds discussed so far, cannot turn a blind eye to the death of an alive person earlier in the novel. After an attack by the swarms, and the swarms killing two of the team members, the others realize that Charley is still alive when they view the attack area from the monitors. Jack insists that someone should go outside and rescue him. However, the swarms could still be outside and this would be a great risk. The initial response of Ricky, who is infected by the swarms, is to leave Charley to die, as going outside is too dangerous. Because he lacks an empathy for the human life, he is calm and brutal; claiming that Charley would be dead by the time any of them reaches him (323). However, Jack does not accept such justification and goes outside once more even though he just survived the swarms' attack, and saves Charley. However, he later realizes how Ricky actually manipulated him to save Charley, with an agenda that Jack could be

killed or infected by swarms. In comparison to Jack's action of saving a live human being, the swarms on the other hand can murder people in cold blood, which again emphasizes the difference between humans and machines:

> Julia walked up to Charley, and kissed him full and long on the lips. Charley struggled, tried to wrench away. Vince grabbed a fistful of Charley's hair and tried to hold his head steady. Julia continued to kiss him. Then she stepped away, and as she did I saw a river of black between her mouth and Charley's. It was only there for a moment, and then it faded. . . . Julia wiped her lips, and smiled. (439-40)

As the swarms realize Charley as a threat to themselves, they kill him immediately. The two scenes can be examined in terms of moral status of actively allowing vs. actively doing a morally wrong action (Foot, "The Problem of Abortion" 20). In the first instance, Ricky intends actively allowing Charley to die and manipulates Jack for the same end as well, which makes the action ethically wrong. In the same manner, their killing of Charley in the second scene is also considered ethically wrong in the novel, creating a contrast between conscience of humans and machines. However, from Jack's perspective, allowing Charley to die is against his conscience, as he would be actively allowing a person to die, and therefore he rejects performing such an action. The author here invites the reader to understand the moral status and difference of actively doing harm and allowing harm by picturing the same character, Charley, in different situations. In this sense, the novel examines what it means to be a human and the line where the qualities of human life is drawn and expects readers to question what the right thing to do would be in Jack's position. The answer the author provides also helps readers to imagine themselves in such a position and shapes their moral understanding in terms of the ethical sanctions of technological inventions and their possible dangers, thus validating the initial argument of this study once more.

Conclusion

Overall, it has been argued that science fiction can function in the same manner as thought experiments and may help us navigate thorough ethical issues or dilemmas. Moreover, as exemplified in above,

a coordinated examination of such literature by using literary ethical analysis reveals this function in an elaborated manner so that one may pinpoint intended ethical invitations of the authors and possible ethical influences their works leave on the readers. Obviously, such use of ethical analysis can be used in any literary genre as it has been practiced for many years; however, its effectiveness for understanding moral issues becomes more apparent when implemented in science fiction literature due to the genre's inherent and vast imaginative nature. "In literary experience we are given the gift of identification without the pathology of delusion" (79) says Marshall Gregory and in this sense science fiction becomes a genre that has limitless space of imagination, consequently allowing the exploration of ethical issues that may not be discussed in anywhere else. Just so, Michael Crichton's *Prey* focuses on the possibility of robots' development through a claim of humanity as well as the consequent emerging necessity of early preemptive measures concerning nanotechnology or any other technological development that may be harmful for humanity if developed too far. In the novel, Jack asks a very congruous question about why they all kept working even though they knew the swarm was dangerous: "If they were all concerned, why didn't they do something about it? But of course that's human nature. Nobody does anything until it's too late. We put the stoplight at the intersection *after* the kid is killed" (64). He criticizes the inability of human beings to do the right action until an accident occurs, and here the autonomy of the swarms is more than just an accident, it threatens all humanity. Within this context, Crichton presents a powerful thought experiment, and wants readers to think about what it means to be human and what they would do if they faced such a threat. The discussion of the ethical issues creates many responses on the side of the readers, and shapes their moral and ethical understanding regarding this type of technology. In the case of *Prey*, after reading the novel, they may perhaps insist on putting that stoplight *before* the kid is killed.

Notes

[1] The hypothetical scenario of molecular nanotechnology in which self-replicating machines eat up the entities in their environment and increase their own population is called "gray goo" (or "ecophagy"). The term "gray goo" was first coined by the engineer K. Eric Drexler in 1986, and it stands as a useful term for the ethics of technology, as it is a thought experiment of worst-case scenarios of low-probability but influential outcomes of current technologies.

[2] The name of the corporation in which the protagonists work.

[3] Singularity describes the hypothetical situation of the "intelligence explosion" and the development of AI reaching to the level of human intelligence, creating their own AI systems and consequently becoming uncontrollable by humanity. Such superintelligent machines then can create other superintelligent machines easily, which makes it hard to predict the future after singularity.

[4] In "Ethical Issues in Advanced Artificial Intelligence" (2003), Nick Bostrom argues that, the superintelligent machines may be the last invention of humanity and their emergence can be sudden. Such superintelligence will lead to other advanced technologies, and eventually more advanced superintelligence, copying artificial minds will be possible, they will potentially be autonomous agents, yet they will not have humanlike motives and psyches. They may even become biology-based superorganisms and function like a global brain. However, AI machines would eventually regard their own utility, which may bring unintended or undesired consequences to human beings such as extinction. It could kill other agents or persuade them to change their behaviors, and block any attempt of interference. Such a hypothetical scenario would create an existential risk from superintelligence. Theoretically, superintelligence itself or a global disaster caused by superintelligence may cause human extinction. Poorly designed initial goals or building it to serve a selected group of humans may be the reason of its malfunctioning. According to this "orthogonality thesis," Bostrom claims that intelligent machines can be programmed into single goal and no ethical or moral rule can stop them from performing their goal.

