From monsters to mediators: The evolution of the theme of altruism in early robotic science fiction texts

Siglé, J.A.

From Monsters to Mediators:

The Evolution of the Theme of Altruism in Early Robotic Science Fiction Texts

Jan Siglé

PhD dissertation manuscript

Leiden University

LUCAS

January 2021

Cover Image Source: *Loss of Sensation* (1935)

From Monsters to Mediators:
The Evolution of the Theme of Altruism in Early Robotic Science Fiction Texts

Proefschrift

ter verkrijging van
de graad van Doctor aan de Universiteit Leiden,
op gezag van Rector Magnificus prof.mr. C.J.J.M. Stolker,
volgens besluit van het College voor Promoties
te verdedigen op donderdag 28 januari 2021
klokke 11:15 uur

door

Jan Albertus Siglé
geboren te Pretoria, Zuid-Afrika
in 1984.

Promotor:	Prof. dr. P.Th.M.G. Liebregts
Copromotor:	Dr. E.J. van Leeuwen

Promotiecommissie:	Prof. dr. M.C.M. Corporaal (Radboud Universiteit Nijmegen)
	Prof. dr. D.A. Pascoe (Universiteit Utrecht)
	Prof. dr. F.W.A. Korsten
	Dr. Michael Newton
	Dr. Sara Polak

Table of Contents

Samenvatting

Dit proefschrift onderzoekt het thema van altruïsme in de representatie van robots in vroege sciencefiction teksten. Terwijl zulke teksten waarin robots figureren vaak thema's aanduiden die betrekking hebben op de complexe relatie tussen mens en technologie, laten verschillende vroege robot-verhalen ook een thematische preoccupatie met vragen over de menselijke natuur en welwillendheid zien. De verhandelingen over humanoïde machines als culturele objecten omvatten een reeks onderwerpen die betrekking hebben op de functies en betekenissen van humanoïde machines in de hedendaagse cultuur en sciencefiction (Haraway; Hayles; Eric. G. Wilson; Kang; Kakoudaki). Er is echter niet veel aandacht besteed aan het specifieke thema van altruïsme met betrekking tot fictieve robots, en dit proefschrift onderzoekt hoe dit thema wordt weergegeven in vroege sciencefictionverhalen met robotpersonages.

Deel I van dit proefschrift bestaat uit vijf hoofdstukken en schetst een theoretisch kader om het thema altruïsme in vroege robotische sciencefictionverhalen beter te kunnen analyseren. Elk hoofdstuk in deel I is gewijd aan een specifiek betoog dat helpt de kloof tussen mensen, machines en altruïsme te overbruggen. Hoofdstuk 1 dient ter verduidelijking van verschillende termen en definities die in dit proefschrift worden gebruikt, en om de relatie tussen mens-, machine- en sciencefictionstudies te verduidelijken. Hoofdstuk 2 onderzoekt kritische posthumanistische overwegingen die de complexe relatie tussen mens en machine, en ook samenleving en technologie verhelderen, evenals vraagstukken over de tweedeling Zelf-de Ander en posthumanistische identiteitsvorming. Hoofdstuk 3 gaat terug op de controversiële verhandeling van La Mettrie, getiteld *L'homme Machine* (1749), waarin werd gesteld dat mensen ontologisch niet te onderscheiden zijn van machines. Door mensen met machines te vergelijken, identificeerde La Mettrie een essentieel verschil, namelijk menselijke welwillendheid, die zijn eigen stelling dreigde te ondermijnen dat zo'n essentieel verschil niet kan bestaan. Als gevolg hiervan betoogde La Mettrie dat menselijke welwillendheid zowel een universele als inherente eigenschap is van mensen (en ook

van dieren), die hij de universele wet noemde. Het lezen van La Mettrie vanuit het perspectief dat in dit proefschrift is aangenomen, dient ook als historische contingentie tussen zijn ideeën en het hedendaagse kritische posthumanisme. Hoofdstuk 4 gaat in op de ideeën van Alan Turing over de relatie tussen menselijke intelligentie en kunstmatige intelligentie. Turing benadrukte dat machine-intelligentie waarschijnlijk niet zal worden gerealiseerd zonder uitgebreide samenwerking tussen mens en machine, gezien de belangrijke rol die menselijke samenwerking speelt in relatie tot menselijke intelligentie. Bovendien wordt de beroemde Turing-test van Turing opnieuw bekeken als een heuristiek die tegelijkertijd de moeilijkheden benadrukt bij het vergelijken en / of onderscheiden van mensen en machines. Dit proefschrift wijst erop dat de Turing-test werd gebruikt in sciencefiction-verhalen voorafgaand aan Turing met een vergelijkbaar effect, namelijk als een troop waarmee de categorieën mens en machine kunnen worden gedestabiliseerd (ook besproken in deel II van dit proefschrift). Hoofdstuk 5 gaat dieper in op het onderwerp altruïsme, inclusief een verduidelijking van het thema van altruïsme en hoe dit thema kan worden weergegeven in de context van robotachtige sciencefictionverhalen. Als we ons richten op recentere debatten in de evolutiebiologie, wordt het duidelijk dat een exacte wetenschappelijke definitie van altruïsme nog steeds niet onbetwistbaar is. Het probleem van altruïsme is terug te voeren op Victoriaanse denkers, waaronder Charles Darwin. Als zodanig is het minder verrassend om de mogelijkheid te beschouwen dat in vroege sciencefiction-robotverhalen verschillende ideeën met betrekking tot altruïsme zijn opgenomen als thematische zorgen.

Deel II van dit proefschrift bestaat uit vijf literair-analytische hoofdstukken van sciencefiction-verhalen over robots. Hoofdstuk 6 gaat terug naar *Frankenstein* van Mary Shelley (1818). Hoewel het kunstmatige wezen van Victor geen robot is, vestigde *Frankenstein* een traditie in robot sciencefiction die Isaac Asimov later het Frankenstein-complex noemde. Bovendien laat een nadere lezing zien dat *Frankenstein* thema's van altruïsme bevat wanneer primaire en secundaire personages altruïstisch tegenover andere personages handelen. Het verhaal bevat ook een Turing-testmoment dat de grens tussen mens en wezen doet vervagen. Hoofdstuk 7 onderzoekt *The Steam*

Man of the Prairies; of The Huge Hunter (1868) door Edward S. Ellis, die beschrijft hoe westerse kolonisten op zoek naar goud, gewapend zijn met een door stoom aangedreven robot, die schermutselingen voert met indianen. Het verhaal leunt sterk op principes van groepsselectie en vertegenwoordigt de machine als de ultieme altruïst (of bron van altruïsme). Hoofdstuk 8 onderzoekt *Tomorrow's Eve* (1886) door August Villiers de l-Isle-Adam, waarin een fictieve Thomas Edison wordt afgebeeld die een vrouwelijke androïde bouwt voor een vriend met tragische gevolgen. Het verhaal toont een aantal altruïstische gebaren tussen personages, terwijl bij nadere beschouwing blijkt dat de aantrekkingskracht van de androïde recht evenredig is met de sociale isolatie die personages ondergaan. Het verhaal bevat ook een Turing-test moment om aan te tonen dat oprechte relaties tussen mensen de tragische gevolgen aan het einde van de roman hadden kunnen voorkomen. Hoofdstuk 9 onderzoekt *R.U.R. (Rossum's Universal Robots)* (1921) van Karel Čapek die de grootschalige productie van robots weergeeft die over de hele wereld als arbeiders worden verkocht. De robots komen uiteindelijk in opstand tegen de mensheid met ernstige gevolgen. Het verhaal toont niet alleen het uitsterven van de mensheid, maar ook het uitsterven van de opstandige robots; de enige overlevenden zijn twee robots die verliefd op elkaar worden. Hun liefde wordt in het verhaal bevestigd als altruïstisch wanneer elk bereid is om namens de ander te worden ontleed (d.w.z. gedood). Als zodanig bagatelliseert het verhaal van Čapek het belang van ontologische verschillen tussen mensen en machines, terwijl het thema liefde (of altruïsme) benadrukt wordt als de bepalende factor voor de overlevingskansen van beide soorten. Hoofdstuk 10 onderzoekt de roman *Metropolis* (1927) van Thea von Harbou. De roman toont een futuristische dystopische stad die wordt onderhouden door een onderworpen proletariaat dat in ondergrondse fabrieken werkt. De stad valt in puin als een vermomde robot een revolutie onder de ontevreden arbeiders veroorzaakt. Zoals vermeld in de epigraaf van de roman, is het belangrijkste thema van de roman het thema van bemiddeling tussen een heersende klasse en degenen die worden bestuurd, vooral in het tijdperk van het industrialisme. Het thema van bemiddeling is echter volledig

gebaseerd op het thema altruïsme, aangezien dergelijke positieve vormen van bemiddeling pas in het verhaal tot stand komen door directe en openlijke altruïsme van verschillende personages.

Uit de analyses in deel II van dit proefschrift blijkt dat deze verhalen meer zijn dan louter literaire reacties op verschillende angsten over technologie in het tijdperk van het industrialisme en / of de gruwelijke nasleep van de Eerste Wereldoorlog. In plaats daarvan probeerden deze verhalen ook de menselijke conditie te verkennen en beter te begrijpen door de aard van menselijke welwillendheid te verkennen. Als zodanig draagt dit proefschrift bij aan ons begrip van de veelzijdigheid van fictieve robots en de vroege thematische ontwikkelingen van dit specifieke subgenre in sciencefiction. Het zou interessant zijn om te onderzoeken of deze thematische preoccupatie na 1930 bleef bestaan (bekend als de Gouden Eeuw van sciencefiction), en of deze thematische preoccupatie nog steeds bestaat in de eenentwintigste eeuw.

Summary

This dissertation examines the theme of altruism in the representation of robots in early science fiction narratives. While such texts often denote themes pertaining to the complex relationship between humans and technology, several early science fiction robotic narratives also reveal a thematic preoccupation with questions regarding human nature and benevolence. Discourses on humanoid machines as cultural objects encompass a range of topics pertaining to the functions and significances of humanoid machines in contemporary culture and science fiction (Haraway; Hayles; Eric. G. Wilson; Kang; Kakoudaki). However, not much attention has been paid to the specific theme of altruism in relation to fictional robots, and this dissertation sets out to explore how this theme is represented in early science fiction narratives that involve robotic characters.

Part I of this dissertation, consisting of five chapters, establishes a theoretical framework with which to better address the theme of altruism in early robotic science fiction narratives. Each chapter in Part I is dedicated to a specific discourse that helps to bridge the gap between humans, machines and altruism. Chapter 1 serves to clarify various terms and definitions employed in this dissertation as well as to elucidate the relationship between human, machines and science fiction studies. Chapter 2 examines critical posthumanist considerations that elucidate the complex relationship between human and machine, society and technology, as well as self-other dichotomies and (post)human identity formation. Chapter 3 revisits La Mettrie's controversial thesis, entitled *L'homme Machine* (1749), which argued that humans are ontologically indistinct from machines. Conflating humans with machines, La Mettrie identified an essential difference, namely human benevolence, which in turn threatened to undermine his own thesis that no such essential difference can exist. As a result, La Mettrie argued that human benevolence is both a universal and inherent trait in human beings (as well as animals), which he referred to as the universal law. Reading La Mettrie from the perspective adopted in this dissertation also serves to constitute historical contingency between his ideas and contemporary critical posthumanism. Chapter 4 turns to Alan Turing's ideas regarding the relationship between human intelligence and artificial

intelligence. Turing emphasized that machine intelligence will likely not be realized without extended human-machine cooperation given the important role that human-human cooperation plays in relation to human intelligence. Additionally, Turing's famous Turing test is revisited as a heuristic that simultaneously highlights the difficulties of comparing and/or distinguishing between humans and machines. This dissertation points out that the Turing test was employed in robotic science fiction narratives prior to Turing with similar effect, namely as a trope with which to destabilize the categories of human and machine (also discussed in Part II of this dissertation). Chapter 5 examines the topic of altruism in more detail, including a clarification of what this dissertation refers to as the theme of altruism and how this theme may be represented in the context of robotic science fiction narratives. Turning to more recent debates in evolutionary biology, it becomes clear that an exact scientific definition of altruism is still not without contestation. The problem of altruism can be traced back to Victorian thinkers including Charles Darwin. As such, it is less surprising to consider the possibility that early science fiction robotic narratives incorporated various ideas pertaining to altruism as thematic concerns.

Part II of this dissertation consists of five literary analytical chapters of robotic science fiction narratives. Chapter 6 revisits Mary Shelley's *Frankenstein* (1818). While Victor's creature is an artificial being and not an automaton, *Frankenstein* established a tradition in robotic science fiction that Isaac Asimov later referred to as the Frankenstein complex. Additionally, a close-reading reveals that *Frankenstein* incorporated themes of altruism when primary and secondary characters behave altruistically towards other characters. The narrative also includes a Turing test moment that obfuscates the boundary between human and creature. Chapter 7 examines Edward S. Ellis' *The Steam Man of the Prairies; or The Huge Hunter* (1868), which depicts Western settlers on the frontier on a search for gold, armed with a steam-powered robot, waging skirmishes with Native Americans in the process. The narrative relies heavily on group selectionist principles whilst representing the machine as the ultimate altruist (or source of altruism). Chapter 8 examines August Villiers de l-Isle-Adam's *Tomorrow's Eve* (1886) which depicts a fictionalized Thomas Edison who

constructs a female android for a friend with tragic consequences. The narrative depicts some altruistic gestures between characters, while a close-reading reveals that the appeal of the android is directly proportional to the social isolation characters suffer. The narrative also includes a Turing test moment in order to demonstrate that sincere human-human interaction could have averted the tragic consequences at the novel's conclusion. Chapter 9 examines Karel Čapek's *R.U.R. (Rossum's Universal Robots)* (1921) which depicts the largescale production of robots which are sold around the world as laborers. The robots eventually revolt against humanity with dire consequences. Not only does the narrative depict the extinction of humanity, it also depicts the extinction of the rebellious robots; the only survivors are two robots who fall in love with one another. Their love is affirmed in the narrative as genuinely altruistic when each is willing to be dissected (i.e. killed) on the other's behalf. As such, Čapek's play downplays the importance of ontological differences between humans and machines, while emphasizing the theme of love (or altruism) as the determining factor to the survivability of either species. Chapter 10 examines Thea von Harbou's *Metropolis* (1927). The novel depicts a futuristic dystopian city which is maintained by a subjugated proletariat working in subterranean factories. The city falls to ruin when a robot in disguise incites a revolution among the disgruntled workers. As stated in the novel's epigraph, the major theme of the novel is the theme of mediation between a ruling class and those who are governed, particularly in the age of industrialism. The theme of mediation, however, is entirely predicated on the theme of altruism given that such positive forms of mediation only come to be realized in the narrative through direct and overt acts of altruism by several characters.

The analyses in Part II of this dissertation show that these narratives constitute more than mere literary reactions or responses to various technophobic anxieties in the age of industrialism and/or the horrific aftermath of the First World War. Rather, these narratives also sought to explore and better understand the human condition by exploring the nature of human benevolence. As such, this dissertation contributes to our understanding of the versatility of fictional robots as well as the early thematic developments of this particular subgenre in science fiction. It would be interesting to

investigate whether this thematic preoccupation persisted after 1930 (known as the Golden Age of science fiction), and whether this thematic preoccupation still persists in the twenty-first century.

Introduction: The Future Repeats Itself

A news article published in *The Guardian*, entitled "We can Beat the Robots – with Democracy" (Bedham 2017), speculates on various technological breakthroughs and the existential threats robots pose to humanity's survival. The article, however, reassures readers that there are essential differences between humans and robots which will prove vitally important. According to the article,

> it's worth remembering that the most meaningful distinction between the machines and ourselves is our human judgment, informed not by algorithms but by values of imagination, empathy, kindness, selflessness and community. Should we not affirm the validity of our own capacity for democratic decision making, then the robots have already won. (Bedham 2017)

What the article refers to as "empathy, kindness, selflessness", this dissertation refers to simply as *altruism*.[1] Thus, the three distinctive features of humanity, namely imagination, altruism and a democratic community, will prove to be humanity's only recourse to survival in the face of a robotic uprising.

What is particularly interesting about the article is the attempt to reformulate the distinction between humans and machines when such distinctions have become increasingly challenged by proponents of transhumanism and posthumanism. While the human mind, whether seen as consisting of reason, consciousness or subjective mental experiences, has stood for a long time as the impassable barrier between humans and machines, various discourses concerning artificial intelligence suggest that the human mind can no longer suffice as a distinctive marker of human exceptionalism. According to Brian Christian, the arrival of the processor and digital mathematics constitute "a huge blow to humans' unique claim to and dominance of the area of 'reasoning'" (49). Now that machines are able to outperform humans physically and cognitively (albeit in a narrow

[1] The Oxford English Dictionary defines "altruism" as: "Disinterested or selfless concern for the well-being of others, esp. as a principle of action. Opposed to *selfishness, egoism*, or (in early use) *egotism*" (OED). The word "altruism" as it is employed in this dissertation is defined in more detail in Chapter 5.

sense), the blueprints for an improved species have been drawn. Scholars from various disciplines (Kurzweil; Bostrom; Yudkowsky) are forced to consider whether humanity will be exterminated, tolerated or integrated in the age of smart machines. While Bedham's article reformulates a set of distinctive markers, reaffirming humanity's rightful place as a species worthy of survival, one should note that machines are already adopting the very features Bedham highlights as uniquely human. The first of these, namely imagination, is being usurped by artificial intelligence in the form of General Adversarial Networks.

In 2018, Ian Goodfellow, a PhD student at the University of Montreal, managed to design a neural architecture that resembles something akin to an imagination (Condliffe). Goodfellow realized that using two neural networks pitted against one another, a generator and a discriminator, can allow General Adversarial Networks (or GANs) to produce, for example, convincing images of nonexistent human faces, cats or bicycles (Condliffe). Goodfellow's ingenious approach allows artificial intelligence the ability to *conjure up* images of things that do not exist.[2] Although the images are not always perfect, they are certainly impressive, and the implications are as terrifying as they are exhilarating:

> some experts believe there's a sense in which GANs are beginning to understand the underlying structure of the world they see and hear. And that means AI may gain, along with a sense of imagination, a more independent ability to make sense of what it sees in the world. (Condliffe)

GANs may one day allow machines to relate to the world through human conceptual understandings. In so doing, GANs will likely play an important role in the development of artificial general intelligence. Bedham's claim that imagination is a distinctive feature belonging strictly to humans is not likely to hold true for much longer.

[2] One can see the results of faces of nonexistent humans generated by GANs online at: <https://thispersondoesnotexist.com/>

Prior to the innovation of GANs, Markus Waibel (and others) demonstrated that machines are also capable of evolving altruism. In 2011, *Wired.com* published an article entitled "Robots Evolve Altruism, Just as Biology Predicts" (Keim). In a Swiss laboratory, robots not only "evolved to help each other", but did so in a manner "just as predicted by a classic analysis of how self-sacrifice might emerge in the biological world" (Keim). The experiment, entitled "A Quantitative Test of Hamilton's Rule for the Evolution of Altruism" (Waibel), involved two hundred groups of robots that evolved for five hundred generations, while inheriting various properties from one generation to the next. The researchers discovered that a biological rule concerning the evolution of altruism, known as Hamilton's rule, also applied to the evolution of robots.[3] Although it was not their primary aim to experiment with robotic altruism, the findings of their experiment nevertheless confirm that "a fundamental principle of natural selection also applies to synthetic organisms when these have inheritable properties" (Waibel 5). Robots can evolve altruism for the same reasons that humans did, namely as a result of natural selection. With such altruistic forms of social cooperation, the reverse of Bedham's title becomes equally true: Robots can beat humanity – with democracy.

In the same year that Waibel discovered robotic altruism, in her book *Alone Together: Why We Expect More from Technology and Less from Each Other* (2011), Sherry Turkle argued that society was (and probably still is) more ready than ever to integrate with robots. Turkle examined what she called the "robotic moment", which "refers to our state of emotional – and I would say philosophical – readiness […] to seriously consider robots not only as pets but as potential friends, confidants, and even romantic partners" (9). Robots can be viewed as a panacea for a wide range of social ills: "People seem comforted by the idea that if we fail each other, robots will be there, programmed to provide simulations of love" (10). Such sentiments are understandable considering the inundation of sensationalist reports about various technological breakthroughs that not merely exceed our expectations, but challenge our conceptions of what is technologically possible.

[3] Hamilton's rule is discussed in more detail in Chapter 5.

The famous 1997 chess match between world champion Garry Kasparov and IBM's supercomputer *Deep Blue* not only constituted an important milestone for artificial intelligence but also served as an ominous foreshadowing of future technological developments (Ford 99-100; Christian 102-3). IBM's next major accomplishment was *Watson*, an artificial intelligence program that defeated two champions, Ken Jennings and Brad Rutter, in the general knowledge quiz gameshow *Jeopardy!* in 2011 (Ford 100-6). In 2014, Eugene Goostman, an artificially intelligent chatbot, became the first program to pass the Turing test (Aamoth; also discussed in Chapter 4). In 2015, Elon Musk, Steve Wozniak, Demis Hassabis and Stephen Hawking, "along with 1,000 AI and robotics researchers", signed an open letter, "calling for a ban on 'offensive autonomous weapons'" (Gibbs); Musk and Hawking are convinced that "AI is our biggest existential threat", whereas Wozniak recognizes the danger but trusts in the benevolence of machines, hoping that one day they may be viewed as pets or extensions of one's family (Gibbs). In 2016, an artificial intelligence program named *AlphaGo* claimed a 4-to-1 victory against world champion Lee Sedol in the ancient Chinese game of Go (DeepMind). In 2017, an intelligent humanoid robot named Sophia obtained citizenship in Saudi Arabia (Weller). In 2018, as mentioned, Ian Goodfellow's GANs enabled artificial intelligence to undertake a variety of creative tasks (Condliffe). In 2019, Elon Musk's company Neuralink unveiled their innovative brain implants which will enable humans and artificial intelligence to "work in 'symbiosis'" (Wong). Also in 2019, a BBC article reported that Pope Francis voiced his concerns about the future of robotics and teamed up with Microsoft in order to foster public awareness (Copestake). When the director of engineering at Google, Ray Kurzweil, states that "Humans will be hybrids by 2030" (Eugenios), one inevitably recalls Donna Haraway's claim: "The boundary between science fiction and social reality is an optical illusion" (149).

This obfuscation between science fiction and reality is also posing problems for roboticists. During the *Social Robotics* conference in Sydney in 2014, Human Robot Interaction researcher Eduardo Sandoval (and others) pointed to a dilemma that roboticists must confront in the age of intelligent machines. The contradiction Sandoval (and others) outlined consists of "the expectations

created by the media", on the one hand, and "the real capabilities of the robots and the needs of people", on the other (59). This places roboticists in a difficult position because consumers "want a mixture of the subtle interactive abilities similar to fictional robots" (61), while such fictional robots are "inspired, consciously or unconsciously, by models of human-human interaction" (56). In other words, the robotic industry must conform to the standards set by science fiction, and humans would rather interact with robots that pretend to be human. As Kathleen Richardson states, in *An Anthropology of Robots and AI* (2015), while Haraway's claim about science fiction and social reality is certainly true, one should not ignore that "the Real is continually asserting itself in the making of robots" (Richardson 4). Sandoval and Richardson share a similar observation, namely that science fiction and technological reality, while informing one another, do not align perfectly and still produce awkward contradictions. Not only are science fiction robots influencing the manufacturing and innovation of actual robots, but they have also come to shape our expectations of the future.

One can make a distinction between fantastic and pragmatic perspectives regarding the future of technology. Pragmatic perspectives tend to focus on near-future events. In his book *The Rise of the Robot: Technology and the Threat of Mass Unemployment* (2015), Martin Ford is concerned with the argument that the automation of physical and intellectual labor is likely to have dramatic consequences in the near future not only because the technology is improving, but also because the cost of implementing such technologies is reducing. Ford points out that if such largescale automation occurs (and results in mass unemployment), the real culprit is not technology but the economy: "The progression toward ever more automation is […] fundamentally driven by capitalism" (Ford 254). Navigating such a future might be prove to be "the greatest challenge for our time" (Ford 285). In *Imagining Slaves and Robots in Literature, Film and Popular Culture* (2015), Gregory Jerome Hampton repeats Ford's concerns in his analysis of literary and cinematic science fiction robots. Hampton argues:

> With the creation of a humanoid robotic workforce America will produce a high-tech form of chattel slavery that will undoubtedly produce many, if not all, of the same detrimental

> effects created by American chattel slavery […] It is safe to assume that initially human unemployment rates will reach all-time highs in industries becoming most dependent on robotic labor […] Ultimately, the production cost of a robot will decrease and the domestic robot will become normative and robot/human relations will truly begin to be tested. The new slavery (techno-slavery) will function as a wedge issue for labor movements of tomorrow. Techno-slavery will force human society to consider how much of its humanity it will forego in order to sustain its standard of living and rate of technological advancement. (80-1)

What Ford and Hampton illustrate is a much more serious cause for concern when it comes to artificial intelligence and robotics. Fantastic scenarios about robotic uprisings or super intelligent minds are not necessarily invalid, but rather premature at this stage.

Fantastic scenarios are typically predicated upon the notion of the Singularity, which Brian Christian summarizes simply as "a moment when we make machines smarter than ourselves, who make machines smarter than themselves, and so on […] toward a massive ultra-intelligence that we can barely fathom" (Christian 263). Such ideas are popular enough that even science fiction narratives have come to satirize them, for example Daniel Wilson's *How to Survive a Robot Uprising* (2005) and Peter Clarke's *The Singularity Survival Guide: How to Get on the Good Side of Your Future Robot Overlords* (2019). The latter novella is narrated, ironically, by an artificial intelligence designed to help humanity survive the arrival of smarter and better artificial intelligences. However, the advice provided is ultimately comedically absurd and fatalistic, similar in style to Douglas Adams' *The Hitchhiker's Guide to the Galaxy* (1979). Fantastic scenarios about the future plight of humanity against smart machines typically come in one of three varieties, which Minsoo Kang summarizes as "the theories of inevitable confrontation", "equivalence through sentience", and "cybernetic mergence" (300). The first, inevitable confrontation, consists of robotic revolts, human resistance movements and an epic battle for survival on both sides, such as those depicted in the *Terminator* and *Matrix* films. The second, equivalence, deals with scenarios in which robots are intelligent or

sentient enough to obtain citizenship and coexist with humans in society, or constitute a labor force within society as second-class citizens; Asimov's robots whose conduct are governed by the Three Laws would constitute such a scenario. The last, cybernetic mergence, deals with scenarios in which humans become technologically enhanced, or cyborgs, as a means of keeping up with the advanced capabilities of machines; Ramez Naam's science fiction novel, entitled *Nexus* (2012), depicts a scenario in which nanotechnology allows human brains to be networked directly to one another. The attempt to uncover which scenario is more viable is also to miss the point. According to Jaron Lanier, such perspectives introduce a very different threat altogether:

> Singularity books are as common in a computer science department as Rapture images are in an evangelical bookstore. […] If you believe the Rapture is imminent, fixing the problems of this life might not be your greatest priority […] In the same way, if you believe the Singularity is coming soon, you might cease to design technology to serve humans, and prepare instead for the grand events it will bring. (25)

The risk is that technological innovation becomes increasingly antihuman in its orientation because the future is not only set but decidedly not intended for humans. This dissertation shares Lanier's concern that "[t]he digital hive is growing at the expense of individuality" (26). This is mainly because "[i]deas that were once tucked away in the obscure world of artificial intelligence labs have gone mainstream in tech culture. The first tenet of this new culture is that all of reality, including humans, is one big information system" (27). Not only are humans part of this big information system, but the future also belongs to it. In the tech industry, the near future has remained stable and predictable for the last fifty years.

In 1965, Gordon Moore, then CEO of Intel, published a paper in which he observed a trend regarding the development of processing power which became known as Moore's law (Ford xii). According to Moore's law, by now far removed from its original context, processing power doubles roughly every two years (Ford xii). For example, if the same rate of development applied to bicycles, a bicycle with an initial top speed of 15 km/h would be capable of 480 km/h after a single decade;

15,360 km/h after another decade; and 491,520 km/h by the end of the third decade. Such was the rate of development for computing power for five consecutive decades, from 1965 to 2015, implying twenty-five doublings during that time. Ford suggests that the actual number is twenty-seven doublings (xiii). However, in 2015, the same Gordon Moore announced: "I see Moore's law dying here in the next decade or so" (Courtland). This rate of innovation is slowing down and it is still unclear what kind of technology will replace silicone processors to reinvigorate this trend (Ford 72-3). Furthermore, despite Moore himself being doubtful of the Singularity (Ford 253), speculations about fantastic futures are still rife.

These insights raise a number of important questions and implications for science fiction studies concerning the role of humanity in the context of advanced technological futures. Russel Blackford, in *Science Fiction and the Moral Imagination* (2017), argues that the future also played an important role in the early development of science fiction. The Industrial Revolution meant that "Western civilization experienced something altogether new: continual […] change that was driven and shaped by advances in technoscience. And so, humanity discovered the future" (Blackford 1). In *The Literary Imagination from Erasmus Darwin to H. G. Wells: Science, Evolution, and Ecology* (2012), Michael Page places a similar emphasis on the importance of discovering the future in relation to Erasmus Darwin's ideas of evolution:

> Not only did the past suddenly expand backwards from around 6,000 or so years to millions (and later billions), but the future expanded from an impending Christian apocalypse to a far-off future that may not even concern itself with human beings. (2)

With the discovery of the future, as Blackford suggests, science fiction can also be understood "as a literary response – not the only one – to humanity's discovery of the future" (8). Peter Clarke's *Survival Guide* (mentioned above) is a contemporary example of science fiction reacting to recent technoscientific innovation and discourse pertaining to the Singularity as his novella satirizes many of the ideas associated with it. Hampton's analysis of literary and cinematic robots (mentioned above) similarly concludes with an emphasis on the relationship between the past and the future: "It

is only through a lens of the past that a future can be accurately imagined" (82). The future of humans and robots seems to have already been largely determined; the near future, according to Ford and Hampton, is a form of techno-slavery; the distant future will be one of three possible outcomes: conflict, equivalence or mergence. However, Bedham's insistence on the viability of human benevolence is a simpler approximation of very similar arguments and concerns raised by various thinkers already mentioned: for Hampton, the fear is that humans will exploit "others," recreating the past by means of slavery; for Richardson, the fear of intelligent machines and human-robotic relations is the result of humans no longer being able to bond with other humans;[4] Turkle's robotic moment is another instance of the failure of human-human interactions in a hypermediated technological society; Lanier expresses the concern of the loss of individuality and the very antihuman approach of the tech industry at large. These anxieties are encapsulated in Bedham's simple claim that human benevolence will prove vitally important in the context of intelligence machines one day; humans will have to employ their imaginations, become altruistic and cooperate with one another if, as her title states, we intend to beat the robots with democracy.

Given that in this hyper-technological age the recurring moral seems to be human cooperation and benevolence, the question arises whether this particular perspective was also present in early science fiction narratives that deal with robots. As Kakoudaki points out: "Whatever fantasies they may facilitate, by their very presence artificial people help naturalize everyone else in a text as more reliably human, before interrogating the very nature of that assumed stability" (213). It is this potential to both assert and question the human condition that makes fictional robotic characters particularly interesting. In addition, one robot implies the possibility of manufacturing many, and therefore also implicitly introduces existential threats to humanity at large. This dissertation examines early science fiction narratives from the early nineteenth to the early

[4] According to Richardson: "A human-robot attachment is only possible because of this mechanistic sociality that underscores contemporary sociality. The mechanical sociality is an outcome of an attachment crisis in how humans bond with others. Attachment wounds are an outcome of a lack of bond between one human and another and the attempts to use machines to help fill these gaps in social relations" (131)

twentieth centuries that feature robot characters and examines the various treatments of the theme of altruism, paying particular attention to altruistic behaviors and various forms of social cooperation. As a result, this dissertation demonstrates that science fiction robots are also very much about humans and not merely about various technological speculations. As Rod Grupen states: "At bottom, robotics is about us […] It is the discipline of emulating our lives, of wondering how we work" (qtd. in Hapgood). As this dissertation will demonstrate, the same idea can apply to early science fiction robots as well.

This dissertation is divided into two parts. Part I establishes a theoretical framework that serves to elucidate the theme of altruism. This framework will function as a lens through which to focus on the theme of altruism in the selected narratives analyzed in Part II. Chapter 1 begins with an overview of the differences between robots, androids and cyborgs before critically exploring science fiction scholarship that examines robotic characters. Chapter 2 turns to certain tenets of critical posthumanism that, much like science fiction, have also come to examine the question of what it means to be human in relation to advanced technologies. Chapter 3 takes a historical turn, looking particularly at eighteenth-century materialism as expounded by La Mettrie's human-machine thesis, which emphasized the importance of morality in a mechanical universe. Chapter 4 revisits Alan Turing's heuristic test for assessing machine intelligence as well as some of his ideas regarding the evolution of intelligent machines and how it relates to humans. Chapter 5 examines historical and contemporary ideas about the dynamics of evolutionary altruism. These chapters together provide a useful framework with which to examine themes regarding altruism in the selection of literary texts.

Part II of this dissertation consists of five chapters, each examining a particular work of early robotic science fiction. The first, Chapter 6, begins with Mary Shelley's *Frankenstein* (1818; 1831) and illustrates that Victor's decision to destroy the creature's mate for the sake of humanity is

motivated by an altruistic concern.[5] Chapter 7 examines Edward S. Ellis' *The Steam Man of the Prairies* (1868) and demonstrates a similar preoccupation with the theme of altruism, but also exemplifies that evolutionary altruism is not everywhere and always a good thing. Chapter 8 examines Auguste Villiers de l'Isle-Adam's *Tomorrow's Eve* (1886) demonstrating a problematic engagement with altruism when the characters' motivations and sincerity are consistently undermined and portrayed as questionable. Chapter 7 examines Karel Čapek's *R.U.R. (Rossum's Universal Robots)* (1921) which reveals a much more direct engagement with the theme of altruism. Lastly, Thea von Harbou's novel *Metropolis* (1927) is examined as an example that engages more directly with the theme of altruism. What these readings demonstrate is not only that altruism was a prevalent theme in early robotic science fiction, but also an interesting and dramatic change in the treatment of this theme in pre- and post-World War I narratives. Although discussed in more detail later, Kang's observation that robotic science fiction narratives took on pejorative connotations after WWI can be correlated to the discoveries of this dissertation, namely that interwar robotic science fiction narratives also began to emphasize altruism as a major theme. When technology becomes monstrous, oppressive and antihuman, one finds the call for the need for human benevolence, a call often ignored in the context of technological discourses. Situating these discoveries next to Bedham's claim in the opening of this chapter also serves to establish historical contingency, revealing that our contemporary fantasies and anxieties of technology are not that different from historical ones. Perhaps the future is repeating itself.

[5] The important question is to consider whether the creature qualifies as a threat to humanity at large or only to Victor himself. Richardson considers Karel Čapek's *R.U.R.* as the "first fiction in which all humanity is destroyed. Not even in our tales of Frankenstein […] is all of humanity under threat" (132). While Richardson's description is more or less correct in that *Frankenstein* does not feature any explicit extinction events, it is important to note that the threat of such an event is certainly present in the narrative. This dissertation argues that the creature does pose a threat to humanity at large, as expressed by Victor himself, which means that his decision to destroy the creature's mate constitutes an act of altruism.

Part I

Introduction to Part I

Chapters 1 to 5 below examine various discourses and thinkers to outline a theoretical framework with which to elucidate the themes of altruism as examined in Part II of this dissertation. Chapter 1, divided into two sections, discusses the differences between fictional robots, androids and cyborgs, and whether these differences are useful when analyzing such characters. In addition, the importance of anthropomorphism is also briefly discussed. The second section of Chapter 1 turns to science fiction scholarship in order to discuss the complex relationship between science fiction and the history of robots. Chapter 2 turns to critical posthumanist discourses that explore the complicated relationships between humans and technology more generally, but also the formation of a (post)human subjectivity more specifically. Chapter 3 revisits La Mettrie's *L'Homme Machine* (1747) or *Man a Machine* which already prefigured many of the same considerations made by posthumanists (as discussed in Chapter 2). What is particularly relevant and often overlooked is La Mettrie's considerations regarding moral behavior and how morality is nevertheless possible despite his insistence that humans are mere machines. Chapter 4 revisits Alan Turing's ideas regarding his Turing test (and its implications) which proposes an experiment with which to assess whether or not a machine can be considered intelligent. Particular attention is also paid to Turing's ideas regarding machine intelligence as dependent on human interactions, as well as the important role of cooperation between machines but also between humans and machines. Chapter 5 turns to what this dissertation describes as the theme of altruism by examining various scientific discourses pertaining to altruism in order to extrapolate a heuristic framework through which various dimensions of altruistic behaviors can be examined in literary narratives. Despite the intricacies of the various technical and scientific discourses, one discovers a simple and intuitive dynamic that helps to account for when and why altruism tends to surface.

These chapters serve to show that questions regarding human altruism in relation to technology are not contemporary developments, but have existed for a long time. In robotic science

fiction narratives, analyzing human-technological relationships can result in a lack of appreciation of the human-human relationships in these texts and their conditioning under the context of technological determinism. Examining the altruistic dimensions of such narratives reminds us that science fiction robots are also about people, and not simply about various anxieties regarding technology.

Chapter 1: Science Fiction and Robots

Robots, Androids and Cyborgs

This chapter examines the differences between robots, androids and cyborgs in the context of science fiction before turning to relevant discourses on science fiction and robots more generally. According to Brian Stableford, the word "robot" was coined by Karel Čapek's brother Joseph to mean forced labor (442).[6] However, Stableford explains that a more contemporary definition came about when the "term was subsequently applied in a looser sense to industrial machines substituting for human laborers on automated production lines […] and to machines capable of self-determined locomotion" (442). There is a certain degree of semantic overlap with its older equivalent, namely the automaton. Kang explains that an automaton can denote three different things, namely a self-moving machine (such as a wristwatch), a machine specifically designed to "mimic a living creature", or a "person who acts like a machine in some way" (7-8). Kang also points out that "In the twentieth century, the word [automaton] has been eclipsed to a great extent by the modern 'robot'" (8). In the context of science fiction, the *Oxford English Dictionary* defines a robot as "An intelligent artificial being typically made of metal and resembling in some way a human or other animal" (OED). Such a definition is sufficiently specific while not being overly prescriptive, and applies to all the artificial characters examined in this dissertation with the exception of the Steam Man, given that the Steam Man is not intelligent. However, the Steam Man character qualifies as a "mecha" (often shortened to "mech"), a term borrowed from Japanese anime and manga to mean: "a large armored robot, typically piloted by a person or creature inside the robot itself" (OED). According to these definitions, the Steam Man is a kind of robot that is piloted by a human (also discussed in Chapter 7).

Stableford explains that the origins of the word android can be traced back to "alchemical literature […] with rumored attempts to create 'homunculi' by such alleged practitioners as Albertus Magnus and Paracelsus" (22). In science fiction, according to Stableford, androids are differentiated

[6] Also discussed in Chapter 8 of this dissertation.

from robots because they have "synthetic flesh rather than inorganic components" (22). However, Stableford is quick to point out that "The usage is not consistent" (22). The *Oxford English Dictionary* similarly posits a vague and short definition of an android: "An automaton resembling a human being" (OED). Both definitions imply that the constitution of an android is less important, so long as it resembles a human. The third category, cyborg, is a "contraction of 'cybernetic organism', contrived to describe products of organic/inorganic chimerisation, particularly the augmentation of the human body with mechanical devices" (Stableford 114). According to the *Oxford English Dictionary*, a cyborg is defined as "A person whose physical tolerances or capabilities are extended beyond normal human limitations by a machine or other external agency that modifies the body's functioning; an integrated man-machine system" (OED). Both definitions allow for an innumerable amount of possible combinations of different assemblages, while remaining specific enough to distinguish cyborgs from robots and androids.

In general terms, one can say that a robot refers to mechanical beings, as the term shares a certain semantic overlap with automaton as its predecessor. Robots may resemble humans, but do not have to, For example, the robots in the Hollywood adaptation *I, Robot* (2004) consists of humanoid machines that resemble humans, whereas *Wall-E* (2008) is also a robot but does not resemble humans. Abraham Merritt's *The Metal Monster* (1920) contains a variety of robotic creatures not resembling human forms, while John Wyndham's "The Lost Machine" (1932) is a first-person narrative about a robot that is described as a coffin with eight legs. Androids imply synthetic humans or artificial beings specifically designed to look like humans; their exact constitutions are less important. The most famous example is arguably Philip K. Dick's *Do Androids Dream of Electric Sheep* (1968), in which the Voigt-Kampff test (measuring empathetic responses) is employed as a means with which to distinguish humans from replicants (with questionable accuracy). Cyborgs may or may not resemble humans, but must consist of assemblages of organic and inorganic materials. For example, the character of Robocop in *RoboCop* (1987) is part human (a brain and vital organs) and part machine (a mechanical chassis); the character of Will Carter played by Johnny Depp in

Transcendence (2014) downloads his consciousness into a computer, and can also be considered a cyborg as a result. While these terms remain popular, this dissertation considers them problematic when it comes to performing literary analyses of such artificial characters. For example, in *R.U.R.*, robots are made of synthetic materials mixed in large vats while remaining indistinguishable from humans in appearance, yet they are famous for being *robots* and not androids (as stated in the title of the play). The prefix robo- in *Robocop* suggests that it is a robot, while the character is more accurately a cyborg. All three terms may be applied to the T-800 character from the *Terminator* franchise. The T-800 refers to itself as a "cybernetic organism" (or cyborg) in the second film (1991), given that it consists of a robotic endoskeleton covered with synthetic skin. While the T-800 infiltrates 1980s and 1990s American society, it dons its synthetic skin and is indistinguishable from humans, qualifying it as an android. In glimpses of the postapocalyptic future, the T-800 is shown without synthetic skin because there is no longer a need to infiltrate human society, thus making it entirely robotic. Such transformations, however, have no bearing on the character which remains stable and uniform. Conversely, the T-800 evolves from being an antagonist to a protagonist in the first two films, but does so without having to change its constitution as a cyborg.

In other words, when analyzing such characters, these designations fail to reveal anything informative about their functions and significances within their respective narratives. The argument is not that these terms are entirely unusable or uninformative. One can certainly surmise in general terms that androids, by virtue of being doppelgängers, are more likely to incorporate themes regarding identity formation; cyborgs, by virtue of enhancing or altering the human condition via technology, are more likely to incorporate themes regarding technological monstrosities; robots, by virtue of emulating humans through artifice, are more likely to incorporate themes regarding free will and oppression. However, in the context of this dissertation, such generalizations are not particularly informative given that any such association is not a matter of veracity but convention. Whether dealing with a cyborg, robot or android, it is important to take each character in turn and not presume anything on the basis of these categories. Given the selection of texts in Part II, the

term robot (and robotic) remains the most appropriate to describe these characters in general terms, and where relevant, further clarification is provided.

Kathleen Richardson, in her book *An Anthropology of Robots and AI* (2015), discusses the role of anthropomorphism in relation to real robots, especially in the context of developing social robots which are designed specifically for the purposes of interacting (socially) with humans. Richardson explains that "anthropocentrism is not merely a frame within which to understand human interactions with technologies that have human qualities, but is also a means by which to rethink the importance of human sociality in underscoring interactions with nonhumans" (74). In the pursuit of social robots, questions regarding the role of anthropomorphizing machines is an important one as it determines the effectiveness with which robots may one day integrate into society. In their article, "Anthropomorphism as a Function of Robot Group Membership" (2012), Friederike Eyssel and Dieta Kuchenbrandt discovered in their experiments that "participants rated the in-group robot more favorably and anthropomorphized it more strongly than the out-group robot. That is, given the same background information about the robot (i.e. a picture), the mere manipulation of group membership affected subsequent evaluations of the robot" (729). The more the robot seems to belong to the same group as humans, the more likely humans tend to anthropomorphize them. While this is true in social robotics research, this dissertation is more focused on how groups are established in early science fiction narratives concerning robots. The term 'group' can denote any loose aggregate of people based on any perceived similarities as exhibited by the characters in the narrative. Although anthropomorphism plays an important role for the development of social robotics, this dissertation agrees with Gary K. Wolfe's realization that the question of anthropomorphizing machines has been present since the first science fiction robot narratives. As Wolfe argues in "Icon of the Robot" (1979), in relation to science fiction robots:

> This implied mythology is evident in even the earliest robot stories of modern science fiction […] once a simple function of human being has been replaced, where will it stop? And if a

> single human function can be supplanted by a mechanism, is it not possible that all human functions might one day be so replaced? (155)

If the implied mythology is to eventually replace all human functions, then social robotics is hard at work to fulfill this very mythology – not merely by replacing specific human functions, but also replacing human-human interactions as well. The ways in which science fiction deals with robots in turn define how humans come to see themselves, and the notion of social robots is merely a continuation of this very old process. As Wolfe argues:

> Technology not only creates new environments for humanity, it also creates new images of humanity itself, which tend to mediate between the natural environment of mankind and the artificial ones it has created, between the past and the future, and between the known and the unknown. (151)

The Singularity would signify the ultimate amalgamation of humans and machines, even if by means of extermination.