[5] See Peter Singer, *Practical Ethics* (1979); Philippa Foot, "Euthanasia" (1977); James Rachels, "Active and Passive Euthanasia" (1975)

[6] According to this thought experiment, a runaway trolley is headed towards five people who would be killed by the trolley and it can be steered to another track where there is only one person. Usually, according to Foot, negative rights outweigh positive rights and one person's negative rights cannot be violated to meet the positive rights of others. However, in this scenario, the choice is between negative rights of one person against negative rights of many, therefore the trolley can be stared to kill the one person and minimize the damage (27).

Works Cited

Anscombe, Elizabeth. "Modern Moral Philosophy." *Philosophy,* vol. 33, no. 128, 1958, pp. 1-19. *JSTOR*, www.jstor.com/stable/3749051. Accessed 16 Jan. 2021.

Aristotle. *Nicomachean Ethics*. Translated and edited by Roger Crisp, Cambridge UP, 2004.

Asimov, Isaac. "A Literature of Ideas." *Today and Tomorrow and…*, Doubleday & Company, 1973, pp. 307-315.

Blackford, Russell. "Introduction: Science and the Rise of Science Fiction." *Science Fiction and the Moral Imagination: Visions, Minds, Ethics*, Springer, 2017, pp. 1-19.

---. "Engaging with Ethical Theories." *Science Fiction and the Moral Imagination: Visions, Minds, Ethics*, Springer, 2017, pp. 75-77.

Bostrom, Nick. "Ethical Issues in Advanced Artificial Intelligence." *nickbostrom.com*, 2003, nickbostrom.com/ethics/ai.html. Accessed 13 Feb. 2021.

Center for the Study of Technology and Society. "The Dust Bites Another One: From Michael Crichton's Prey to the Department of Nanotechnology," *The New Atlantis*, vol. 1, 2003, pp. 142. *JSTOR*, www.jstor.org/stable/43152867. Accessed 14 Jan. 2021.

Crichton, Michael. "Artificial Evolution in the Twenty-first Century." *Prey*, Harper Collins, 2002, pp. ix – xv.

---. "Could Tiny Machines Rule the World?" *Parade*, 24 Nov. 2002. *The Official Site of Michael Crichton*, www.michaelcrichton.com/prey/. Accessed 12 Mar. 2021.

---. *Prey*. Harper Collins, 2002.

Dolan, Timothy. "Science Fiction as Moral Allegory." *Journal of Futures Studies*, vol. 24, no. 3, 2020, pp. 105–112.

Driver, Julia. "The History of Utilitarianism." *Stanford Encyclopedia of Philosophy*, 22 Sept. 2014, plato.stanford.edu/entries/utilitarianism-history/. Accessed 16 Jan. 2021.

Elgin, Catherine Z. "Fiction as Thought Experiment." *Perspectives on Science*, vol. 22, no. 2, 2014, pp. 221-241.

Foot, Philippa. "Euthanasia." 1977. *Virtues, Vices, Other Essays in Moral Philosophy*. Oxford UP, 2002, pp. 33-61.

---. "The Problem of Abortion and the Doctrine of Double Effect." 1967. *Virtues, Vices, Other Essays in Moral Philosophy*. Oxford UP, 2002, pp. 19-32.

Gregory, Marshall. "Ethical Criticism: What It is and Why It Matters." *Ethics, Literature, Theory: An Introductory Reader*, edited by Stephen K. George, A Sheed & Ward, 2005, pp. 56-86.

Hursthouse, Rosalind. *On Virtue Ethics*. 1999. Oxford UP, 2002.

Kahn, Peter H., et.al. "What is a human? Toward Psychological Benchmarks in the Field of Human–Robot Interaction." *Interaction Studies*, vol. 8, no. 3, 2007, pp. 363-390.

Mach, Ernst. "On Thought Experiments." *Knowledge and Error: Sketches on the Psychology of Enquiry*, translated by Thomas J. McCormack and Paul Foulkes, D. Reidel Publishing, pp. 134-147.

Müller, Vincent C. "Ethics of Artificial Intelligence and Robotics." *Stanford Encyclopedia of Philosophy*, 30 Apr. 2020, plato.stanford.edu/entries/ethics-ai/. Accessed 22 Mar. 2021.

Nussbaum, Martha C. *Frontiers of Justice: Disability, Nationality, Species Membership*. Harvard UP, 2007.

Rachels, James. "Active and Passive Euthanasia." *The New England Journal of Medicine*, vol. 292, no.2, 1975, pp. 78-80.

Schneider, Susan, editor. "Introduction: Thought Experiments: Science Fiction as a Window into Philosophical Puzzles." *Science Fiction and Philosophy: From Time Travel to Superintelligence*, 2nd ed., Wiley Blackwell, 2016, pp. 1-16.

Singer, Peter. *Practical Ethics*. 1979. Cambridge, 2011.

Wiltsche, Harald A. "The Forever War: Understanding, Science Fiction, and Thought Experiments." *Synthese*, vol. 198, 2021, pp. 3675–3698.