Science Fiction and Robots

In *Science Fiction and the Moral Imagination* (2017), Russell Blackford briefly discusses the overlap between science fiction and posthumanist discourses. Many science fiction authors, according to Blackford, indirectly engage with posthumanist considerations, such as "the demise of human exceptionalism" and "the future of humanity as a species" (179). He points out that science fiction authors "were doing this before posthumanism and transhumanism were heard of" (179). What is relevant for this dissertation is Blackford's observation that while transhumanism originated "in something like its current form during the 1980s", one should not neglect "proto-transhumanist ideas expressed […] early in the twentieth century" (179). Blackford's analysis of science fiction is premised on its relation to what he calls "the moral imagination" (14). This is not to be confused with philosophies of morality or ethics, as he explains, but rather to make the claim that science fiction can be studied by examining its "contributions to humanity's moral imagination", and thereby "examine SF's engagement with moral questions mainly for its own sake" (15). In this context

outlined by Blackford, the question of science fiction robots and altruism becomes all the more pertinent.

Approaching the emergence of science fiction from a different perspective, John Rieder in *Colonialism and the Emergence of Science Fiction* (2008) provides many innovative readings of early science fiction narratives. According to Rieder, the nature of humankind is central to any body of literature, but "scientific accounts of humanity's origins and its possible or probable futures are especially basic to science fiction" (2). He examines the "complex mixture" of Darwinian ideologies concerning "competition, adaptation, race and destiny" as prevalent "thematic material of early science fiction" (2). It is no surprise, then, that science fiction originated in countries "most heavily involved in imperialist projects – France and England" and later, "the United States, Germany and Russia" (3). Rieder is careful to point out that his fundamental argument is not that "colonialism is science fiction's hidden truth", but rather: "science fiction exposes something that colonialism imposes" (15). Through science fiction's complex treatments of such themes, it is important to "determine how and to what extent the stories engage colonialism" (3). Given Rieder's insights, one must distinguish between early science fiction and texts involving fictional robots. While Rieder identifies a link between the emergence of science fiction and colonialism, which is likely to play a part in robotic science fiction during this era, robots as cultural and fictional objects also precede the history of science fiction. Examining early robot narratives, then, introduces a complex mixture of meanings in a vast network which cannot be fully elucidated in a single dissertation. While this dissertation therefore does not deny the expansive histories of both science fiction and robots, it must by necessity narrow its focus onto a specific thematic concern which, as mentioned, is the various engagements with themes of altruism.

When analyzing fictional robots more generally, some scholars adopt a transhistorical perspective, spanning from antiquity to Hollywood films. Such transhistorical approaches include Eric G. Wilson's *The Melancholy Android: On the Psychology of Sacred Machines* (2006); Minsoo Kang's *Sublime Dreams of Living Machines: The Automaton in the European Imagination* (2011); and

Despina Kakoudaki's *The Anatomy of a Robot: Literature, Cinema, and the Cultural Work of Artificial People* (2014). Other approaches that may be described as more synchronic include a much larger body of science fiction scholarship, such as Patricia S. Warrick's *The Cybernetic Imagination in Science Fiction* (1980), focusing particularly on robotic science fiction from 1930 until 1977; and Adrienne Mayor's *Gods and Robots: Myths, Machines and Ancient Dreams of Technology* (2018) which focusses on transcultural myths of artificial humans in antiquity.

Warrick's analysis, ranging from 1930 to 1977, covers as many as "225 short stories and novels" specifically related to robots and computers (xv). For Warrick, cybernetics "comprises all systems, mechanical and biological, in which information plays a role", and "is concerned with governance or control in social systems and in mechanical systems" (9).[7] She identifies three kinds of systems, namely isolated, closed and open, and subsequently categorizes her selection of science fiction narratives in relation to each of these categories. The closed-system model is a reductionist model, concerned with mass and energy, and predicated on classical mechanics (Warrick 97); many of these narratives can also be described as dystopian. In contrast, the open-system model is closer to biological or living systems, "steady-state systems that exist in nonequilibrium", and capable of transformation that always involves "moving toward increased complexity" (Warrick 98). The isolated-system model is an ideal that does not exist in the real world, and consists of an "assemblage or combination of things that is uninfluenced by anything outside it" (Warrick 98-9). Warrick posits that traditional literary devices (plot, character, setting, etc.) are "secondary to other elements that are the central concerns of SF" (81). The central concerns of science fiction, according

[7] It is worth mentioning that Warrick also discusses J. C. Maxwell's paper presented in 1868 to the Royal Society discussing the use of "governors" on steam engines, and links the etymological root of the Latin *gubernator* to the Greek *kybernētēs* (9). Hayles also retraces the history between governors and cybernetics when explaining the role of feedback loops in the early developments of cybernetics: "Feedback loops had long been exploited to increase the stability of mechanical systems, reaching a high level of development during the mid-to-late nineteenth century with the growing sophistication of steam engines and their accompanying control devices, such as governors" (8). Norbert Wiener, in his book *The Human Use of Human Beings* (1954), explained his motivation for employing the term cybernetics: "Hence, 'Cybernetics', which I derived from the Greek word *kubernētēs*, or 'steersman', the same Greek word from which we eventually derive our word 'governor'" (15).

to Warrick, can be identified as five elements that determine the strength of a science fiction narrative, namely scientific knowledge, novelty, dislocation in space or time, an "*awareness of unity* [original emphasis]", and an emphasis on the mind (as a genre of ideas) (82-4). With these five elements successfully combined in a narrative, a sixth criterion measures the success of the overall narrative, namely that the reader experiences a new awareness, surpassing one's "previous perceptions of time and space" (Warrick 84). Although such criteria are perhaps overly prescriptive for a general approach, Warrick's analysis of robots and computers highlights the importance of the relationship between science and fiction: "We need to remind ourselves constantly that science and poetry are not separate universes; the working of the scientific imagination and the literary imagination parallel each other" (235). Thus, science fiction must "grow with science" and not oppose it: "it cannot react against science and still be *science* fiction" (237). Such a union between "the literary imagination that immerses itself in science" can "lead us, intelligently and humanely, into our future" (237).

Eric G. Wilson's analysis also employs a transhistorical approach from a psychoanalytic perspective. Connecting the practice of mummification in ancient Egypt, the figure of the golem and various representations of automata, Wilson's focus is primarily on the relationship between creators and their creations: "the androids emerging from human imagination constitute psychic projections as much as physical collections" (2). The practice of creating artificial doubles represents the "physical activities (the descent into the unconscious)" which "are analogous to spiritual activities (the ascent to soul)", and despite their opposition, "they reach the same end" (4). In both cases, one finds the recursive movements, "from visible to invisible, outside to inside, known to mysterious" (4). Thus, to "make an android" is to "walk this razor's edge between transcendence and neurosis" (4). His discussions ultimately conclude open-endedly:

> From these tensions stems the nervousness most feel in this hyper-technological technological age. Those who love an organic cosmos suffer sporadic guilt from their repressed affection for their machines. Those who laud a mechanistic universe suspect they

> are divorcing themselves from the vitalities that haunt their dreams. Both types attempt to suppress this chronic disorientation. But to repress is to grant control. […] Both nature lover and computer maven risk becoming somnambulists: organic machines, mechanistic organs, living systems unaware of the forces driving them. (133-4)

Wilson argues that by embracing these figures, bringing them "into the playful light of the mind, to examine alike their accursed qualities and sacred potentials" will "force us to acknowledge seemingly inhuman beings […] as our own" (139). Wilson's analyses are particularly poignant at examining the very human psychological dimensions and tensions in artificial characters.

Minsoo Kang's transhistorical approach also highlights the various positive and negative reactions to automata throughout different historical periods. His analysis similarly reflects on the inherent duality of automata as transgressors of preestablished oppositions:

> […] for a full understanding of the automaton motif in the Western imagination as a whole, one must take into account both aspects of the mechanical entity, to see how the object functioned in different historical contexts as the representation of *both* human empowerment *and* oppression, liberation *and* subjugation, transcendence *and* debasement. (305)

As Kang summarizes, during the medieval and Renaissance periods, the automaton reflected "the danger of demonic agency" and/or "the marvelous potential of natural magic"; during the Enlightenment, it became stripped of its "magical aura" and reflected "the rational order of the world, the state, and the body"; during the late Enlightenment, it reflected "a shabby representation of people lacking freedom either through oppression or conformism"; during the Romantic era, the uncertainty of the times meant the arrival of the "uncanny automaton appearing in literature"; during the industrial era, machines were "portrayed as living superhuman creatures", which became "significantly darker and cautionary in the period after World War I" (304-5). Kang emphasizes that this ability "to hold such disparate meanings" is what makes the automaton "such a powerful and enduring conceptual object" (305).

Despina Kakoudaki in *Anatomy of a Robot: Literature, Cinema, and the Cultural Work of Artificial People* (2014) adopts a transhistorical approach, examining robots from antiquity to contemporary Hollywood cinema. She employs four "networks of meaning" spanning two millennia of artificial people; these networks include "artificial birth", the "mechanical body", artificial slaves and "the interpretation of artificiality as an existentialist trope" (26). These four elements can be found throughout historical and modern narratives of artificial people, and serve to reveal how "ancient ideas of animation and transformation inform modern legal and political narratives of objectification and subjectification, while the mechanical understanding of the body becomes associated with questions of control, agency, and psychological interiority" (212-3). Tracing the modes of thought throughout various historical periods, Kakoudaki concludes that the "contemporary theories of post-humanism" which seek to "redefine the human in a wider, more embodied, and more ecological frame of reference" must "counter nostalgic or apocalyptic trends that promote new body-phobic dualisms, notions of high tech transcendentalism and cerebral supremacy" (218-9). Kakoudaki convincingly demonstrates that artificial people have functioned, for a long time, as tropes with which to navigate "intimate philosophical questions", such as: "What does it mean to be alive? […] And how we can regard and treat one another without the prospect or threat of objectification?" (28). This latter question, in the context of this dissertation, is similarly preoccupied with ideas of altruism.

While these examples of artificial people and robotic scholarship inform the underlying perspective of this dissertation, it is also important to point out the differences in intention and methodology. Given the expansiveness of such transhistorical approaches, one rarely encounters close readings of specific texts. As Kakoudaki notices in relation to Hayles and Haraway, such "philosophical interpretations tend to reference literary and cinematic texts but quickly move on" (8), whilst later acknowledging herself that: "I use quick readings of well-known texts and films in order to ground complex philosophical or transhistorical questions" (27). While the theoretical authors discussed thus far have contributed invaluable research towards the broader aim of

understanding the various modes and semiotic structures of meaning of automata throughout history, this dissertation makes a far more localized and specific contribution to these considerations, one that Warrick, Wilson, Kakoudaki and Kang do not engage with, namely the ways in which early robotic science fiction deals with themes of altruism. Such a contribution necessitates close readings as opposed to "quick readings" of the selected texts. Whereas Warrick deemphasized the importance of traditional literary devices, this dissertation approaches robotic science fiction from the opposite direction: the various manners in which altruism features in these texts only become evident when focusing on characters' behaviors, sentiments, motivations and interrelationships with other characters.

It should also be explained that, as many of these authors similarly explain (Warrick xiv; Kakoudaki 27), the prevalence of automata throughout history in the Western imagination inevitably means the elimination of many science fiction narratives for the sake of scope. This dissertation similarly selected five well-established works of robotic science fiction in its early years, namely *Frankenstein* (1818), *The Steam Man of the Prairies; or The Huge Hunter* (1868), *Tomorrow's Eve* (1886), *R.U.R.* (1921), and the novel (not the film) *Metropolis* (1927). This selection is made on the basis of the various idiosyncratic treatments of altruism by each narrative. This selection does not suggest that other texts from the same period fail to incorporate this thematic dimension, nor to assert that all such narratives include these themes, but merely that this particular selection of well-known early science fiction robots already serves to demonstrate the fundamental claim that early science fiction robots engaged with themes of altruism in idiosyncratic ways.

Examples of science fiction narratives that were omitted include texts such as Edgar Allan Poe's "The Man That Was Used Up" (1839), in which a general, consisting merely of a "large and exceedingly odd looking bundle of something", is physiologically reassembled with prosthetics (40). Herman Melville's "The Bell Tower" (1856) includes an unfortunate accident when the protagonist is killed by a mechanical human figure. Ambrose Bierce's "Moxon's Master" (1899) is about an inventor who is killed by his own chess-playing machine. Although one can speculate whether this is

an early warning of artificial intelligence, the reliability of the narrator casts doubt as to the validity of such an interpretation: “This sequence of events is usually interpreted as a story about a vicious robot that murdered its inventor, but it is much more likely to be a murder mystery in which the young man was skillfully misled by the actors” (Bleiler 65). George Haven Putnam’s *The Artificial Mother; A Marital Fantasy* (1894) is about a mechanical babysitter who competes with the biological mother, although the entire event is revealed to be a dream. E. V. Odle’s *The Clockwork Man* (1923) features a cyborg amalgamation of a human and clockwork mechanics, sent from the future to warn humanity not to create the very future from whence it came. Gaston Leroux’s *The Bloody Doll* (1924) is a precursor to the Hollywood film *Child’s Play* (1988). In *The Bloody Doll*, the brain of a recently executed criminal is transplanted into a machine (or puppet) and subsequently goes on a killing spree. S. Fowler Wright’s “Automata” (1929) is an early depiction of “the machine as the next stage in the evolutionary process” (Warrick 51), although concerning the loss of humanity at the hands of machines, does not portray explicit scenes of altruism (Wright). E. M. Forster’s *The Machine Stops* (1909) certainly deals with humanity’s transformation into a hive (as the opening sentence demonstrates), as well as exhibiting the augmentation of humanity and the disintegration of family relations when mediated by machines, yet it does not feature any automata. In “The Clericomotor,” published anonymously in the Detroit Free Press newspaper in 1884, a robot gives a sermon to a congregation, until the boy responsible for turning the crank starts to turn it backwards, causing the robot to malfunction, resulting in the congregation to flee in horror (“The Clericomotor”). The December 1899 issue of *The Black Cat* magazine featured the comical story by Elizabeth Whitfield Croom, entitled “Ely’s Automatic Housemaid,” about robots being employed for domestic housekeeping (Croom). Although humorously depicting a few mishaps, it ultimately seems optimistic about such socio-technological prospects. A follow-up story published in the same magazine in October 1900, entitled “Mr. Corndropper’s Hired Man” by W. M. Stannard, features a robot from the same inventor as Croom’s narrative, this time working on a farm (Stannard). Stannard’s depiction is similarly very enthusiastic about the prospects of robots working in the farmlands. Clare Winger

Harris' "The Artificial Man" (1929), is about a man who, after a series of accidents, replaces his limbs and organs with artificial prostheses and organs, effectively turning him into one of the earlier representations of a cyborg.

As discussed in Chapter 6, this dissertation begins its investigation with *Frankenstein* as the first major robotic science fiction narrative, and stops at *Metropolis* (1927). Kang similarly stops his study in the interwar period and admits it "might seem arbitrary, but there are number of good reasons for concluding the narrative proper here" (297). After World War II, as Kang explains, there was a "veritable explosion" (297) of science fiction robots: "The sheer bulk of the material that would have to be covered for a comprehensive study of the robot symbolism in the postwar era would require another full volume" (298). This dissertation certainly agrees but would also emphasize that the explosion of robots started to occur prior to World War II, with the inception of the so-called Golden Age of science fiction in 1930 (Warrick xvii; Blackford 27; MacArthur 13). Warrick begins her study at 1930, and adopts a very different approach to deal with the vast amount of narratives available. As Sian MacArthur similarly observes: "In the 1930s science fiction reached what is now widely regarded as its 'Golden Age' whereby the 'rules' and expectations of the genre and its many sub-genres became more readily identifiable and easier to classify" (13). In the case of robotic science fiction, one thinks of Asimov's three rules of robotics as the pinnacle of formalized humanoid robots (also discussed in the conclusion of this dissertation).[8] Since the 1930s, the sentimental robot meant that 'rules' became necessary with which to mediate human-machine

[8] While the 1920s gave rise to *R.U.R.*, *The Clockwork Man* and *Metropolis*, the 1930s became preoccupied with sentimental robots. John Wyndham's "The Lost Machine" (1932), Lester del Rey's "Helen O'Loy" (1938), and Eando Binder's "I, Robot" (1939), are the first robotic narratives that explore the emotional dimensions of their characters. In "The Lost Machine", the first robot narrative in first-person, is about a lonely robot that eventually commits suicide; "Helen O'Loy" is about a robot that marries and, after the death of the robot's husband, asks the engineer to be shut down; Binder's "I, Robot", also in first-person, is about a robot that is wrongfully accused of killing its maker and is persecuted by an angry mob. After reading *Frankenstein*, the robot understands why it is persecuted but decides to commit suicide as an act of altruism rather than fight and possibly injure several humans. Both "Helen O'Loy" and "I, Robot" were inspirations for Asimov's first robot narrative "Robbie" (1939). According to Asimov, his first robot narrative, although not directly referencing the three laws of robotics, nevertheless reveals that he had the three laws already in mind: "Asimov […] says that clearly he was already thinking about The Three Laws of Robotics since the story makes a rambling reference to the First Law" (Asimov "Robbie" 68-9).

relationships when machines acquired human emotional dimensions; as the robot character of Adam Link puts it: “Ironic, isn’t it, that I have the very feelings you are so sure I lack” (18). However interesting it would be to continue the approach outlined in this dissertation into this era of popular science fiction robots, such an analysis falls outside the scope of this dissertation. As it stands, this dissertation contributes to a large body of scholarship and discourses on both science fiction studies as well as various discourses on the nature of conceptual robots more generally and hopes to stimulate further research and discussion on the nature of, and relationships between, humans, robots and altruism.

Chapter 2: Critical Posthumanism and Robots

The term posthumanism denotes a variety of possible meanings. While transhumanism and posthumanism are often employed synonymously, denoting discourses that embrace human-technological enhancements, critical posthumanism denotes a discourse that is more skeptical of such endeavors. This dissertation, interested in critical posthumanism, distinguishes between transhumanism and posthumanism as oppositional discourses, such that posthumanism is synonymous with critical posthumanism. This chapter explores certain tenets of critical posthumanism that serve to establish a theoretical framework with which to approach and analyze the themes of altruism.

According to Blackford, transhumanists are optimistic about technological augmentations of the human condition, and hold that one "should *welcome* the use of technology to expand human capacities" (178). In so doing, transhumanists hope that human-technological integration will result in human "abilities raised far beyond the current human range", and at that point become posthuman: "The suffix 'trans' in 'transhumanism' refers to a desired *transition* to a posthuman level of competence" (178). The variety of posthumanism that this dissertation employs, also referred to as critical posthumanism, is not always diametrically opposed to transhumanism. Blackford defines critical posthumanism as "a set of philosophical positions that attempt to understand the world and guide our lives while repudiating the idea that human beings are ontologically exceptional" (178). Critical posthumanism traces such ideological assumptions regarding "ontological exceptionalism", or anthropocentrism, back to the liberal humanist subject.[9] While transhumanism is primarily concerned about exceeding the limitations of the human condition through the use of technology, critical posthumanism "is best understood as a critique of *humanism*" (Blackford 178). In other words, posthumanism engages with the legacies of liberal humanism as a means of averting the

[9] In the context of critical posthumanism, the liberal humanist subject denotes a conception of a universal human condition, derived from Enlightenment values, beginning with René Descartes. This particular conception of human subjectivity emphasizes the individual as a disembodied rational being that is substantively and/or ontologically distinct from nature.

dangerous combination of anthropocentrism coupled with transhumanist fantasies of self-empowerment.

While posthumanism and transhumanism are interested in the eventual realization of the posthuman figure, critical posthumanism considers it "lethal" when "grafting […] the posthuman onto a liberal humanist view of the self" (Hayles 286-7). In *Bodies of Tomorrow; Technology, Subjectivity, Science Fiction* (2007), Sheryl Vint shares a similar anxiety: "My two key concerns with the liberal humanist tradition as it persists in posthumanism are the emphasis on universality and the emphasis on individuality" (12). Universality implies the notion that a certain "human 'essence' [is] shared by all" (Vint 12), while individuality implies a self-ownership and isolation that "evacuates our model of society from any ethical sense of intersubjectivity and collectivity" (Vint 13). Instead, Vint focusses on the ways "we might be posthuman *and* embodied" [original emphasis], and argues that embodied posthumanism "has the power to expand our capacity for responsibility and our connections with others" (26). If the liberal humanist subject is a self-contained disembodied entity, Vint's embodied posthumanism advocates for a human subject that is more aware of its situatedness in the world and open to interconnectivities with others. Vint also states that "It is important to stress that in making this critique of liberal humanism, I am not ignoring or denying the many benefits that can be associated with humanism and liberalism" (13). This dissertation adopts a similar perspective, namely that critical posthumanism is not a complete dismissal of all liberal humanist values, but (as is employed in this dissertation) constitutes a careful reconsideration of liberal humanist values in the context of modern hyper-technological society.

According to Blackford, it is possible to "adopt a stance that is simultaneously posthumanist and transhumanist" (178). While this dissertation agrees with Blackford, such a perspective is not common among critical posthumanists. However, Blackford explains that an engagement with humanism's anthropocentrism is not necessarily "logically inconsistent with welcoming technological efforts to enhance [the human condition]" (178). In *The Posthuman Condition: Consciousness Beyond the Brain* (2003; originally 1995), Robert Pepperell adopts a similar

perspective of welcoming technological advancements without endorsing traditional liberal humanist conceptions of humans. Pepperell stresses contingency between "the mind, the body and the world" (20), and such a continuity signifies "the end of 'humanism'," which he defines as "that long-held belief in the infallibility of human power and the arrogant belief in our superiority and uniqueness" (171). Andy Miah in "Posthumanism: A Critical History" (2007) discusses Pepperell's posthumanism and states that

> [Pepperell] betrays the particular history of posthumanism [when] he appears comfortable to discuss posthumanism as if it were a temporal, progressive concept […] and […] about using technology to achieve even greater productivity and functionality. (5)

Indeed, Pepperell often "invites a posthumanism that involves embracing human enhancement" (Miah 6), which is connotative of transhumanist ambitions. However, Miah does point out that Pepperell does so "in a way that rejects traditional technological determinism" (6). Thus, while Pepperell may be more comfortable incorporating transhumanist ideas and attitudes, it would be unfair to categorize the oeuvre of his posthuman condition as a transhumanist fantasy. This dissertation views Pepperell's "betrayal of the particular history of posthumanism" merely as a betrayal in certain attitudes towards technology, but not as a serious deviation from critical posthumanist ambitions. His ideas about the fluidity of human subjectivity and embodiment are equally pertinent topics as discussed by other critical posthumanists: "the recognition that none of us are actually distinct from each other, or the world, will profoundly affect the way we treat each other, difference species and the environment. To harm anything is to harm oneself" (Pepperell 172).

N. Katherine Hayles' *How We Became Posthuman* (1999) is a well-received contribution to the field of critical posthumanism. Retracing historical developments in cybernetics that led to the loss of the body to information, Hayles emphasizes the importance of embodiment to contemporary and future developments of information technologies: "I see the deconstruction of the liberal humanist subject as an opportunity to put back into the picture the flesh that continues to be erased

in contemporary discussions about cybernetic subjects" (5). Hayles envisions two potential outcomes of realizing a culture inhabited by posthumans:

> my nightmare is a culture inhabited by posthumans who regard their bodies as fashion accessories […] my dream is a version of the posthuman that embraces the possibilities of information technologies without being seduced by fantasies of unlimited power and disembodied immortality. (5)

By retracing the early developments of cybernetics and revisiting important early proponents of cybernetics including Shannon, McCulloch, von Neumann and Wiener, Hayles convincingly demonstrates how "humans were to be seen primarily as information-processing entities who are *essentially* similar to intelligent machines [original emphasis]" (7). This dissertation describes such a perspective – positing humans as "*essentially* similar" to "information-processing entities" – as a cybernetic metaphor. As Hayles explains: "Cybernetics was born when nineteenth-century control theory joined with the nascent theory of information" which eventually "signaled that […] information, control and communication […] were now operating jointly to bring about an unprecedented synthesis of the organic and the mechanical" (8). Hayles also stresses the importance of embodiment as a means of countering the ever-increasing modes of disembodied interactions between humans and machines. Jaron Lanier (also quoted in the Introduction of this dissertation) describes this particular trend as a cultural ideology among technologists in Silicon Valley today: "The first tenet of this new culture is that all of reality, including humans, is one big information system" (27).

The first topic relevant for this dissertation, namely human-machine relations, can be extrapolated from the thoughts of Hayles and Pepperell. By human-machine relations, this dissertation does not mean interactions between humans and machines, as that pertains to the second topic of self-other relations discussed below. Rather, human-machine relations deal with (what is referred to in this dissertation as) cybernetic metaphors that conflate or equate the human condition to artificial conditions, or viewing a human as "essentially similar" to "information-

processing entities" as Hayles describes (quoted above). Although Miah's summary emphasizes the differences between Hayles and Pepperell, this dissertation emphasizes the similarities between Hayles and Pepperell's considerations in the context of human-machine relations. Both share the same basic concern of grafting the posthuman onto a liberal humanist subjectivity. While Pepperell may at times exhibit a transhumanist leaning, it should be noted that Hayles similarly is not antagonistic towards embracing information technologies either; instead, both oppose an embrace that is premised on beliefs of human exceptionalism. Thus, while they employ a monist paradigm that equates human and machine ontologies, they do so without re-inscribing the liberal humanist subject. This allows both to consider the possibility of creating artificial consciousness, given that human consciousness is the product of material conditions and there is no reason to assume that such conditions cannot be replicated artificially. Thus, while machines may be conscious one day, both thinkers deny the possibility of transferring consciousness from one medium to another. This leads Pepperell to make the claim that "Complex machines are an emerging form of life" (177), and not merely a technology. In relation to digital artificial life simulations, Hayles states: "When *form* is triumphant, Tierra's 'creatures' [i.e. digital organisms] are, in a disconcertingly literary sense, just as much *life-forms* as are any other organisms" (233). Although this dissertation is informed by these discussions, it will not elaborate upon speculations on machine consciousness or ontologies of artificial life-forms; Hayles and Pepperell already provided such considerations that would be difficult to improve upon. What is relevant in the context of this dissertation is that their works demonstrate the many ways in which human-machine ontologies can be deconstructed and reimagined. The liberal humanist subject becomes a posthumanist subject as the various distinctions between humans and machines collapse, destabilizing traditional conceptions of subjectivity, and thereby allowing for a posthumanist view to come more clearly into focus.

Neil Badmington's considerations serve as a good example of posthumanism's ability to examine "Otherness as it is manifested within culture" (Miah 7). In the context of this dissertation, Badmington's approach is exemplary when it comes to applying a critical posthumanist framework

to literary/cinematic analyses. As with Hayles and Pepperell, Badmington pays close attention to the Cartesian legacy of dualism, anthropocentrism and liberal humanism by examining its origins through a close-reading of Descartes ("Theorizing Posthumanism" 16-9). However, Badmington differs from Hayles and Pepperell when emphasizing the continuity between posthumanism and humanism itself.[10] Posthumanism must, according to Badmington, "take the form of a critical practice that occurs *inside* humanism, consisting not of the wake but the working-through of humanist discourse [original emphasis]" ("Theorizing Posthumanism" 22). In his book, entitled *Alien Chic; Posthumanism and the Other Within* (2004), Badmington applies his posthumanist framework to cinematic representations of extraterrestrial beings as forms of Otherness. He examines the evolution of cultural sentiments towards extraterrestrials from 1950 to 2000; initially, aliens were represented antagonistically and, eventually, they were represented more positively. However, Badmington is skeptical of this change in cultural attitudes and sentiments, and argues that such a newfound openness towards extraterrestrial otherness, or "alien love" as he calls it (4), is perhaps little more than "a reaction to the contemporary crisis in humanist discourse" (89). As such, the opposition between self and other persists regardless of negative or positive attitudes:

> Alien Chic, then, is a defence mechanism, a trend with which 'we' reassure 'ourselves' about who 'we' are at a moment of immense uncertainty. When 'our' difference from machines and animals is no longer obvious, 'we' turn to the alien for its instant difference ('I may be a cyborg, but at least I'm not one of *those*'). 'We' love 'Them', and loving 'Them' as a 'Them' confirms 'Us' as 'ourselves'. (90)

Badmington already connects his analysis of aliens to machines and animals in the above quotation, and this dissertation continues to expand on this particular insight in the context of science fiction robots. A readiness to embrace robots positively or antagonistically can still be considered a "defense mechanism" in the wake of "immense uncertainty" which serves to reaffirm traditional

[10] This is not to suggest that Hayles and Pepperell fail to notice such a continuity between humanism and posthumanism, but merely that Badmington puts more emphasis on this continuity.

anthropocentric values that regard humans as ontologically exceptional beings. Thus, a positive and enthusiastic representation of robots does not by default imply a subversion of liberal humanist subjectivity. These insights can be correlated with Donna Haraway's ideas as expressed in her well-known "Cyborg Manifesto" (1991).

Donna Haraway's cyborg is foundational to critical posthumanist discourses. The cyborg "advances the notion of a post-gender world where being a cyborg is preferable to being a goddess" (Miah 8), and does not constitute a blind obedience nor preference for cyborg ontology in a literal sense. As Richardson observes: "The cyborg is an anti-dualistic and anti-essentialist symbolic construct, in the sense that Haraway in her essay critically attacks patriarchy, colonialism, and capitalism, drawing the lines between these positions and social theorizing of lived realities" (9). However, some have found it alluring to interpret Haraway's cyborg as a literal solution to overly determined conceptual topologies. As Miah explains, Haraway has also expressed concern that "her ideas have been appropriated by a particular vein of posthumanism [or transhumanism] that expresses biological transgressions as a utopian break with evolution" (8). As a symbolic construct, the intention is not to "enhance humanity", but rather "to disrupt uniform ideas about what it means to be human and the social and political entitlements this might imply" (Miah 8). Hayles' discussion of Haraway's cyborg reiterates that the cyborg serves to "disrupt traditional categories", including "human/machine", "human-animal", and "animate/inanimate" distinctions (84), while Haraway provides many more examples of such oppositions (177). This dimension of the cyborg connects Haraway's ideas with the long history of automatons (as cyborgs, robots and androids) as constitutive of trans-categorical figures throughout various socio-historical contexts. As Kang explains:

> The use of the automaton idea, with which people have meditated for centuries on the fearful consequences of the collapse of such binary oppositions as well as the positive possibilities of transcendence of such strictures, has been one of the central themes of

> [Kang's] study. The cybernetic organism [including Haraway's cyborg] is just the most recent figure in that ongoing meditation in the Western imagination. (304)

It is this combination of "fearful consequences" and "positive possibilities of transcendence" that defines not only Haraway's cyborg but conceptual, cultural automatons more generally. In the context of critical posthumanism, "Haraway, Hayles and Pepperell each emphasise the dis-integration of the liberal humanist subject as the core characteristic of posthumanism" (Miah 8). The cyborg destabilizes uniform ideas pertaining to the liberal humanist subject in contemporary society which serves as a useful example for how historical automatons functioned similarly in various socio-historical contexts.

Rosi Braidotti also emphasizes the importance of empathy in relation to critical posthumanist discourses. In *The Posthuman* (2013), Braidotti identifies the "common denominator" of posthumanist discourse as a monist perspective that emphasizes "the vital, self-organizing and yet non-naturalistic structure of living matter itself" (2). Progressing from dualistic to monist perspectives means, according to Braidotti, that posthumanism should no longer rely on the "social constructivist approach" that "posits a categorical distinction between the given (nature) and the constructive (culture)" (3). Rather, as Braidotti points out, "the binary opposition between the given and the constructed […] is currently being replaced by a non-dualistic understanding of nature-culture interaction" (3). This "nature-culture continuum" is the "shared starting point" of Braidotti's posthumanism (2). While other critical posthumanists such as Cary Wolfe, in *What is Posthumanism?* (2010), examines the relationship between critical posthumanism and animal studies, Braidotti discusses the work of "post-anthropocentric neo-humanist" Frans de Waal, "who extends classical humanist values, like empathy and moral responsibility, to the upper primates" (77). This dissertation agrees with Braidotti that such an "emphasis on empathy accomplishes several significant goals in view of a posthuman theory of subjectivity" (78). Braidotti describes three such accomplishments: first, communication is reevaluated as an evolutionary tool; second, emotion rather than reason is identified as "the key to consciousness"; third, such a view serves to take

"critical distance from the tradition of social constructivism and situates moral values as innate qualities", which Braidotti claims is "a significant addition to the theory of the nature-culture continuum" (78). In addition, Braidotti explains that "Empathy as an innate and genetically transmitted moral tendency […] is in fashion, whereas selfish genes and greed are definitely out" (78). What is particularly relevant for this dissertation is the connection between empathy and posthuman subjectivity. Connecting Braidotti's claims with Pepperell and Hayles', one can say that machines may one day be as conscious as humans, and if moral values and empathy are innate qualities, machines might one day possess these qualities as well.

To summarize, critical posthumanism identifies several points of connection between humans and machines. Hayles, Pepperell and Haraway have successfully demonstrated the futility of attempting to define humans and machines in perfect oppositional terms; conversely, they also demonstrate the dangers of conflating these conditions as being identical. As Braidotti argues, empathy serves to accomplish the goals of a critical posthumanist view of the self. The theme of altruism from a critical posthumanist perspective offers ways of evaluating a posthuman theory of subjectivity without necessitating technological augmentations or other cybernetic metaphors that seek to reduce humans and machines to the same basic essentialities. However, as Badmington's study cautions, sentiments towards aliens (such as alien love) can be applied, in the context of this dissertation, to Turkle's robotic moment (discussed in the Introduction) as a desire not to overcome but to reformulate the self-other dichotomy. In both cases, aliens or robots, one can say a fear or love of robots serve the same end: loving one's robot is a transhumanist ambition that repositions a techno-humanist view of the self as opposed to fostering a critical posthumanist view of the self.

Chapter 3 below turns to Julien Offray de La Mettrie's thesis on humans and machines. Arguing that humans (and animals) are machines, La Mettrie is remembered as a major proponent of materialism. While his work is frequently mentioned in robotic discourses, what is ignored is his justification for morality despite the notion that humans are mere machines. In so doing, his arguments can be seen as a historical example of certain aims of critical posthumanism.

Chapter 3: Mechanical Morality

Julien Offray de La Mettrie's controversial 1747 treatise, entitled *l'Homme Machine* (or *Man a Machine*), warrants closer examination as a materialist treatise that takes up many of the same concerns expressed by critical posthumanists (as discussed in the previous chapter). Such concerns include the relationship between disembodied and embodied subjectivities, the relationships between humans, animals and machines, and the question of morality in a materialist universe. With these topics in mind, La Mettrie's thesis can be considered a precursor to contemporary critical posthumanism for its insistence on a monist perspective of human, animal and machine ontology as well as advocating innate morality within such a paradigm. However, before turning to La Mettrie, it is worth mentioning Jacques de Vaucanson and his automata which shortly preceded La Mettrie's *Man a Machine*.

According to Aram Vartanian, in *La Mettrie's l'Homme Machine: A Study in the Origins of an Idea* (1960), Vaucanson's automata "embodied the engineering equivalent of the man-machine theory and undoubtedly prepared the imagination of La Mettrie and his contemporaries" (68). Gaby Wood in *Edison's Eve* (2000) provides a detailed overview of Vaucanson and the creation of his automata. Wood explains that Vaucanson, according to his biographers André Doyon and Lucien Liaigre, was "an early cybernetician" and "his wildest and most secret ambitions were to remain in the realm of artificial life" (17). Fascinated by mechanics from an early age, Vaucanson became "a novice in the religious order of the Minimes in Lyon" as a means of pursuing "his scientific study" (Wood 18). In 1727, Vaucanson decided to "make some androids, which would serve dinner and clear the tables" during a visitation of one of the "heads of the Minimes" (Wood 19). While the visitor was impressed, he subsequently declared Vaucanson's machines to be "profane", and "ordered that his workshop be destroyed" (Wood 19). As a result, Vaucanson "pleaded with the Bishop to be withdrawn" and was free to continue his personal pursuits (Wood 19). In Paris in 1738, Vaucanson's Flute Player was exhibited for the first time, with an entry fee that was equivalent to a week's wages for a manual laborer (Wood 21). Within three months, the exhibition attracted

seventy-five visitors per day (Wood 21). In 1739, Vaucanson added two more automata, one was a "pipe-and-drum figure" and the other "a mechanical duck" (Wood 26). The duck became Vaucanson's most famous invention (Wood 27), given its ability to eat, digest and defecate. It was only later revealed that instead of a "chemical laboratory" that mimicked the digestive process, the mechanical duck was fitted with a compartment containing a "separate substance made to look like the digested version" of the food, near "the bird's rear end" (Wood 31-3). However, the impact of Vaucanson's automata was firmly established by that time.[11] It is no surprise that La Mettrie would label him a "new Prometheus" (Wood 17).

Trained as a physician, La Mettrie went to the Netherlands and studied under "the renowned Boerhaave, at […] Leyden" (Vartanian 2). After leaving the Netherlands, he returned again in 1746 after having "ridiculed many of the bigwigs of medicine in France" (Vartanian 6). He completed *Man a Machine* roughly in August in 1747, and copies began circulating in November or early December (Vartanian 6). The reception was "met with angry protests from all classes", while the publisher, Elie Luzac, was "summoned on December 18 before the Consistory of the Eglise Wallone de Leyde" and ordered to deliver all copies, reveal the author's identity and make apologies (Vartanian 7). The church also ordered all copies to be burned (Wood 12). Fortunately, Luzac also "surreptitiously placed in circulation enough copies […] to gratify the increasing curiosity of the reading public" in 1748 (Vartanian 7). La Mettrie was eventually revealed as the author and subsequently fled to Berlin in February in 1748 (Vartanian 7). Although he would later credit Descartes for his own ideas in the treatise, it is an oversimplification to assert that La Mettrie was merely continuing the Cartesian and Newtonian traditions of mechanical philosophy. Rather, Descartes postulated two substances (matter and mind), while Newton postulated that matter could still be animated by subtle spirits; La Mettrie's position is more controversial when stating that humans and animals are nothing more than machines.

[11] Despite such trickery, Vaucanson's mechanical ingenuity should not be underestimated. The wing of the duck alone contained more than four hundred moving parts (Wood 27).

According to Kang, "La Mettrie's materialist tract marks the end point of the classical mechanist movement that commenced in the 1630s" (131). As Kang explains, most classical mechanists were dualists, and restricted their "mechanistic descriptions to the body, excluding the soul" (131). La Mettrie's contribution to mechanical philosophy was, in some sense, almost a departure from traditional mechanical philosophies because he aimed to show that between soul and matter, the soul became redundant. As such, La Mettrie was far more controversial than Descartes when affirming that a human is nothing more than a machine (without the need for any kind of soul). In order to make his claim, La Mettrie had to overcome three shortcomings of mechanical philosophy: the problem of language, the problem of souls, and the problem of morality.

La Mettrie argued that language is a function of nature. His argument almost seems to anticipate the theory of evolution, as well as many insights pertaining to the nature of language, as he wrote:

> All was done by means of signs. Each species understood what it was fitted to understand. And thus have men acquired what our German philosophers still call *symbolic knowledge*. […] Everything is reduced to sounds or words which fly from the mouth of one through the ear of the other into the brain, which receives at the same time through the eyes the shapes of the bodies of which these words are the arbitrary signs. [original emphasis] (41)

The divide between human and animal is no longer premised on a physical-metaphysical duality but simply on a difference in evolution, and this leads La Mettrie to argue that one could hypothetically teach apes how to communicate with sign language.[12] In so doing, La Mettrie overcame the problem of language from a materialist perspective.

[12] One can consider contemporary successes such as teaching Koko the gorilla sign language. Although La Mettrie's conjectures were overly optimistic as to these possibilities (surmising that an ape might be as fluent as a deaf person), it is interesting that they are nevertheless possible. For more information on Koko, see: <http://www.koko.org/sign-language>

When it comes to the problem of souls in a materialist philosophy, La Mettrie observed a logical paradox in dualism. Badmington (discussed in Chapter 2), a critical posthumanist, also notices a logical paradox inherent in Descartes' formulations, when Descartes wrote:

> For whereas reason is a universal instrument which can be of use in all kinds of situations, these [bodily] organs need some particular disposition for each particular action; hence it is impossible to conceive that there would be enough of them in a machine to make it act in all the occurrences of life in the way in which our reason makes us act. (qtd. in Badmington "Theorizing Posthumansim" 18)

Descartes, it would seem, would have disagreed with posthumanists such as Hayles and Pepperell, ironically by pointing to human embodiment and suggesting that machines could never replicate or imitate such levels of complexity. However, Badmington responds that, according to Descartes' own formulations, the implication is that "Given enough organs, a machine would, after all, be capable of responding, and responding in a manner utterly indistinguishable from that of a human being. Reason, no longer that which 'distinguishes us from the beasts', would meet its match, its fatal and flawless double" (18). As Badmington's close-reading reveals, Descartes' negation of mechanical humans is not premised on ontological limitations but rather on technological limitations. If there is a machine with enough organs in the right places, the result would be identical to that of a human being. As a result, when dualism begins to collapse, the approximation of humans, animals and machines means that if one possesses a soul, the others must logically follow; if one is a machine, then all must be machines. This is a realization that La Mettrie certainly also had in mind.

His convictions were supported in his own time by Abraham Trembley's experiments with freshwater polyps: "long classed a plant rather than animal, [polyps] had the ability to regenerate itself when divided: it would, without intercourse, turn into as many polyps as there were parts" (Wood 13). La Mettrie describes this phenomenon as "reproduction, which takes place by division alone without coupling" (40). Another contemporary scientist of La Mettrie was Albrecht Haller, who demonstrated that "muscles move of their own accord" and "respond individually if directly

stimulated" (Wood 13). Creating new life through material division along with the body's ability to move autonomously constituted enough evidence for La Mettrie with which to assert that life is a property of material conditions. As a result, La Mettrie no longer considers any need for the postulation of a soul.

The third problem of morality that La Mettrie had to address is a more difficult claim given the idea that humans and animals are mere machines. If only matter exists, such an argument could easily be construed as a denial of morality, or imply that morality is a superfluous dimension to human existence. La Mettrie set out to argue that materialism does not equate to a state of amorality, for similar reasons that critical posthumanists emphasize (as discussed in Chapter 2). In order to make his argument, La Mettrie relies on what he refers to as the natural law. First, La Mettrie argued for an innate sense of morality:

> Criminals, the wicked […] those, in short, who do not feel the pangs of conscience, wretched tyrants unworthy of being born, let them try to wring cruel pleasures from their barbarous behavior. Willy nilly, during calm moments of reflection, their vengeful conscience rises up, testifies against them, and sentences them to nearly ceaseless self-laceration. He who torments other men is tormented by himself, and the ills he causes himself is fair measure of that he causes in others. On the other hand, there is so much pleasure in causing good and making others happy, so much contentment from being virtuous, kind, humane, tender, charitable, compassionate, and generous […] that I maintain that whoever is so unfortunate as not to be born virtuous is punished enough. (53)

With these words, La Mettrie claimed that morality is an inherent property to humans. His argument is not simply that there is "much contentment" from "making others happy," but there are underlying mechanisms, endowed from nature, as to why this would be the case. He refers to this principle of inherent morality as the natural law:

> Now, how shall we define the natural law? It is a feeling that teaches us what we must not do on the basis of what we would not like someone else to do to us. Dare I add to this

> common notion that it seems to me that this feeling is only a kind of dread or terror as beneficial to the species as to the individual? […] You see that the natural law is only an inner feeling of the imaginations, as are all other feelings, among which is thought. By consequence, the presence of the natural law obviously presupposes neither their education, revelation, nor legislator, unless you confound it with civil law in the ridiculous way theologians do. The weapons of fanaticism can destroy those who witness these truths, but never the truths themselves. (53-4)

For La Mettrie, the human body is "an immense clock" (69), which is governed by his natural law, and this law is as beneficial to the species as to the individual. Altruism is not an abstract metaphysical virtue, but an evolutionary imperative. As he reiterates in his conclusion, there exists the possibility (as critical posthumanists also argue) to be both a materialist and a moral being: "the convinced materialist, though his own vanity whispers in his ear that he is only a machine or an animal, will not mistreat his fellows […] following the natural law given to all animals, he does not want to do unto others what he would not want them to do unto him" (76).

In so doing, La Mettrie manages to disarm the threat of vulgar social Darwinism, or any kind of anti-humanist sentiments one might derive from such a worldview. La Mettrie discusses a thought experiment concerning artificial intelligence and life, and similarly comes to the same conclusions contemporary scholars do:

> On these grounds, the precious gift of the natural law certainly would not have been refused to animals. They offer obvious signs of their repentance and intelligence. So why would it be absurd to think that such beings, machines nearly as perfect as us, are, like us, made to think and to feel the natural law working in them? (49)

Within La Mettrie's considerations, one finds the same enumerations of embodied minds, human-animal equality, dualism's self-alienation, and an emphasis on the nature-culture continuum. In addition, he links altruism with evolution and extends this dimension of behavior to autonomous machines.

Aram Vartanian's study on La Mettrie, entitled *La Mettrie's l'Homme Machine: A Study in the Origins of an Idea* (1960), concludes by considering the relevance of La Mettrie's views to cybernetics. According to Vartanian, *Man a Machine* gained new popularity during the 1940s and 1950s as a result of cybernetics: "it is easily understandable that [La Mettrie's] man-machine idea would be raised to a new level of meaning […] with the construction of multipurpose digital computers, logical calculators, and a variety of complex self-regulatory devices embodying the servomechanism principle" (134). After mentioning various cybernetic developments and insights, including W. R. Ashby's "Homeostat", Shannon's maze-runners, and Grey Walter's "tortoise" and "CORA", Vartanian explains that in view of such devices, "it has become more difficult than ever for the vitalists to attach to the specific phenomena of life an objective idea that will transcend all mechanistic interpretation" (134-5). Vartanian explains that although cybernetics does not "prove that a man is literally a machine", it does constitute the "most convincing illustration" of La Mettrie's thesis (136). In the context of this dissertation, it is not so much the question of whether humans are machines that is interesting, but rather whether La Mettrie's arguments for innate morality and the early science fiction texts involving robots engaged with similar considerations. While applying an early approximation of a cybernetic metaphor – humans are essentially similar to machines – served to further the development of materialist philosophies, La Mettrie's thesis should also be remembered for its recognition of the possibility of morality. La Mettrie's natural law which he defines as "do unto others" anticipated Pepperell's morality (discussed in Chapter 2) in the context of critical posthumanism: "To harm anything is to harm oneself" (172).

The following chapter turns to Alan Turing's considerations on intelligent machines and his famous Turing Test. While Turing was not particularly concerned with altruism, some of his ideas regarding machine intelligence emphasize cooperation between humans and machines. Revisiting the Turing test also enables one to better examine what this dissertation refers to as "Turing test moments" in robotic narratives. Turing test moments are instances in which robotic characters become momentarily indistinguishable from human characters as a means to legitimize the

humanity of such robotic characters, and/or as a means through which to explore human-machine relations in more detail.

Chapter 4: Alan Turing's Test; More or Less without a Body

In his biography of Alan Turing, entitled *Turing: Pioneer of the Information Age* (2012), B. Jack Copeland notes that "Time Magazine listed [Turing] among the 20th century's hundred greatest minds, alongside the Wright brothers, Albert Einstein, DNA busters Crick and Watson, and the discoverer of penicillin, Alexander Fleming" (3). To better understand Turing's contributions to the general fields of mathematics and computer science, one should recognize that "There is a direct line from the universal Turing machine of 1936 […] and onwards to the first personal computer" (Copeland 143). Turing believed that machines would one day become intelligent, and predicted that by the year 2000 the first machines would pass the so-called Turing test (Christian 4). The Turing test is a heuristic with which to decide or assess whether a machine is intelligent. Simply put, if one talks to a machine without being able to distinguish whether it is artificial or human, then one must assume the machine to be intelligent. The Turing test has become an annual international competition known as the Loebner Prize, and any artificially intelligent software program than can fool 30% of the judges is considered to have passed. In 2014, the first artificially intelligent program, called *Eugene Goostman*, simulating a 13-year-old boy from Ukraine, fooled 33% of the judges into thinking it was human ("Computer AI Passes Turing Test").

After being prosecuted for "homosexuality", which was only "decriminalized" in 1967 in the UK (Copeland 194), Turing was subjected to therapy that involved the injection of female hormones for the duration of one year (Copeland 195). Jaron Lanier explains the underlying motivations as to why medical practitioners would believe that injecting female hormones could treat or cure male homosexuality. The answer, as Lanier explains, has to do with implicit human-technological metaphors. That is, we perceive our world analogously through the pervasive technology surrounding us; today, it is not uncommon to find various analogies comparing the mind to software, while the body is posited as the corresponding hardware. In the case of Turing, in the 1940s, "the steam engine was a preferred metaphor for understanding human nature" (Lanier 30). During this time, homosexuality was viewed as the result of having all "that sexual pressure […] building up and

causing the machine [or human] to malfunction, so the opposite essence, the female kind, ought to balance it out and reduce the pressure" (Lanier 30). In 2009, fifty-five years after his death, the British Government offered an official apology for the inhumane treatment Turing received. Lanier speculates that this ordeal might have influenced Turing's conception of the famous Turing test:

> It is impossible for us to know what role the torture Turing was enduring at the time played in this formulation of the test. But it is undeniable that one of the key figures in the defeat of fascism was destroyed, by our side […] because he was gay. No wonder his imagination pondered the rights of strange creatures […] I can imagine it might have been a comfort to imagine a form of life apart from the torments of the body and the politics of sexuality. (30-1)

As Lanier explains, Turing's ordeal should also serve as a reminder of the dangers of analogizing ourselves or others in relation to various technologies. While Lanier claims that it is impossible to know, he nevertheless surmises that machines may have provided some form of comfort from the sufferings of the politics of sexuality. Kathleen Richardson similarly writes: "Turing's otherness, his alterity and difference, was influential in his theorizing about machine thinking – his life must have seemed like an imitation game of sorts, with revelations and secrets, coding and decoding" (44). Richardson also quotes Paul Strathern's claim that "To regard himself as a machine provided a great psychological relief from the continuing turmoil of his inner life" (qtd. in Richardson 44). Such speculations seem to be supported by the realization that Turing's Imitation Game was based on the separation of genders prior to a separation of humans and machines.

The Turing test is derived from his well-known essay "Computing Machinery and Intelligence", originally published in 1950, and was first introduced as the "imitation game" (Turing 441). The game consists of three players, namely a man, a woman and a judge (which may be of either sex), and the judge must subsequently distinguish between their genders through interlocution alone (Richardson 43; Hayles xii-I; Turing 441). Hayles discusses the implications of overlaying gender differences with human-machine differences in more detail:

> In the paper itself, however, nowhere does Turing suggest that gender is meant as a counterexample; instead, he makes the two cases rhetorically parallel, indicating through symmetry, if nothing else, that the gender and the human/machine examples are meant to prove the same thing. (xiii)

When downplaying the significance of gender in the original formulation of the test, one fails to do justice to the various implications of the test. By including gender, the participants are already embroiled in a circuit of predefined parameters in which their gender is entirely predicated on language, as Hayles explains:

> It would also necessarily bring into question other characteristics of the liberal subject, for it made the crucial move of distinguishing between the enacted body, produced through the verbal and semiotic markers constituting it in an electronic environment. This construction necessarily makes the subject into a cyborg, for the enacted and represented bodies are brought into conjunction through the technology that connects them. (xiii)

Reading the role of gender in the Turing test through a posthumanist perspective, as Hayles provides, suggests that the test merely proves, or serves to reaffirm, one's own preconceptions: "Think of the Turing test as a magic trick […] the test relies on getting you to accept at an early stage assumptions that will determine how you interpret what you see later" (xiv). Richardson similarly mentions that flesh loses its meaning in the context of the Turing test. When Turing considers the possibility of making artificial human skin, he claims that "we should feel there was little point in trying to make a 'thinking machine' more human by dressing it up in such artificial flesh" (qtd. in Richardson 44; Turing 442). This statement, according to Richardson, "informs us that Turing did not see the mind as an embodied mind, but as a disembodied cognitive system" (44). Much of the allure of the test derives from the illusion of presenting the liberal subject in a distilled form, namely as a disembodied mind. However, peering through the illusion of the test, as Hayles cautions, the reality is not pure communication but rather the formation of a cyborg.

According to Copeland, the question of whether the mind is nothing more than a machine is something that Turing was "open minded" about (219). As far as the notion is concerned that the human mind is nothing more than an elaborate digital computer running on the body's hardware, Turing both agreed and disagreed. As Copeland explains:

> [Turing] did manage to prove on paper […] that a large enough network of his synthetic neurons will function as a universal Turing machine […] and he speculated that the human cortex is a 'universal machine or something like it'. So the soft machine [i.e. the human brain] might really be a computer. There is, though, an extra something to the human brain, Turing thought. If the brain were nothing more than a universal Turing machine, then once a person had executed the program of instructions stored in the brain's equivalent of the paper tape, he or she 'would sink into a comatose state or perhaps obey some standing order, such as eating'. He called this extra ingredient of human intelligence (whatever it is) 'initiative'. […] [As a result of his premature death] Turing offered no further explanation of the nature of initiative. (202-3)

Turing's concept of "initiative" can also be linked to one of his earlier contributions in the field of mathematics, called "intuition."

In 1937, Turing worked on a phenomenon that mathematicians call "intuition," which is defined by Turing as an activity which "consists in making spontaneous judgments which are not the result of conscious trains of reasoning" (Copeland 28). This means that "the more skilled the mathematician, the greater is his or her ability to apprehend truths by intuition" (Copeland 28). This view was contested, as some, like German mathematician David Hilbert, considered the realm of mathematics to be completely fixed, rigid and perfectly systematic (and certainly no place for something like intuition) (Copeland 28). However, Turing would help vindicate the problem of intuition, as Copeland explains:

> Thanks to Turing and [Kurt] Gödel […] it was now clear that Hilbert's dreamed-of foundation was a figment. No matter which systematic procedure is picked, there will always be

> statements that mathematicians can see intuitively to be true but which cannot be shown by using the rules. Intuition cannot be eliminated as Hilbert thought. (29)

If there is always room for intuition in the realm of mathematics, then a computer based on mathematics is ultimately a poor analogy of the human brain: there must be something extra-mathematical about the human brain. Turing merely recognized that contemporary models and modes of understanding were insufficient to account for human and/or artificial intelligence. Something was still missing from our conceptual equations.

Kurt Gödel postulated the possibility of constructing "a society of machines that are able to prove mathematical theorems", and unless "the society develops in ways that are essentially uncomputable, the whole race can be reduced to a single master-machine" (Copeland 219). Such a machine would be able to "perform the work of all the other machines belonging to the race" (Copeland 219). The idea of a universal master machine harked back to David Hilbert's idea that mathematics represented a perfect system. When Max Newman "once suggested that the whole of formal mathematics was indeed about finding proofs that could be produced by a single master-machine" (Copeland 219), Turing responded by calling Newman an "extreme Hilbertian" (qtd. in Copeland 219). However, while Turing did not believe in the notion of a single master machine, he did tell Newman that a race of machines might indeed be able to achieve the same result, as Copeland explains:

> Although no single machine can master the whole content of mathematics, there is no argument that the entire endless race of machines might not collectively be able to do so. Hilbert was wrong that mathematics is one giant machine, but that is no obstacle to regarding human mathematicians as being akin to Gödel's race of machines. (220)

Using collections of agents to solve problems has developed into a subfield in artificial intelligence research known as distributed artificial intelligence, which studies "how logically and physically distributed agents cooperate with each other to perform intelligent behaviors" (Shi 20). According to Zhongzhi Shi: "Studies of human intellectual behaviors show that most human activities involve

social groups consisting of multiple individuals, and large-scale complex problem solving also involves cooperation of several professionals or organizations" (20). Shi also explains that "'Cooperation' is a major aspect of human intelligence pervasive in the human society, and thus the motivation for research in Distributed Artificial Intelligence" (20). As discussed below, Turing's ideas were also inspired by human social behavior.

In his essay "Intelligent Machinery" (1948), under section 6 entitled "Man as a Machine", Turing provides some important considerations for the aims of this dissertation. Drawing on the similarities between humans and machines, Turing begins with the body and moves towards the brain when explaining that: "A great positive reason for believing in the possibility of making thinking machinery is the fact that it is possible to make machinery to imitate any small part of a man" (420). Examples include a microphone as a replacement for the human ear, a "television camera" for the eyes, and "remote controlled Robots whose limbs balance the body with the aid of servo-mechanisms" (420). In addition, for the machine to learn and discover things on its own, it must be free to roam where it pleases, which not only poses a serious "danger to the ordinary citizen", but also introduces another problem: "the creature would have no contact with food, sex, sport and many other things of interest to the human being" (420). This observation is particularly paradoxical, given that this "creature" is the result of mechanizing the human body, yet it fails to connect to those interests that specifically deal with human embodiment; it amalgamates the human and machine embodiment whilst simultaneously highlighting the impossibility of such an amalgamation. Although Turing admits that mechanizing the body in order to achieve a "thinking machine" might be successful, it is still considered too "slow and impractical" (420). Turing therefore offers a different solution: "to try and see what can be done with a 'brain' which is more or less without a body", and Turing subsequently restricts the activities of such a brain to five suitable branches of thought (420). The "great positive reason" for believing in the possibility of artificial intelligence actually proves to be the most difficult and cumbersome challenge: mechanizing embodiment. The easier alternative is to actually build intelligent machines that are restricted to

formal and rational exercises, such as playing games, learning and translating languages, cryptography and mathematics (420).

When considering the possibility of teaching or educating a machine, Turing points out that humans have a tremendous advantage: "It would be quite unfair to expect a machine straight from the factory to compete on equal terms with a university graduate" (421). The reason is because humans benefit from human-human interactions. Humans are therefore subject to "interference" from other humans (421). Turing still points out that "although a man when concentrating may behave like a machine without interference, his behavior when concentrating is largely determined by the way he has been conditioned by previous interference" (421). Later in the essay, Turing reiterates the same point by stating:

> the isolated man does not develop any intellectual power. It is necessary for him to be immersed in an environment with other men […] From this point of view the search for new techniques must be regarded as carried out by the human community as a whole, rather than by individuals. (431)

Turing's realization of the importance of communication and cooperation between members of a group is fundamental in the context of this dissertation. By analogizing human intelligences, and how humans achieve such intelligence, to machines (and machine learning), cooperation becomes key. In the case of distributed artificial intelligence as explained by Shi (mentioned above), researchers are still studying and employing these ideas today.

Susan G. Sterrett, in her essay entitled "Turing and the Integration of Human and Machine Intelligence" (2004), discusses these implications in further detail and provides the example of IBM's *Watson*, which won the gameshow *Jeopardy*, as a convincing demonstration of Turing's claims. Sterrett explains that the gameshow tests one's general knowledge, and "Watson was able to outperform humans in a question-answering task that was (seemingly) unrestricted with respect to topic" (14). However, Watson's designers admit that out of "3500 questions, all but 4.53% of the answers were Wikipedia titles" (qtd. in Sterrett 15). In other words, Watson was heavily dependent

on Wikipedia entries for its astounding performance. One should keep in mind that Wikipedia, as Sterrett explains, "is constructed by humans who enjoyed writing and sharing the information – and were joined by others who added to and revised it" (16). The implication is that "the case of Watson bears Turing out; the machines that are closest to developing intellectual power on parity with humans are those that are not only trained by humans, but that are in frequent communication with them" (17).

The annual Turing test competition, although lighthearted and good for innovation, might be exacerbating the very dimension Turing wanted to downplay, namely, the dimension of casting judgments onto other entities while forgetting to evaluate our own subjectivity in the process. As Turing pointed out: "With the same object therefore it is possible that one man would consider it as intelligent and another would not" (431). The question is not how one can ascertain the intelligence of another entity, but rather to emphasize the importance of interlocution and cooperation when it comes to the development of intelligence in the first place.

To conclude, whenever the human-robot metaphor occurs, whether in fiction, politics, philosophy or religion, so too does (usually) the misrepresentation of a systems theoretical perspective. The misrepresentation is one of humans and robots being essentially similar as opposed to behaviorally similar. Turing's insights help to remind us that the more useful focus should be on behavior rather than existential dimensions. For example, consider Ayn Rand's words in 1964, in her book *The Virtue of Selfishness*:

> […] try to imagine an immortal, indestructible robot, an entity which moves and acts, but which cannot be affected by anything, which cannot be changed in any respect, which cannot be damaged, injured or destroyed. Such an entity would not be able to have any values; it would have nothing to gain or lose […] Only a *living entity* can have goals or can originate them. And it is only a living organism that has the capacity for self-generated, goal-directed action. On the *physical* level, the functions of all living organisms, from the simplest to the most complex – from the nutritive function in the single cell of an amoeba to the

> blood circulation in the body of a man – are actions generated by the organism itself and directed to a single goal: the maintenance of the organism's *life*. An organism's life depends on two factors: the material or fuel which it needs from the outside, from its physical background, and the action of its own body, the action of using that fuel *properly*. (12-3)

For Rand, values cannot apply to such an indestructible robot, but only to living things; morality is thereby connected to living organisms despite subsequently reducing humans to being essentially similar to amoeba. In addition, both humans and amoeba are dependent on fuel and the ability to consume that fuel properly, much like a mechanical system. Here, in the context of arguing for the virtue of selfishness, using a metaphorical robot undermines the idea that values only apply to living systems when those systems begin to approximate mechanical systems being merely dependent on fuel and the consumption of fuel.

Turing would likely have advised Rand to imagine a race of robots instead of just a single indestructible one. Turing's conclusion in "Intelligent Machinery, A Heretical Theory" (1951) gives us a much more accurate understanding of what indestructible machines would portend. As Turing explains:

> Let us now assume, for the sake of argument, that these machines are a genuine possibility, and look at the consequences of constructing them. […] There would be no question of the machines dying, and they would be able to converse with each other to sharpen their wits. At some stage therefore we should have to expect the machines to take control, in the way that is mentioned in Samuel Butler's 'Erewhon'. ("Intelligent Machinery, A Heretical Theory" 475)

Both Rand and Turing's machines are indestructible, but Turing's machines exist as a group while Rand's robot is alone. The difference between individual and community is important, because Turing's machines are still capable of evolving, adapting and communicating to "sharpen their wits". Hence, Rand's emphasis is on existential dimensions, as she posits the notion that values are related to living organisms; Turing illustrates that existentialism is irrelevant, and that the important

difference is one of social behavior. Whether these machines are intelligent, conscious or indestructible life-forms is immaterial given the fact that they can converse with one another and adapt. The implication for robotic science fiction is that the archetypal robot should be employed as a metaphor of human plurality, social behavior and interaction, and not human totality or existentiality (despite the fact that these inevitably tend to overlap in fictional discourses).

The following chapter, Chapter 5, examines systems theoretical perspectives more closely in order to elucidate the inner workings of group dynamics and altruism. These discussions implicitly reiterate Turing's speculation about intelligent machines working together, sharpening their wits and replacing humans on the evolutionary scale. The underlying reason has nothing to do with embodiment and disembodiment, but group dynamics and altruism. Through mutual collaboration, individual members of a species can form a superorganism, in much the same way that a race of machines can achieve what a single master machine cannot. Indeed, Turing's race of machines might one day kill us with their kindness – kindness not towards us, but towards each other.

Chapter 5: How Robots Evolved Altruism

In evolutionary biology, the problem of altruism resulted in two opposing perspectives, namely kin and group selectionism.[13] As discussed below, kin selection holds that the evolution of altruism is contingent upon genetic relatedness between individuals. Group selection, on the other hand, maintains that natural selection can occur at a group level whereby altruism becomes an important factor determining the survivability of the group as a whole. The fundamental difference, then, is whether natural selection occurs at an individual or group level.

Edward O. Wilson, a biologist specialized in myrmecology (the study of ants), explains in *The Social Conquest of Earth* (2012) that he originally ascribed to a kin selectionist view, but later abandoned kin selection in favor of group selectionism. The idea of kin selection was originally introduced by J. B. S. Haldane in 1955, yet "the foundation of a full theory was laid out by […] William Hamilton in 1964" (Wilson, E. O. 167). Hamilton's formula became the "'$e=mc^2$' of sociobiology" and "was stated by Hamilton as an inequality, $rb > c$" (Wilson, E. O. 167). In this formula, *b* is the benefit gained by the recipient of the altruistic act, while *c* is the cost to the altruist, and *r* represents their genetic relatedness; for example, siblings share half of their genes, meaning *r* is equal to 0.5, and cousins share one-eighth of their genes, meaning *r* is equal to 0.125 (Wilson, E. O. 167-8). This relationship describes the limitations of the evolution of altruism, so that altruism will evolve, for example, if "the benefit to a brother or sister is 2 times the cost to the altruist" (Wilson, E. O. 168). If this threshold is not met, then the gene responsible for altruistic behavior will not be inherited and therefore not evolve. As Wilson explains, the explicative power of Hamilton's formula was that it "let credence to a superstructure of sociobiological theory based on the presumed key

[13] Note that such a claim is an oversimplification when reducing these debates into merely two opposing interpretations. The aim of this dissertation is not to provide a history of such debates, nor is it concerned with which particular definition (or model) of natural selection best describes the dynamics of evolutionary altruism. The aim is merely to use evolutionary altruism as a theoretical lens, connoting generosity, cooperation, kindness, self-sacrifice and other philanthropic ideas, and how such a theme was represented, positively or negatively, in early science fiction texts about robots. As such, a technical discussion, although fascinating, exceeds the scope of this dissertation. For more information on the topic of altruism, see Samir Okasha's *Evolution and the Levels of Selection* ([2006] 2008); Oren Harman's *The Price of Altruism* (2011); David Sloan Wilson's *Does Altruism Exist* (2015); Abigail Marsh's *The Fear Factor* (2017).

role of kinship" (168). Despite that Wilson and Hamilton defended this theory together in 1965 before the Royal Entomological Society of London (169), Wilson would later come to the realization that this theory is flawed:

> The misadventure of inclusive-fitness theory originated in the belief that a single abstract formulation, in this case the Hamilton inequality, has implications that can be unpacked layer by layer to account for social evolution in ever-growing detail. This belief can be refuted by both mathematical logic and empirical evidence. (182)

Richard Dawkins is a supporter of kin selection, and the debates between these two schools have also become known as the Wilson/Dawkins debate (discussed below).

Samir Okasha, in *Evolution and the Levels of Selection* ([2006] 2008), provides a philosophy of science exploring these topics in great detail. He discusses the Price equation, named after George Price, published in 1972 (Okasha 18).[14] Okasha explains that the power of the Price equation, which describes inheritance from one generation to the next, "lies in its generality […] it rests on no contingent biological assumptions, so always holds true" (19).[15] According to Okasha, the Price equation holds significance to the "levels-of-selection question" for several reasons, including its generality; that it "lends itself naturally to a description of *multi-level* selection, as Price himself realized [original emphasis]"; and that it has "historical significance" in "shaping the debate over group selection" (3). Okasha distinguishes between two kinds of multi-level selection, namely MLS 1 and MLS 2, and argues that both are relevant but at different "temporal stages of a transition" (9). Initially, MLS 1 is the relevant model while "collectives are loose aggregates of interacting particles",

[14] Oren Harman, in *The Price of Altruism: George Price and the Origins of Kindness* (2011), relates the history of Hamilton and George Price in great detail: "Whereas others, in their hunt to fathom goodness, pitted different levels of organization of life against one another—the gene conniving against the individual, the individual subverting the group, one group fighting doggedly against another—this lonely outsider [George Price] understood that they would all have to be part of a single equation. It was a dramatic flash—a penetration that would forever change our view of the evolution of life. Unknown, untrained, in a foreign country, dejected and alone, he had caught a glimpse of the great canvas of natural selection and seen its splendor and broadness. And, writing the elegant equation, he literally came off the street, anonymous, to present it to the world" (4-5).

[15] Some have pointed out certain problems regarding the Price equation, for example, Matthijs van Veelen: "Statements that are 'derived' with the help of the Price equation are […] in many cases not the answers they seem to be" (412). For more information, see van Veelen (2005).

and later, MLS 2 is the relevant model when "collectives are cohesive units" (9). In 2015, however, Okasha writes on his blog that "Recently, a number of scholars have argued that the opposition between kin and multi-level (or group) selection is misconceived, on the grounds that the two are actually equivalent" (Okasha "Kin Selection"). Okasha explains that, according to these scholars, kin and group selection are "alternative mathematical frameworks for describing a single evolutionary process, so the choice between them is one of convention [and] not empirical fact" (Okasha "Kin Selection"). Not only does this imply that both Wilson and Dawkins are wrong, but that the controversy "can in part be attributed" to the difference between statistics and causality; this means that there is a "mismatch between the scientific explanations that evolutionary biologists want to give, which are causal, and the formalisms they use to describe evolution, which are statistical" (Okasha "Kin Selection").

Aylet Shavit similarly explains that much of the controversy and debate relates to the overall generality and ambiguity of these terms. Speaking about the difference in perspectives between Wilson and Dawkins, Shavit explains:

> the broad definitions used by Dawkins and Wilson are more likely to talk past each other without resolution. Nonetheless, the use of broad concepts seem to be dominating the field, perhaps partly due to the political images and memories that everyday terms such as 'altruism', 'group', and of course 'selection' carry into science from society at large. (Shavit)

As mentioned, Charles Darwin was very aware of this particular problem of altruism from an evolutionary perspective and addressed it by stating, in *Descent of Man* (1871), that:

> […] a standard of morality gives but a slight or no advantage to each individual man and his children over other men of the same tribe, yet that an increase in the number of well-endowed men and an advancement in the standard of morality will certainly give an immense advantage to one tribe over another. A tribe including many members who, from possessing a high degree of the spirit of patriotism, fidelity, obedience, courage, and sympathy were always ready to aid one another, and to sacrifice themselves for the

common good, would be victorious over most other tribes, and this would be natural selection. (166)

Wilson reiterates the same idea: "In colonies composed of authentically cooperating individuals, as in human societies, and not just robotic extension of the mother's genome, as in eusocial insects, selection among genetically diverse individual members promotes selfish behavior" (Wilson E. O. 162); in other words, selfish individuals generally survive altruistic individuals. However, the inverse is also true for Wilson as it was for Darwin: "On the other hand, selection between groups of humans typically promotes altruism among members of the colony. […] but colonies of cheaters lose to colonies of cooperators" (Wilson 162-3). In other words, altruistic groups survive selfish groups.

A detailed examination of the scientific debates and discourses of altruism goes beyond the scope of this dissertation. When dealing with altruism as a literary theme, the word altruism as employed in this dissertation necessarily denotes a wider range of meanings, including empathy, kindness as well as selflessness. All that is required for a literary reading of science fiction is the understanding that "Individual-versus-group selection results in a mix of altruism and selfishness, of virtue and sin, among the members of a society" (Wilson, E. O. 163).[16] As discussed below, nineteenth- and early twentieth-century scholars were similarly intrigued by this paradox.

In her historical account on the cultural histories of ants, entitled *Ant (Animal)* (2004), Charlotte Sleigh states that in 1851 the "Society for the Promotion of Christian Knowledge (SPCK) published two volumes concerning natural history and animal morals" (67). The second volume "was devoted to insects, with ants forming a major part of it" (67). Long before ants and bees captured the interests of systems theorists and artificial life researchers, Sleigh points out that "Conservative natural theologians of the nineteenth century saw the order of the ants' nest as God's way of

[16] It would be interesting for future research to examine these evolutionary theories in more detail, and correlate them to more contemporary works of science fiction to see if there is a more direct engagement with such theories. One can speculate, as this dissertation does, that early science fiction authors may have come across the problem of altruism as expressed by Darwin and others and incorporated them as themes, metaphors or intuitive conceptions of humanity/society as a species/group; however, it would not be logical to assume that early science fiction writers would have had any notion of these evolutionary dynamics to the extent that they have been debated since the nineteen sixties.

teaching a person his or her place within society" (73). One example pertains to the labor force: "The worker, clearly, should not aspire to anything greater, or easier, but should modestly fulfil his or her allotted role" (Sleigh 73). George Cruikshank's copper plate etching entitled *The British Bee Hive* (1867) exemplified Sleigh's observation of the development of using insects as models for social organization (Mellby; Sleigh 68). As Sleigh explains, Victorians were also impressed by the levels of altruism exhibited by these insects:

> Another class-related feature of ant life discussed by Victorian moralists was their [the ants'] kindness to one another or 'mutual aid'. Ants were observed to feed one another, to groom one another, to tend to their wounded companions, and even, according to some, to bury or mourn their dead […] In just the same way philanthropists encouraged members of the nineteenth-century working class to set up Mutual or Friendly Societies for their savings, capital loans, and for their support in old-age. Charles Darwin himself engineered such an organization for the poor in his Kent village. All this behavior could be read into the behavior of ants and bees; it was recommended [to both religionists and secularists] by nature itself. (74-5)

Thus, several metaphors conflating human society as insect hives became prominent during this time, while altruism became a measurement of how well evolved the organism (or hive) is at that particular point. What is also interesting is that the model of ants were seen as natural examples, equally relevant for both secularists and religionists. As Sleigh explains:

> If animals, as Descartes claimed, were effectively automata, then ants were one of the more intriguing models of nature's machines. Curiously, atheist mechanists and theists produced virtually indistinguishable accounts of ant behavior. Whether or not they were designed by God, these little creatures provided examples of incredibly precise actions. Whether or not God was pulling the strings of these tiny marionettes, no human artificer could replicate the results. (143)

Much in the way that La Mettrie argued for a materialist morality, ants seen as machines were viewed (from a secularist perspective) as evidence of a materialist altruism. From this perspective, the distinctions between soul and soulless, mechanical and organic, natural and artificial, become less interesting, while social organization and altruism become more interesting.

Diane M. Rodgers' insightful study, entitled *Debugging the Link Between Social Theory and Social Insects* (2008), confirms the perspective put forward here. She explains that "The organicism that informed many biological metaphors in the nineteenth and early twentieth century has experienced a resurgence in current analogies involving the intersection of computer systems, human groups, and insect societies" (3). In addition, she points out the tendency to anthropomorphize these insects as exemplary models of human social organization. While this may be a "convenient way of describing behavior […] it also creates sociobiological analogies that then contribute to the naturalization of social theories and concepts" (9). The result is that such "anthropomorphism shapes a particular view of nature through a social lens and naturalizes social structure through a reinforcing loop" (9). As Rodgers explains, these metaphors were ingrained in the nomenclature of studying ants, some which are still used today:

> [employing descriptions such as] marriage, armies and army maneuvers, slave raids, robbery, and altruism. These activities are described as taking place in cities, villages, factories, the royal chamber, and bivouacs. The insect colonies are perceived to be overseen by all types of governments (socialistic, democratic, republican, anarchic, and nepotistic). This elaborate anthropomorphism makes eusocial insect colonies especially conducive to being utilized as analogues for human society. (11)

Thus, when applying such anthropomorphisms to human social organization, one is continuing the legacy of Victorian, secular and religious, thought. According to Rodgers, "Eusocial insect colonies have been termed a 'superorganism' […] from Herbert Spencer, who applied it to human societies but also felt that it could be applied to describe a universal condition for any complex organism" (11). The implication, eventually, is that during the nineteenth century the concept of hierarchy

"became a part of the concept of evolution," and "proved useful in justifying the colonization of other people" (Rodgers 17). These worldviews were popular enough that H. G. Wells wrote his short story "The Empire of Ants" (1905) about a new intelligent ant species threatening to take over the world: "These are intelligent ants. Just think what that means!" (648). These intelligent ants from the Amazon have evolved the capacity to use weapons and record history, and the protagonist concludes forebodingly that the ants will spread: "I fix [by] 1950 or '60 at the latest for the discovery of Europe" (648), by which he means the colonization of Europe.[17]

These discussions and examples should at the very least give pause for concern whenever human-machine metaphors are employed. Rodgers similarly observes that:

> As self-organizing systems deemphasize the individual, the computer simulation programs and artificial intelligence that utilize these self-organizing models are also viewed as neutral sources of evidence about natural and social systems. However, computer programs and AI are constructed by humans and subject to the cultural biases of these humans. (186-7)

In *Complex Adaptive Systems: An Introduction to Computational Models of Social Life* (2007), John H. Miller and Scott E. Page similarly discuss eusocial insects as examples of complex adaptive systems. Miller and Page define a complex adaptive system as including anything which "maintain[s] a recognizable form and activity over long periods of time, even though their constituent parts exist on time scales that are orders of magnitude less long lived" (7).

One should not confuse the words "complicated" and "complex" in these contexts: "In a complicated world, the various elements that make up the system maintain a degree of independence from one another" (9). Complexity, on the other hand, "arises when the

[17] The film *Phase IV* (1974) also depicts super intelligent ants colonizing the globe and incorporating humans into their world. The character of Dr. Ernest D. Hubbs explains what humanity is up against when saying: "They're not individuals. They're individual cells; tiny, functioning parts of the whole. Think of the society, James, with perfect harmony, perfect altruism and self-sacrifice, perfect division of labor, organized for preordained roles. Think of the building of elaborate and complex structures, according to plans that they know nothing of, and yet execute perfectly. Think of their ability to evolve and adapt in ways that are… so beautiful… and still so unknown. And all contained in one simple form. So defenseless in the individual… so powerful in the mass".

dependencies among the elements become important" (9). This means that complexity is "a deep property of a system, whereas complication is not. A complex system dies when an element is removed, but complicated ones continue to live on" (9). For example, a single human does not make humanity a complex system, but merely a more complicated system. If one human dies, humanity continutes to live on. As the biblical account of the Tower of Babel (Genesis 11:1-9) demonstrates, communication makes humanity a complex system. If one removes humans' ability to communicate, social organization collapses and the system of humanity ceases to exist. This means that communication is a "deep property" of the human system, much like Turing described with his race of machines being able to converse with one another and thereby sharpen their wits (discussed in Chapter 4).

Genetic diversity is important, not necessarily to the individual bee, but rather for the stability of the hive. As explained by Miller and Page: "When the hive gets too cold, bees huddle together, buzz their wings, and heat it up. When the hive gets too hot, bees spread out, fan their wings, and cool things down" (15). Each individual bee's "temperature threshold for huddling and fanning are tied to a genetically linked trait," meaning that genetically similar bees will experience hot and cold at the same temperatures. Hives "that lack genetic diversity in this trait experience unusually large fluctuations in internal temperatures" (15). This is due to the fact that genetically similar bees react to temperature deviations simultaneously, resulting in an all-or-nothing response. Conversely, with greater genetic diversity one finds that, when the temperature deviates, only a select percentage of bees (say five or ten percent) will notice the change in temperature and subsequently attempt to rectify it. Should that small percentage be insufficient, the temperature will continue to deviate, which will cause more bees to steadily react as the temperature exceeds their individual thresholds. Because of a greater genetic diversity, the overall response is more proportional to the deviation in temperature. As Miller and Page explain, it is "very misleading" to consider the average behavior of bees, seeing as though "average behavior leads to wide temperature fluctuations whereas heterogeneous behavior leads to stability" (15). In other words, it

is difficult to speak of a complex system by merely examining individual agents as the system's overall behavior might be very different from one agent's personal properties. This is not to argue that diversity is always and everywhere a good thing. Indeed, all systems must establish their own balance between homogeneous and heterogeneous elements in order to maintain stability. This equilibrium is typically referred to as the edge of chaos: "This balance point—often called the edge of chaos—is where the components of a system never quite lock into place, and yet never quite dissolve into turbulence, either" (Waldrop 12).

When it comes to complex behavior and the human mind, D. S. Wilson discusses, in his book *Does Altruism Exist* (2015), a particular experiment involving rhesus monkeys. The monkeys are trained to follow dots moving across a screen, while scientists monitor their brain activity. The brain is divided into what Wilson describes as "factions," where some "neurons fire at the sight of right-moving dots and others fire at the sight of left moving dots. Their firing rates are unequal because more dots are moving in one direction than the other" (Wilson D. S. 14). The monkey subsequently turns its head towards the faction of neurons that is strongest, implying that its decision is made on the basis of competing groups of neurons. As Wilson points out: "Remarkably, the interactions among the neurons resemble the interactions among the bees" (14). As Turing foresaw, the brain is a computer but, as he suspected, not a simple universal Turing machine nor master machine. As D. S. Wilson concludes: "Once we regard an individual organism as a population of socially interacting cells, the possibility of a group mind, comparable to an individual mind, becomes less strange" (Wilson D. S. 14). Here, Hayles also derives the same conclusion from a critical posthumanist perspective, described as the plurality of individual consciousness:

> Speaking for myself, I now find myself saying things like, 'Well, my sleep agent [or faction of agents] wants to rest, but my food agent says I should go to the store'. Each person who thinks this way begins to envision herself or himself as a posthuman collectivity, an 'I' transformed into the 'we' of autonomous agents operating together to make a self. (6)

While Hayles emphasizes the single mind as a collection of agents, Wilson expands the collection of individuals into a larger social mind.

To conclude, whenever this dissertation employs the term altruism, it is employed to mean a wide range of benevolent sentiments and behaviors. Although technical scientific definitions are interesting, they also exceed the scope and intention of this dissertation. It is sufficient to realize that conceptual links between individuals, groups, altruism and selfishness existed during the period of early science fiction robotic texts, and therefore is not unjustified to speculate that such early robotic texts may have incorporated these ideas as thematic concerns. When reading early science fiction robotic narratives, this dissertation is particularly interested in moments where group identifications, social isolation, altruism and selfishness become particularly important. Furthermore, this dissertation defines the term 'group' as any collection of characters that share a particular interest, motive or affiliation that binds them together; should such a group also happen to possess a clear complex property, then the group may be described as a complex adaptive system. The theme of altruism as explored in this dissertation refers to altruistic actions positioned on the cusp of individual and social interests; one sacrifices one's ability to survive when competing against individuals when acting altruistically, but as a group, such generous forms of behavior benefits the survival of the group while diminishing the importance of the individual.

Conclusion to Part I

The discussion in Chapter 1 highlighted several key concepts in studying fictional robots in the context of science fiction. What this dissertation refers to as robotic characters have a longer history than that of science fiction, and such a history reveals that such conceptual robots engage with unstable boundaries between binary oppositions. Chapter 2 discussed critical posthumanist considerations relevant to this dissertation, including embodiment and disembodiment; the self-alienation of a liberal humanist subject, human-machine and human-animal relations, and anthropocentrism as human ontological exceptionalism. It also highlighted that critical posthumanism does not deny the importance of empathy in relation to these discourses. As discussed in Chapter 3, La Mettrie's human-machine thesis had to address the problem of morality in a materialist universe. The theme of altruism – or the need to advocate for morality in a materialist philosophy – reveals many of the same considerations that critical posthumanists are reiterating contemporarily. While these considerations make a case for empathy and morality in the context of materialist philosophies and critical posthumanism, Chapter 4 turned to Alan Turing's insights on the nature of intelligent machines and the importance of cooperation. The two ideas presented in this chapter important for literary analyses of robotic characters are (a) the self-reflexive and self-critical dimensions of the Turing test, and (b) the connections between cooperation, machines and intelligence. While Turing posited the idea of looking at brains more or less without bodies, he nevertheless did recognize the importance of interactions between humans, as well as between humans and machines. Chapter 5 revealed that discourses by evolutionary biologists on the nature of altruism have been similarly preoccupied by questions of individual and group selection. Although the aim is not to evaluate the veracity of competing scientific discourses, these debates offer a useful model of the dynamics of altruism which Darwin also reiterated. This model of the dynamics of altruism serves as a practical framework through which to analyze early robotic science fiction narratives and the theme of altruism.

This overview provides sufficient information with which to formulate a number of theoretical assumptions and perspectives pertaining to fictional robots. The overarching impression is that technology and altruism are intricately linked in relation to human evolution. Additionally, fictional robots can serve as archetypes that compensate for the lack of interhuman relations particularly since the Industrial Revolution. The proliferation of technology does not merely imply personal empowerment, but a fundamental restructuring of society. In his book *Science Fiction* (2005), Luckhurst explains that the term mechanism "should not be understood as limited to the machine", but implies a much wider and more profound impact:

> When Thomas Carlyle wrote in 1829 about the onset of 'the Mechanical Age', he began a hugely influential discourse in which Mechanism was felt to pervade not only 'the external and physical … but the internal and spiritual also'. (3)

Not only did machines invade internal spaces, but it effectively defined the nature of society in mechanical terms. As Luckhurst explains: "In the 1880s, urban life was *itself* a machine ensemble, with everyday communications, public spaces and popular culture increasingly routed through machines" (29). The theme of altruism helps to understand the relationship between individual and society and the interplay between how the mechanization of one serves to define the other. Whether the human becomes an automaton, or society a machine ensemble, the loss of embodiment occurs concurrently with the concretization of a natural/artificial hierarchy in which altruism between humans is ultimately redirected as servitude towards a cosmic machine. The robot is often celebrated as a symbol of human intellectual prowess and technical ability, a means of overcoming death and suffering, but can also reveal implicit metaphors and conceptions of how humans define themselves in relation to one another. The themes of group selection and altruism become more prevalent as time passed by in the selected texts discussed below. Following the chronology of these fictional robots, the development of such themes from 1818 to 1927, ultimately culminate in and return to the realization of the theme of altruism.

Part II: Literary Analyses

Introduction to Part II

The following readings of five texts in this Part will demonstrate that altruism was a prevalent theme in early robotic science fiction, which may help us to appreciate the dramatic change in the treatment of this theme in pre- and post-World War I narratives. Although interwar robotic science fiction narratives placed a particular emphasis on the theme of altruism, its roots go back as far as the early nineteenth century.

Chapter 6 shows how the theme of altruism is presented in Mary Shelley's *Frankenstein* (1818), a text which introduces the artificial being in conjunction with altruistic concerns. Victor eventually becomes an altruist, if only for an instant, when he must choose between his own personal benefit or the benefit of humanity. Chapter 7 examines Edward S. Ellis' *The Steam Man of the Prairies* (1868) which demonstrates a similar preoccupation with the theme of altruism, but also exemplifies (from a modern perspective) the darker side of the dynamics of evolutionary altruism. Chapter 8 looks at Auguste Villiers de l'Isle-Adam's *Tomorrow's Eve* (1886) which demonstrates a problematic and peculiar engagement with altruism, and emphasizes the importance of human-human interactions and importance of sincerity. Chapter 9 examines Karel Čapek's *R.U.R. (Rossum's Universal Robots)* (1921) which reveals a much more direct engagement with the theme of altruism, represented quite clearly by incorporating a direct human-vs-machine scenario, in which humans become extinct and replaced by robots. The only survivors are robots capable of love. Lastly, Chapter 10 turns to Thea von Harbou's novel *Metropolis* (1926), *not* the eponymous film, which explicitly engages with the theme of altruism on multiple levels, not only in its representation through overt scenes of altruistic actions between characters, but also in showing the development of one of the protagonists from a self-interested industrialist towards a philanthropist.

Chapter 6: Mary Shelley's *Frankenstein*

This chapter revisits Mary Shelley's *Frankenstein* (1818) which has received much attention both within and outside science fiction discourses.[18] However, some of the specifically robotic nuances of her text may have been overlooked, given that her text is polemical and comprehensive in its treatments of both science and gothic fiction. This chapter examines *Frankenstein*'s treatment of a Turing test moment as well as the theme of altruism. The creature, being the first of its kind, like any robot, constitutes a binary opposition to humanity, and eventually orbits problems relating to intergroup competition. *Frankenstein* is not about an automaton in the strict sense, but the novel deals explicitly with the creation of an artificial humanoid, while it also in relation to this artificial creation engages with themes of group selection and altruism.

According to Kang, the novel "is commonly considered the first work of science fiction" (218) because of Percy Shelley's preface which distinguishes *Frankenstein* from conventional Gothic narratives that incorporate supernatural elements.[19] Sian MacArthur, while also identifying *Frankenstein* in *Gothic Science Fiction* (2015) as the "[…] earliest example of a science fiction narrative" (1), emphasizes its role as a subgenre to the Gothic tradition: "Shelley is moving away from the realms of traditional Gothic and into something new, and that is the beginnings of Gothic science fiction, a sub-genre of the Gothic" (2). This swerve from the traditional Gothic is strongly signified by the eight-foot tall creature as an artificial being that begins to move away from traditional alchemical notions of homunculi, and one that approaches more modernized representations of robots. Simultaneously Gothic and science fiction, the force of the text's

[18] A Routledge sourcebook, entitled *Mary Shelley's Frankenstein; A Sourcebook* (2002), contains excerpts from numerous articles classified under different section headings from "Body, Medicine and Science", "Commodity Culture and Social Structure", "Gender and Queer Theories", "Genre, Literary Form and Literary History", "Language and Psyche" and "Race, Colonialism and Orientalism".

[19] Isaac Asimov agreed with Brian W. Aldiss that *Frankenstein* constituted "the first genuine science fiction novel" (Blackford 9). Despite many earlier examples of science fiction narratives, Blackford similarly states: "This seems, I submit, as plausible a starting point as any" (9). However, one should keep in mind, as Blackford also states, that "no definition of the genre commands universal scholarly assent" (9). Cp. Paul Alkon's observation: "Others agree with Darko Suvin's suggestion that *Gulliver's Travels*, not *Frankenstein*, is science fiction's very first archetype" (12).

presentation of the creature is derived from its chimeric constitution of various ambivalences and anxieties, its ability to enforce and subvert rational order, the progress of technology and the subjective identity formations. Indeed, Robert Miles's response to the question of 'what is Gothic' offers a perspective of fragmentary subjectivity in the case of the creature: "My short answer is that the Gothic is a discursive site, a carnivalesque mode for representations of the fragmented subject" (4). As such, the novel also has strong sociological dimensions which Kakoudaki explores in more detail. Discussing the creature's artificial birth, Kakoudaki argues that such a trope has the effect of comingling the ontological and social dimensions of the creature's subjectivity: "the monster cannot be complacent about the fact of life and the fact of social rejection, as the two mysteries, one ontological and the other social or political, are irrevocably intertwined for him" (39).

The aim of this chapter is not to uncover the origins of science fiction as a genre, but rather to establish a literary tradition between *Frankenstein,* literary robots, the social environment and the theme of altruism. This dissertation sees Shelley's novel as the first modern robotic science fiction narrative, as it represents the moment when two separate discourses – science fiction and robots as cultural objects – came together in a single narrative. In the words of Michael Szollosy, in "Freud, Frankenstein and Our Fear of Robots" (2017): "Frankenstein and his monster emerge, if not at the very beginning, at a specific point very early in our modern (mis)understanding of robots" (434). Szollosy explains that the novel established a familiar trope: "Frankenstein, like Faust, is a victim of *hubris* and demonstrates that human endeavor, science and technology, whatever their noble intentions, inevitably create a monster that will gain autonomy and return to haunt us" (434). This trope, according to Szollosy, set the standard for subsequent works of robotic fiction: "We see this fundamental archetype time and again in our fantasies of monstrous robots" (434). Discussing *Frankenstein*, Warrick makes a similar observation: "These issues appear again and again in modern SF about robots and computer" (38-9). Arguably the most famous robotic science fiction author, Isaac Asimov, is remembered for his Three Laws of Robotics (discussed below), but he also coined the phrase "the Frankenstein complex", which he used as a means with which to critique a particular

trope or cliché in robotic science fiction (Asimov "Introduction" 5). In this way Asimov definitively sealed the connection between robotic science fiction and Shelley's novel.

This chapter is divided into three subsections. The first examines the relationship between Asimov's Frankenstein complex and the broader cultural myth of *Frankenstein* in the popular imagination which serves to illustrate its longstanding influence. The second subsection turns to the role of alchemy in Shelley's text in order to examine how the creature is manufactured. The last subsection turns to the theme of altruism as it is adopted and treated in the novel.

The Frankenstein Complex and Boris Karloff's Monster

Asimov defines his Frankenstein complex as a cliché in robotic fiction that deals with (or is derived from) technophobia. According to Asimov, writing in 1984,[20] technophobia operates "against change and technological advantage generally, [and] operate[s] against robots in particular" ("Introduction" 4). Asimov explains that because robots are "usually visualized as at least vaguely human in shape" or "pseudo human beings", their creation "by a human inventor is therefore perceived as an imitation of the creation of humanity by God" ("Introduction" 4). Given that robots function as pseudo-human beings, from religious perspectives, such forms of creation inevitably come to be seen as blasphemous (Asimov "Introduction" 5). As a result, Asimov argues, a particular cliché developed, namely a didactic moral that "there are some things man was not meant to know", a view rejected by Asimov: "as though it were perfectly all right for human beings to learn a thousand ways of ending life through every gradation of pain, misery, and unspeakable humiliation, but wrong and sinful to learn even one new way of creating life" ("Introduction" 5). The first work of fiction to impart this moral, according to Asimov, is Mary Shelley's *Frankenstein*: "Victor Frankenstein creates the Monster, who turns on Frankenstein and those he loves, and kills them. […] The success of *Frankenstein* was such that the basic plot of 'man creates robot; robot kills man' was repeated over

[20] Originally stated in 1964 in an introduction to a collection of his works, entitled *The Rest of the Robots* (1964). His arguments were repeated and elaborated in another introduction to a 1984 collection of robotic short stories, entitled *Machines That Think* (1984), also published under the title *War with the Robots* (1984).

and over again in uncounted numbers of science fiction stories" ("Introduction" 5). Asimov takes credit for personally helping to destroy that cliché through his own fiction: "It became one of the more unbearable clichés in the field (one that I successfully fought and destroyed, I am proud to say, with the establishment of my 'Three Laws of Robotics'" ("Introduction" 5). Asimov is not claiming that his robotic science fiction is somehow superior to Shelley's novel, but merely that his Three Laws helped to innovate and nudge the genre of robotic science fiction in new directions. In his view, this particular trope/cliché "has helped exacerbate this particular variety of technophobia, the fear of technological advance as 'blasphemy,' in connection with robots, and the consequent fear of robots above and beyond other products of technology. It is why I referred to such fear in my stories as the 'Frankenstein complex'" ("Introduction" 5). Asimov also explains that due to the popularity of the "story of Frankenstein", he "never felt the need to define the meaning of the term in any of [his] stories" (5-6).

For Asimov, then, the Frankenstein complex as discussed in 1984 is a combination of technophobia aimed at robots, as well as a pervasive cliché in robotic science fiction of having robots revolt against their makers. However, the Frankenstein complex is not only a reference to Shelley's novel, but also the broader myth of *Frankenstein*, as he explains: "People who know of the Monster only from the movie do not fully appreciate that the Monster [in Shelley's text] was rather movingly virtuous and became a killer only because he was unbearably ill-treated" ("Introduction" 5). In addition, Asimov claims that the success and popularity of the story of Frankenstein can be attributed *more* to the film than the novel: "Such is the power of the story of Frankenstein […] thanks even more to the movie than to the book" ("Introduction" 5). However, Asimov never clarifies which film he has mind. The assumption, in this dissertation, is that Asimov is referring to the 1931 *Frankenstein*, starring Boris Karloff, directed by James Whale.

Haynes has noted that although that film "ended with the Monster being burnt to death […] the box-office success indicated a sequel. The final scenes of the 1931 film were cut from all prints in circulation and *Bride of Frankenstein* (1935) opened" (18). In turn, *Bride of Frankenstein* was

"followed by a long succession of Frankenstein derivatives whose titles are sufficiently indicative of their content and of the way in which Frankenstein has been integrated into Western culture as an ever-contemporary by-word" (19). Indeed, there are as many as nineteen film adaptations from 1938 to 1993 (Haynes 19), with titles such as *Frankenstein Meets the Wolf Man* (1943) and *Frankenstein Meets the Space Monster* (1965). When Asimov wrote about the Frankenstein complex in 1964, there had already been twelve film adaptations. All of these adaptations can be traced back (as Haynes does) to the success of the 1931 film.

Despite Asimov's explications regarding his Frankenstein complex, this dissertation, instead, proposes that Asimov's ideas about the Frankenstein complex in robotic science fiction was primarily influenced by one particular robotic science fiction narrative. The text in question is Eando Binder's short story "I, Robot" (1939); the title *I, Robot* is more commonly associated with Asimov's own collection of short stories published in 1950. Binder's title was appropriated by Asimov's publisher against Asimov's wishes ("Robbie" 69). In addition, Asimov has stated that Binder's short story was a direct influence on him prior to writing his first robot story ("Robbie" 69). Asimov praised Binder's robot story (and later novel) by saying: "To anyone fond of the robot story in science fiction, ADAM LINK is of extraordinary interest. The robot-with-emotion has rarely been so well-handled" (Binder *Adam Link*). In Binder's originally short story, the robot, named Adam Link, is wrongfully accused of killing its maker, Dr. Link. Pursued by an angry mob, the robot discovers a copy of Shelley's novel, which was hidden from him by Dr. Link, and robot Adam strongly identifies with the plight of Frankenstein's creature. Adam Link shouts at the pursuing mob: "I hear you now – shouting outside – *beware that you do not drive me to be the monster you call me*" (17). When discovering a copy of *Frankenstein*, the robot wonders why its inventor never made him read it: "He had kept the book from me. Why? I read it now, in half an hour, by my page-at-a-time scanning. And then I understood!" (18). Ironically, the robot considers Shelley's text to be "the most stupid premise ever made: that a created man must turn against his creator, against humanity, lacking a soul. The book is all wrong. *Or is it?*" (18). The robot, now surrounded by an angry mob, explains in its confession that

it still possesses the power to escape, but "it would only be at the cost of several of your lives. And that is the reason I have my hand on the switch that can blink out my life with one twist" (18). Instead of killing humans, the robot decides to commit suicide, proving its humanity in the process: "Ironic, isn't it, that I have the very feelings you are so sure I lack?" (18). Given that Asimov was directly inspired by this narrative (and praised it), it is not unfounded to argue that Asimov had Binder's narrative in mind when setting out to overcome the Frankenstein complex in his own narratives. Despite Asimov's own explications, these insights already make it clear that his own robot stories would not adequately destroy the complex as he claims.

Given that Asimov praised Adam Link as a "robot-with-emotion", it is telling that Asimov describes his own robotic characters as follows: "My robots are almost invariably sympathetic, and if villains there be (though my stories rarely contain villains – only people), they are human" ("Introduction" 9). Asimov, writing about sympathetic robots, continued a tradition that was already instantiated by Binder, but also Lester del Rey's "Helen O'Loy" (1938) which was equally influential to Asimov prior to writing his first short story ("Robbie" 69). In del Rey's short story, a robotic wife wishes to be decommissioned (or euthanized) after the death of her (or its) husband, demonstrating that a robot is capable of real love. These two narratives, "Helen O'Loy" and "I, Robot" (and one can also include John Wyndham's "The Lost Machine" (1932)), are all about sympathetic machines. The 1930s robot narratives departed significantly from their 1920s counterparts in *R.U.R.* and *Metropolis* which presented machines as destroyers. Overcoming the Frankenstein complex was thus not a feat that Asimov can take personal credit for, as he does, although Asimov does admit: "I have never been notable for my modesty" ("Introduction" 10). Rather, it is more accurate to state that the Frankenstein complex is a bit of a misnomer, for what Asimov actually meant by that term was the *unsympathetic* and *technophobic* approach towards robots as mindless (or heartless) monsters; a similar misappropriation that occurred with the 1931 film adaptation of *Frankenstein* (discussed below). What Asimov, along with Wyndham, del Rey and Binder, was more interested in was to write about robots as sympathetic creatures, more in keeping with, rather than opposed to,

Shelley's creature. This serves to illustrate the importance of the theme of altruism in robotic science fiction.

The Frankenstein film adaptations took things in a different direction. While the robotic narratives sought to make their characters more sympathetic, Boris Karloff's (and subsequent) depictions played on ideas of surgical horrors. This is where another important point of connection can be made between robots and their archetypal siblings, zombies. Karloff's monster, also memorable for the metal bolts protruding from its neck, walking mechanically and having a squarish head, is aesthetically or visually situated somewhere between the aesthetics of being zombified and/or being an uncanny robot.[21]

Jeffrey Allan Johnson, in "Dr. Frankenstein, I Presume? Revising the Popular Image of Frankenstein" (2018), explores the impact of Karloff's depiction in more detail. According to Johnson, the 1931 film depiction did more than popularize a misconception of the creature, as it also served to "have made it difficult to visualize Shelley's characters as she originally intended" (287). Johnson provides an important explanation as to the aesthetic choices made for the 1931 film. In his view, the major difference between Shelley's original composition and the 1931 film's impact is the idea that Victor was a medical student: "Given the post-1931 identification of Frankenstein as a physician or at least a medical student, it is worth noting that in the 1818 novel medical study plays no obvious role" (293). Unfortunately, "most readers, unable to escape the influence of the Karloff Monster and his successors, have nevertheless assumed that the Creature is a 'stitched-together meat puppet' composed of used human body parts" (296). However, in Shelley's text, the creature is created through chemistry, and thereby should be considered as more holistically or grown despite the fact that the exact details remain obscured and mysterious (discussed later). Johnson also accounts for the decisions behind the Karloff monster: "Its stiches, scars, misshapen head, and

[21] It is worth noting that Masahiro Mori's concept of the uncanny valley positions zombies at the very bottom of the valley, making them entirely uncanny.

bolted neck resonated as a maimed and mentally incapacitated veteran of the First World War, then only thirteen years gone" (301). In addition, Johnson explains that:

> Despite some initial optimism about curing such men [veterans], and a general sympathetic attitude among the public in the immediate postwar era, the limitations of treatment had become increasingly obvious by the late 1920s as 'uncured shell-shocked veterans … became increasingly associated with crime, suicide and unpredictable […]'. By 1929, the short-staffed and under-funded United States Army Surgeon General's office 'chose to abandon treatment altogether'. Reports of violence by shell-shocked veterans were appearing in increasing numbers by the early 1930s, just in time for the appearance of *Frankenstein*. […] If the Monster is intended to evoke a war-ravaged veteran, Whale's film manages to elicit sympathy for him as well as horror. Like the veteran, the Monster is not simply a killer, but also a victim of the incompetence of modern medicine. (303)

As Johnson argues, it is this particular connection between the film and the First World War that also indirectly served to foster a connection between medical horrors and Shelley's creature. The Karloff Monster served as a vehicle for the metaphor of the veteran's plight. Johnson also emphasizes the connection to zombies: "As many observers have suggested, Victor's Creature [from the novel] thus falls into the 'uncanny valley' first imagined in 1970 (without reference to *Frankenstein*) by the Japanese roboticist Masahiro Mori. In his original diagram, Mori indicated the deepest part of the valley as a zombie; this would be roughly equivalent to Karloff's subhuman Monster" (304-5). Not unlike robots, the zombie trope can also result in intergroup competition between humans and zombies which may denote various themes of altruism. In the context of the 1931 film, one finds a strong thematic overlap between zombies and robots. At the same time as Karloff's hideous monster begged for sympathy, inhuman robots, via Asimov, Binder, del Rey and Wyndham, became sympathetic as well.

Alchemy and Robots

The great mystery of Shelley's novel is the manner in which the creature comes to be created. While the narrative provides some insights, Victor ultimately tells Walton that: "you expect to be informed of the secret with which I am acquainted; that cannot be" (46). Indeed, Victor's secret remains a secret throughout the narrative. However, it is important to note that the creature is not an automaton. It is, therefore, not merely a continuation nor direct application of La Mettrie's thesis that humans are machines. Rather, it is La Mettrie's other argument, his idea of the universal law, that is more specifically relevant to *Frankenstein.* The creation of the creature occurs in two phases; the first phase involves Victor's acquisition of the necessary knowledge, followed by the application of that knowledge in the second phase.

Victor explains that he became interested in "the structure of the human frame, and, indeed, any animal endued with life" (44). He searched for the primal cause of life: "Whence, I often asked myself, did the principle of life proceed?" (44). This led Victor to examine the causes of death in order to understand the causes of life: "To examine the causes of life, we must first have recourse to death. I became acquainted with the science of anatomy: but this was not sufficient; I must also observe the natural decay and corruption of the human body" (44).[22] He studied the "cause and progress of this decay", spending "days and nights in vaults and charnel houses" (45). While Victor examined "all the minutiae of causation, as exemplified in the change from life to death, and death to life," he experienced his eureka moment: "until from the midst of this darkness a sudden light broke in upon me" (45). Not only did he succeed in "discovering the cause of generation and life," he also became capable of "bestowing animation upon lifeless matter" (45). Once Victor acquired this secret knowledge, he "hesitated a long time concerning the manner in which I should employ it" (46). He pondered whether to create a simpler form of life or "a being like myself" (46). Although Victor acknowledged the tremendous undertaking of creating something as complex as a human, he

[22] Victor also makes clear that there are no supernatural elements involved in his creation: "In my education my father had taken the greatest precautions that my mind should be impressed with no supernatural horrors […] a church-yard was to me merely the receptacle of bodies deprived of life" (44-5).

hoped that his work "would at least lay the foundations of future success" (46). While the "minuteness of the parts formed a great hinderance to my speed", Victor set out to create a being "of a gigantic stature; that is to say, about eight feet in height, and proportionably large" (46). The creation of a single artificial being implies the ability to produce many, and Victor similarly imagines the potential outcomes of his endeavor: "A new species would bless me as its creator and source; many happy and excellent natures would owe their being to me" (47). In addition, Victor also hopes to fulfill the alchemical ambition of creating some kind of an elixir or panacea: "if I could bestow animation upon lifeless matter, I might in process of time (although I now found it impossible) renew life where death had apparently devoted the body to corruption" (47). Hidden in his isolated workshop, Victor explains that he also "collected bones from charnel houses; and disturbed, with profane fingers, the tremendous secrets of the human frame" (47). In addition, the "dissecting room and the slaughterhouse furnished many of my materials" (47). In Baconian fashion, he "pursued nature to her hiding places" (47).

Given the long legacy of *Frankenstein* with its innumerable adaptations and reinventions, the creature is often described (or depicted) as being some kind of an assemblage of corpses (and/or animal parts), stitched together and electronically animated. Such a perspective aligns the creature closely with modern conceptions of robots. For example, Kang describes the creature as "not a mechanical automaton but a reanimated being consisting of patched-up pieces of corpses" (218). Kakoudaki similarly views the creature's make-up as "a collection of disparate body parts" (34). Haynes writes that "In a macabre parody of Julien Offray de La Mettrie's *L'Homme machine* […] Frankenstein assembles his eight-feet-tall 'child' from the components of corpses and brings it to life with an electrical discharge" (25). However, if the creature constitutes a "macabre parody" of La Mettrie's thesis, then the question follows as to why organic materials were employed rather than mechanical components. Additionally, if the narrative were merely an endorsement (or parody) of La Mettrie's thesis that humans are machines, then the function of putrefaction and decay, along with the use of animal parts, would also be superfluous.

In “Dr. Frankenstein, I Presume? Revising the Popular Image of Frankenstein” (2018), Jeffrey Allan Johnson presents a different perspective on how the creature comes to be created. Johnson, “writing as a historian of chemistry in Germany” hopes to “clarify the relationship of chemistry to medicine in the German universities around 1800 in order to give a clearer picture of what Victor actually might have been doing as a student” (287). According to Johnson:

> For medieval and early modern scholars, the creation of an artificial version of human life in the form of a homunculus was a central concern. This in itself was not primarily medical. Nevertheless, Paracelsus is of particular interest in connection with Frankenstein, because Paracelsus’s work combined alchemy with medicine, and he believed that while classical humoral approaches to disease were worthless, chemical treatments […] could cure diseases like syphilis, which had been unknown to the ancient authorities. He thereby created a German tradition of medical chemistry that by the late eighteenth century was flourishing, so that university administrators were funding chemical laboratories particularly at medical schools. Thus around 1780 there were some twenty German chemists doing regular research, and a couple of hundred more doing occasional research. (292)

For Victor’s question, “Whence […] did the principle of life proceed?” (44), the answer would have been more closely associated with the study of chemistry as opposed to medicine (Johnson 292). Johnson suggests that “instead of surgically joining body parts, Victor used an alternative approach, using physiological chemistry to create the body parts he needed, with the help of a key modification of the conventional theory of spontaneous generation” (297). In other words, Johnson argues that the necessary parts were actually grown, manufactured or produced through chemical means from the materials Victor collected.

The problem that Victor faced, alluded to by both Shelleys in their prefaces, is the same problem Erasmus Darwin faced, as Mary Shelley explains: “[Darwin] preserved a piece of vermicelli in a glass case, till by some extraordinary means it began to move with voluntary motion. Not thus, after all, would life be given” (243). However, Darwin discovered “that only microscopic organisms

could apparently generate spontaneously" (Johnson 298). Consequently, Darwin himself stated: "the existence of spontaneous vitality is only to be expected to be found in the simplest modes of animation […] complex ones have been formed by many successive [sexual] reproductions" (qtd. in Johnson 298). The challenge posed in this context is to overcome the gap between micro-organisms and macro (or complex) organisms. As Johnson points out, chemistry provided the answer to Darwin when he discussed the *primordium of life*:

> So the spontaneous production of alcohol or of vinegar, by the vinous and acetous fermentations, as well as the production of a mucus by putrefaction which will contract when extended, seems almost as difficult to understand as the *spontaneous production of a fibre from decomposing animal or vegetable substances, which will contract when stimulated, and thus constitutes the primordium of life*. (qtd. in Johnson 299)

This insight is based on the work of Xavier Bichat (1771-1802) who proposed that "each tissue type" had a "particular organization and its own life" (qtd. in Johnson 299). Thus, decay can produce "not just a one-celled organism or a microscopic eel, but a fiber, which, if mass-produced, could become muscle or other types of tissues" (Johnson 299). These fibers can be grown into tissues and a large collection of them could be used to create a human: "If such fibers could be stimulated into life, then anything produced from masses of those fibers could also be brought to life, including a complex organism" (Johnson 299). While Johnson's account neatly explains the importance of decay for Victor's studies, there is also a historical connection between Bichat, Darwin and Shelley: "Bichat's ideas had been popularized in London in the spring of 1816 by none other than Percy Shelley's personal physician, William Lawrence, in lectures supporting a materialist interpretation of organic life" (Johnson 299). Although the creature is not an automaton, it is supportive of "a materialist interpretation of organic life", not unlike La Mettrie's materialism.

If the creation of the creature is (partially or wholly) synthetically produced or grown, then it also shares a closer kinship with Karel Čapek's robots in *R.U.R.* (discussed in Chapter 9). Johnson also notes this connection: "The idea of a corporate Frankenstein is much older, however, going back to

the 'Czech *Frankenstein*,' Karel Čapek's play *R.U.R.*" (305). As Johnson points out, while the play is famous for introducing the word "robot" into the English language, "the contemporary association of the word with non-living machines may lead scholars to miss the point that the Rossum robots are organic beings" (305). This dissertation agrees that there is tremendous overlap between Victor's concerns for the survival of humanity (discussed in the following section below), and the events that transpire in *R.U.R.* (discussed in Chapter 9). According to Johnson: "Utilizing this new form of living matter [in *R.U.R.*], the corporation founded by Rossum's nephew mass-produces the living robots that will fulfill Victor Frankenstein's worst nightmare by exterminating humanity" (306). In both narratives, sexual reproduction by artificial beings introduce the threat of human extinction, as Johnson also points out: "Although the typical robot cannot reproduce, we learn at the end [of *R.U.R.*] that a few specially modified ones are in fact capable of doing so, thus becoming the new species that will replace humanity" (306). While *R.U.R.* portrays the extermination of humanity, Victor also expresses the same anxiety of the extermination of humanity which, as a result of destroying the creature's mate, never comes to be realized. In both narratives, intergroup (or interspecies) competition between humans and artificial beings establishes a foundation for various altruistic actions and themes.

Before turning to the theme of altruism in *Frankenstein*, it is nevertheless insightful to retrace some of the historical developments that would likely have informed Shelley's conceptions of spirit, matter, automata and alchemy. What follows below is a discussion of three historical phases (for the purposes of this chapter, rather than comprehensive chronologically of historical ideas). First, a brief overview of the three alchemists that Victor discovers are discussed. Second, the transition from mechanical philosophies to vitalism plays an important role in Victor's education. Third, the philosophies of German Romanticism also reveal some of the underlying ideas and themes in the narrative.

The three alchemists that Victor studies are Cornelius Agrippa, Albertus Magnus and Paracelsus. Other than being three alchemists, they have another common element in their works,

namely automata. Albertus Magnus proposed the existence of "natural magic" which enabled one to "perform marvelous acts through the manipulation of forces inherent in the world", without the aid of demons or angels (Kang 58). Alchemy, for Albertus, also constituted a nonspiritual pursuit of knowledge given that it was based upon experimentation (Kang 74). What is particularly interesting about the legacy of Albertus Magnus is the connection to the word "android":

> One can find faint echoes of the items from the lists of wonders even in Denis Diderot and Jean le Rond D'Alembert's eighteenth-century *Encyclopedia.* In the articles 'Automate' and 'Androide,' the two traditional automata mentioned are the wooden dove of Archytas and the work of Albertus Magnus. […] The origin of the word 'android,' defined as an automaton specifically in the shape of a human being (as opposed to other living creatures), is obscure, but it is a medieval coinage from Greek roots ('andros,' man, and 'eides,' species) and is commonly linked to Albertus Magnus. The Renaissance historian Paolo Giovio (Latin name Paulus Jovius) in his work *Vitae Illustrium Virorum* (1549-1557) writes: 'Having become the master of the magical sciences, Albertus began the construction of a curious automaton, which he invested with the powers of speech and thought. The Android, as it was called, was composed of metals and unknown substances chosen according to the stars and endowed with spiritual qualities by magical formulae and invocations, and the labor upon it consumed thirty years'. (Kang 99-100)

While legends of the talking head are commonly associated with Roger Bacon, Kang provides some interesting precursory narratives of the same legend.[23] One such legend, written by Matteo Corsini in 1373, entitled *Rosario della vita*, involves Albertus Magnus as a great inventor (Kang 70). In this tale, Corsini explains that Albertus constructed a "metal statue modeled after the course of the planets, and endowed with such a capacity for reason that it spoke" (qtd. in Kang 70). Corsini makes it clear that there is nothing demonic about this statue, "it was not from a diabolical art or

[23] For a detailed account, see "Chapter 2: Between Magic and Mechanics: The Automaton in the Middle Ages and the Renaissance" (Kang 55-102)

necromancy – great intellects do not delight in such things" (qtd. in Kang 70). However, in Corsini's narrative, a monk visits Albertus' chambers and, upon discovering the speaking metal statue, assumed it to be "an idol of evil invention" and destroyed it (qtd. in Kang 70). A fifteenth-century variation of the story identifies the monk as "none other than […] Thomas Aquinas" (Kang 71). What makes this particular narrative interesting is that it is a morality tale about creating automata that deals with the dangers of ignorance or superstition rather than dark forms of magic.

Cornelius Agrippa's expanded three-volume version of *De Occulta Philosophia* in 1533 became the "central text of reference on all things magical during the Renaissance" (Kang 84). In Book 2 one finds a list of artificial wonders which is "of enormous importance to the history of the automaton idea" (Kang 84). Notably, Agrippa "uses the obscure Greek word 'automata', quoting Artistotle in his *Politics*, who was in turn quoting Homer in his reference to the self-moving tripods of Hephaestus" (Kang 84). Because of the popularity of Agrippa's text, Kang explains that the word became increasingly associated with an "artificial and mobile device" (Kang 85). The list of wonders "became a popular reference for automata that played a significant intellectual role in the early modern period" (Kang 85). Agrippa, like Albertus, was also a supporter of more natural (or empirical) forms of magic, which he called "celestial" magic, relating to "mathematics, mechanics, music and astronomy" (Kang 85). Although he considered such forms of magic inferior to "religious" magic, which deals with spirits, sacred rituals, and miracles, he nevertheless "argues for a naturalistic explanation of artificial wonders" (Kang 86). The automaton was therefore an important intellectual concept during the Renaissance, as the concept offered a "vivid example to illustrate the power of natural magic" when people "pleasurably pondered the blurring of the boundary between the natural and the artificial, the animate and inanimate, and the living and dead" (89).

Agrippa's contemporary, Paracelsus, "who gave a detailed formula for the alchemical creation of a smaller but very much living version of a human being" (Kang 89), namely the homunculus, is equally relevant in this context. The creation of the homunculus, as prescribed by Paracelsus, does not require any supernatural interference. As Haynes explains, one "can

understand the appeal of the homunculus-peddlers better" when realizing "that robots are of the same conceptual family" (10). The attempt to create a tiny human being "was an example of extreme hubris, since it claimed to by-pass both the Creator and the divinely ordained method for reproduction" (Haynes 10). It was provocative as it "challenged the Church's teaching that the soul was created at the moment of conception and mimicked both the Greek legend of Prometheus molding humans from clay and breathing life into them" (Haynes 10). In other words, the homunculus offers a means with which to create artificial humans through entirely materialistic means. In the case of Victor's 'homunculus', the size of his eight-feet-tall creature becomes symbolic of his hubris, as Victor himself explains that the creature's size was the result of his efforts to simply speed up the process: "As the minuteness of the parts formed a great hindrance to my speed, I resolved, contrary to my first intention, to make the being of a gigantic stature" (46). This is, of course, in direct contrast to the alchemical tradition, and also serves to show that his creation no longer belongs to that tradition.[24]

An important legacy of alchemy deals with the figure of the automaton (and homunculus) as conceptual objects that played a major role in the development of mechanics and experimental methodologies. Kang explains that more "nuanced approaches to the intellectual history [of the scientific revolution] […] have pointed to significant overlaps in areas of interest and methodology in the works of Hermetic thinkers and mechanistic philosophers" (100). From 1637 to 1748, classical mechanical philosophers were dualists, advocating the importance of mechanism, until "La Mettrie's materialist tract [which] marks the end point of the classical mechanist movement" (Kang 131). As discussed in Chapter 3, La Mettrie's thesis is distinct from Descartes (and other classical mechanical thinkers) when he negates the existence of the soul altogether; the human body is not unlike a wholly materialistic creature such as the one Frankenstein creates, devoid of souls or other

[24] Paracelsus was also interested in what might today be described as empirical or natural medicine: "In part owing to the long-lasting renown of Paracelsus as a founder of modern medicine, mercury remained the drug of choice for treating syphilis among other maladies well into the nineteenth century" (Peterfreund 83). Similarly, Victor explains that he was originally inspired by these alchemists to "banish disease" (32); an ambition that gave way eventually to hubris.

supranatural dimensions. However, at face value, it is also obvious that Shelley's novel is not a blind endorsement of La Mettrie's thesis (or mechanistic philosophies) given that Victor's creature is *not* an automaton.

We know that Victor Frankenstein's early studies were permeated by alchemy rather than 'modern science'. When Victor happens upon a volume of Cornelius Agrippa, his father dissuades his interest by saying: "My dear Victor, do not waste your time upon this; it is sad trash" (31). However, this has the opposite effect, as Victor explains:

> If, instead of this remark, my father had taken the pains, to explain to me, that the principles of Agrippa had been entirely exploded, and that a modern system of science had been introduced […] under such circumstances, I should certainly have thrown Agrippa aside, and, with my imagination warmed as it was, should probably have applied myself to the more rational theory of chemistry which has resulted from modern discoveries. It is even possible, that the train of my ideas would never have received the fatal impulse that led to my ruin. (31-2)

After witnessing the effects of lightening which "utterly destroyed" an oak tree, Victor learns of the power of electricity, which "completed the overthrow" of the alchemists in his mind (33). However, Victor still was disinclined to study "any modern system" of science or knowledge because of a course he attended on "natural philosophy" at his father's request (33). Due to some unexplained incident, Victor could only attend the final lectures of the course, which meant that he could not understand it: "I became disgusted with the science of natural philosophy, although I still read Pliny and Buffon with delight, authors, in my estimation, of nearly equal interest and utility" (34). Peterfreund considers Victor's judgment of Buffon to be "highly questionable" (81). Indeed, given the linkage between some of Buffon's ideas and alchemy, Frankenstein's remark can even be seen as rather ironic, as I will make clear in the following paragraphs.

Peter Reill, in his book *Vitalizing Nature in the Enlightenment* (2005), begins his analysis of "Enlightenment Vitalism" with Buffon, not because he was its founder, but because he "emerged as

an iconic figure in its elaboration" (10) by challenging the mechanical philosophers, such as Descartes, Newton, Boerhaave (whom La Mettrie studied under), and others (34). While these mechanical philosophers did not agree on everything (and cannot easily be placed in the same category), "none denied Cartesian dualism" (Reill 36). As Reill explains: "It is well known that the machine became the dominant metaphor to characterize nature during the first half of the eighteenth century" (36). For such thinkers, "the world machine was, 'composed of inert bodies, moved by physical necessity, indifferent to the existence of thinking beings'" (36-7). However, by the mid-eighteenth century, "some younger intellectuals considered mechanism's basic assumptions as neither satisfying nor self-evident" (37).

The problem with mechanical philosophy was its primary focus on surfaces as opposed to interior realities, and this was an issue that Buffon sought to address. For Buffon, "mathematical systems were hermetically sealed, closed to the realities of observable nature […] Physical truths in contradistinction were based on actual events" (Reill 40). Mathematics, and the branches of study associated with it, denote mechanical philosophers such as Descartes. Buffon praised Descartes for "reducing the explication of all phenomena to mechanical principles" as "grand and beautiful", yet he questioned whether "these mechanical principles [are] the extent of matter – its impenetrability, its movement, its external appearance […]" (Reill 42). For Buffon, the answer was clearly negative as his position was, as Reill explains, that: "Instead of concerning ourselves with such phenomena, we should attempt to penetrate beneath the surface to the inner nature of a body" (43). This affirms Peterfreund's notion that Victor misreads Buffon, but for different reasons than those Peterfreund outlines. The difference between exterior (surface) and interior (hidden force) is exactly the gap that Victor cannot peer across when he sees the creature for the first time. Horrified by the visual appearance of the creature, for example, whose soul is actually benign, Victor flees instead. His behavior accords more to the tenets of mechanical philosophy than Buffon's vitalism, which makes his remarks about taking delight in Buffon rather ironic.

This irony and misunderstanding also pertains to Victor's studies in relation to his meeting with Waldman, who "heard with attention my little narration concerning my studies, and smiled at the names of Cornelius Agrippa, and Paracelsus, but without the contempt that M. Krempe had exhibited" (42). Instead, Waldman's attitude is one that recognizes the historical contingency between chemistry and alchemy: "these [alchemists] were men to whose indefatigable zeal modern philosophers were indebted for most of the foundations of their knowledge. […] The labours of men of genius, however erroneously directed, scarcely ever fail in ultimately turning to the solid advantage of mankind" (42). Given that this statement "was delivered without any presumption or affectation" (42), and that Waldman also delivers his lectures "with an air of frankness and good nature" (43), Victor's attitude towards his scientific instruction at the university begins to change. Victor explains that it was "the amiable character of this man that inclined me more to that branch of natural philosophy which he professed, than an intrinsic love for the science itself" (43). The manner in which natural philosophy – with which Victor was previously *disgusted* – is communicated means that it now becomes appealing to Victor thanks to Waldman who seems to embody Buffon's ideals. However, despite Waldman's amiable character inspiring him to pursue science more rigorously, Victor's approach seems unable to go from the mechanical exterior to the interior, that is, its possible consequences: "the more fully I entered into the science, the more exclusively I pursued it for its own sake" (43). The pursuit of knowledge for its own sake is the beginning of Victor's downfall, as he acquires knowledge without any real regard for its effects on others than himself.

While Buffon was a figurehead of Enlightenment Vitalism, and read by Victor at university, one should also recognize the important connection between vitalism and Romantic *Naturphilosophie* worldview which Victor also would have been exposed to at Ingolstadt. Kang explains: "In the last decades of the eighteenth century, however, German thought tended to follow the larger trend of the late Enlightenment in moving away from classical mechanism in favor of more vitalist and organic views of the world" (188). The development of a new Romantic worldview,

namely that of *Naturphilosophie*, associated with figures like Johann Wolfgang von Goethe" owes a great deal to Enlightenment vitalist thought" (Kang 189). The Romantic worldview was not merely a continuation of vitalist thought, nor a complete rejection of Enlightenment; it constituted its own separate, although related, worldview. What the Romantics rejected were "ideas associated with mechanistic philosophy […] with its conception of reason in terms of the mathematical, the logical, and the quantifiable and its static models of the world, the state, and the body" (Kang 189). Yet for these Romantics, even vitalism was too modest in its rejection of mechanistic philosophies as it seemed to lack "metaphysical certainty and a clear vision of an eventual goal" (Kang 190). The result was that Romantic *Naturphilosophen* adopted "the basic vitalist model of the world as animate", but also "superimposed on their vision a grand idealist narrative that was both metaphysical and teleological" (190). Kang summarizes this grand narrative as follows:

> The central story they told was of a primordial unity that was broken at the birth of the cosmos, resulting in the splitting of all things into countless sets of opposing forces. In the manner of the dialectic, the antipathetic elements in every pair come into constant conflict with each other but ultimately resolve themselves into a new unity that, once again, comes into conflict with its opposite to eventually form another union. At the heart of the dynamic is the binary pair Spirit and Matter – contradictory in nature but endowed with an essential desire for reunion, a longing to a return to the original oneness. […] For the early nineteenth-century thinkers, this vision provided not only an explanation of why events in the world can be so violent and disorienting but also hope for an ultimate unity in which all things will eventually be healed and made whole. (191)

This grand ideal narrative is thoroughly at work throughout *Frankenstein*. If one looks at Victor and Elizabeth's education, one finds the competing forces of Spirit and Matter. Elizabeth is taught to draw, and her art is a combination of her internal creativity (Spirit) combined with the external act of painting (Matter) in service of pleasing her aunt. Victor, however, is a seeker of knowledge, who adopts the Baconian approach of forcibly extrapolating Spirit from Matter, or pursuing his

experiment merely for the sake of the science itself. This is where, as Peterfreund points out, both Waldman and Victor fail: "Waldman speaks of how scientists 'penetrate into the utmost recesses of nature, and show how she works in her hiding places' […] A few pages later, Victor confides, […] I pursued nature to her hiding places'" (Peterfreund 84).

As Peterfreund explains, Victor's failure is his inability to understand the relationship between natural knowledge and self-knowledge: "Victor could not have studied [Paracelsus] and not been made aware of the relationship" (86). Here, then, is the lesson of another connection between Paracelsus, vitalism and *Naturphilosophie* that is lost on Victor. According to Reill:

> late Enlightenment thinkers [including Buffon] argued that humans, being part of living nature, could acquire an intimate understanding of it through self-reflection, and vice versa […] In arguing for the strict identity between mind and nature, the *Naturphilosophen* collapsed the distinction between them, denying Enlightenment Vitalism's mediating logic. […] Thus, the only true path to discovering truth lay in investigating mind. Self-reflective consciousness directed toward nature produced self-evident knowledge. (Reill 210)

Thus Victor should have been become aware of the same lesson even if he had not read Paracelsus. Self-reflexivity is therefore certainly his failure. The relationship between the pursuit of knowledge and the state of one's mind was clearly cautioned to Victor by his father. As Victor recounts his father's moral lesson:

> A human being in perfection ought always to preserve a calm and peaceful mind, and never to allow passion or a transitory desire to disturb his tranquility. I do not think that the pursuit of knowledge is an exception to this rule. If the study to which you apply yourself has a tendency to weaken your affections, and to destroy your taste for those simple pleasures in which no alloy can possibly mix, then that study is certainly unlawful, that is to say, not befitting the human mind. If this rule were always observed; if no man allowed any pursuit whatsoever to interfere with the tranquility of his domestic affections, Greece had not been enslaved; Caesar would have spared his country; America would have been discovered more

> gradually; and the empires of Mexico and Peru had not been destroyed. *But I forget that I am moralizing in the most interesting part of my tale; and your looks remind me to proceed.* [emphasis mine] (48-9)

In this last sentence, Victor is addressing Walton, but the sentence also functions on a meta-narratological level, reversing the role between reader and mad scientist Victor himself. In the midst of his explanations of how enormous historical human catastrophes could have been avoided, the reader, like Walton and Victor originally, is only interested in the outcome of the creation of the creature, unlocking the secrets of nature. The reader, Victor and Walton, care little for the historical contingency and pursue knowledge for its own sake at this point; all that matters is the result, and not the process. Similarly, once he discovered how to bestow "animation upon lifeless matter", Victor forgets the intermediary steps and only focusses on the outcome: "But this discovery was so great and overwhelming, that all the steps by which I had been progressively led to it were obliterated, and I beheld only the result" (45).

The creature is the realization or result of such a superficial pursuit of knowledge; knowledge derived without contingency, through pure mechanical contrivance, a disregard for nature, humans and communication. It is the realization of a power without responsibility, and acquiring knowledge in such a way disrupts the "tranquility of mind" and "interferes with domestic affections" that his father warned about. Just as Victor was repulsed with Krempe's attitude and contempt for his studies of the alchemists, he becomes Krempe himself rather than Waldman; his pursuit of knowledge was as dismissive to his father's advice, his domestic affections (and relations to others), as Krempe was of alchemy; conversely, Victor is also guilty of Waldman's Baconian approach to nature. Victor's ugly soul is mirrored in the creature's matter, and Victor's beautiful material body is mirrored in the creature's (initially) benign soul. The Romantic grand ideal narrative implies that their destinies are now entwined. As mentioned, when the creature becomes animated, it does so with "a convulsive motion [which] agitated its limbs" (51). When Victor flees, and dreams of his dead mother in the next room, he too wakes up with "horror […] and every limb became

convulsed" (52). In other words, when Victor animated the creature – unlocking nature's secret – he simultaneously unlocked a secret of his own mind and acquired a terrifying self-realization; a realization that he cannot come to terms with. The dream of Elizabeth morphing into his dead mother, according to Irving H. Buchen, is significant: "Whatever sexual or Oedipal elements may or may not be at work here, what is clear is that Victor is reacting to his own act of creation as one that dispossesses the woman-mother from the entire process" (108). His mother's death takes on cosmic significance now that he created an unnatural form of life, circumnavigating the need for motherhood; his original intentions of helping humanity has turned into the very destruction of humanity. His psychological approach towards his pursuit of knowledge was equally *dead* and mechanical, as he described himself at that point: "I appeared rather like one doomed by slavery to toil in the mines, or any other unwholesome trade, than an artist occupied by his favourite employment" (49). He became the instrument of Matter or material forces, and thereby lost his proverbial soul, "I seemed to have lost all soul or sensation but for this one pursuit" (47). Victor's self-realization, as much as the creature's, is that he is himself a form of artificial life, separated from humanity, and compared to the gigantic creature, he has become the homunculus.

The Theme of Altruism

The theme of altruism is particularly present in *Frankenstein*. The novel opens with an explicit account of altruistic behavior in the second letter from Walton to his sister Margaret. Walton explains that his lieutenant, an Englishman, "in the midst of national and professional prejudices" nevertheless "retains some of the noblest endowments of humanity" (14). He is also "heroically generous", given that he met a Russian lady who loved another, but whose father would not consent to their engagement because her beloved was poor (14). The noble Englishman "abandoned his pursuit" of her and "bought a farm" which he "bestowed the whole on his rival, together with the remains of his prizemoney to purchase stock, and then himself solicited the young woman's father to consent to her marriage with her lover" (15). In this account, already, one encounters altruistic behaviors, signifying a concern for humanity over "national and professional prejudices".

Also early in the novel, when Walton and his crew discover Victor who is nearly dead, Walton explains: "if any one performs an act of kindness towards him, or does him any the most trifling service, his whole countenance is lighted up, as it were, with a beam of benevolence and sweetness that I never saw equaled" (21). This description serves to show the role-reversal between Victor and his creature, in that Victor is equally appreciative of such forms of generosity as those that the creature sought. It is not difficult to imagine that the creature would have reacted the same under similar circumstances (and momentarily does during the De Lacey scenes, discussed below).

There are acts of altruism between Elizabeth and Caroline (Victor's mother). When Elizabeth was sick with scarlet fever, Caroline tended to her despite being discouraged from doing so, risking her own life, and indeed that decision resulted in Caroline's death. After Caroline's death, Elizabeth "was continually endeavouring to contribute to the happiness of others, entirely forgetful of herself" (38). Shortly after, Victor leaves for Ingolstad, and the dichotomy between knowledge-and-isolation versus experience-and-companionship begins to form: "I, who had ever been surrounded by amiable companions, continually engaged in endeavouring to bestow mutual pleasure, I was now alone. In the university, whither I was going, I must form my own friends, and be my own protector" (39). Frankenstein spends some time on explaining his relationship to countenances: "My life had hitherto been remarkably secluded and domestic; and this had given me invincible repugnance to new countenances. I loved my [family and friends]; these were 'old familiar faces;' but I believed myself totally unfitted for the company of strangers" (39). However, Frankenstein explains that his spirits soon were raised by prospect of learning: "I ardently desired the acquisition of knowledge" (39). It is not merely knowledge to which he is drawn, but knowledge as compensation to his isolation. He exchanges human companionship, bestowing "mutual pleasure", for an isolated pursuit of knowledge. This method or form of learning is in direct opposition to the more natural way (and social forms of learning) that he and Elizabeth experienced earlier. According to Buchen:

> Like many before and after her, Mary Shelley clearly was enthralled by the issue of what was the essential nature of human nature. […] Was man basically good or evil, altruistic or self-

> serving? The dramatic difference of Mary Shelley's inquiry is its scientific framework. Her starting point is not some noble savage in a far-off country but an artificial person placed in a fictional laboratory in order to explore under controlled and observable conditions the development of a creature as an individual version of the human race (109)

Given these opening scenes of altruistic actions, this dissertation would like to simply add that Shelley's inquiry was not an inquiry but a clear and obvious statement. The creature is born virtuous and in Shelley's experiment, to use Buchen's perspective, this is clearly attested during the De Lacey scenes.

Between the creature and the De Lacey family, one finds a fictional Turing test moment. When the creature visits the cottage and converses with De Lacey (a blind man) whose children are away on a long walk, he explains: "I am an unfortunate and deserted creature, I look around and I have no relation or friend upon earth," to which De Lacey responds: "To be friendless is indeed to be unfortunate, but the hearts of men, when unprejudiced by any obvious self-interest, are full of brotherly love and charity" (137). De Lacey applies the Turing test directly when stating: "I am blind and cannot judge of your countenance, but there is something in your words which persuades me that you are sincere" (138). De Lacey can only judge the intentions and sincerity of the creature by his use of language, and the creature is perfectly capable of passing the test. When De Lacey offers to help the creature, he explains his own altruistic motivations, namely the personal gratification of being able to help another: "I am poor and an exile, but it will afford me true pleasure to be in any way serviceable to a human creature" (138). This kind of generosity affirms a group selection perspective in the sense that it is afforded to strangers. The creature responds happily: "You raise me from the dust by this kindness; and I trust that, by your aid, I shall not be driven from the society and sympathy of your fellow creatures" (138). The creature's appreciation reveals his hope of belonging to the society and sharing in the sympathy of fellow creatures. When the creature admits that he is seeking the companionship of De Lacey and his family, Felix, Agatha and Safie return. Horrified, Felix attacks the creature: "he dashed me to the ground and struck me violently with a

stick. I could have torn him limb from limb, as the lion rends the antelope. But my heart sank within me as with bitter sickness, and I refrained" (138). This realization that the creature could have torn him limb from limb is echoed by Binder's robot Adam Link (discussed earlier), stating that it could free itself from a persecuting mob but instead chooses to preserve life (and kill itself instead).

After his escape, the creature fluctuates between love and rage given the mixed results of his experiment: "When I thought of my friends […] a gush of tears somewhat soothed me. But again, when I reflected that they had spurned and deserted me, anger returned" (143). The irony here is that De Lacey coincidentally forewarns the creature about Felix's reaction when saying that love and charity of the human heart, or altruism, is only possible when the heart is "unprejudiced by any obvious self-interest" (137). Felix is spurred into action as a means of protecting his self-interest, namely his father, wife and sister. The creature's own self-interest, however, lies with his hope of being accepted into the De Lacey family. After that hope is lost, his conscience compels him towards death: "I bent my mind towards injury and death" (143). Between Felix and the creature, there is a clear line drawn in the sand between included and excluded members of their respective groups. From Felix's perspective, the creature falls outside of his family group; from the creature's perspective, his relation is simultaneously towards *any* human and therefore to *all* humans. This is the same cusp on which Asimov's robots come to misinterpret or confuse the meaning of the First Law: not having particular affiliations to any specific human being, the machines interpret the First Law to mean *all of humanity* and subsequently act as guardians to all of humanity.

The creature's experience with the De Lacey family affirms to the creature that it is not human, and does not belong to the human species. According to Buchen: "Frankenstein did not create a human being; he created a being who had the potential to be human" (109). This Turing test affirms that the creature is – like an intelligent machine – perfectly intelligent, but not human. Josh Bernatchez similarly points out: "this scene marks the moment when the Creature begins to consider himself categorically alienated from the human community" (76). The creature is also altruistic on two separate occasions. Initially, he anonymously provides firewood to the De Lacey family prior to

his expulsion from the human community. Shortly after this expulsion from the De Lacey family, the creature rescues a drowning girl and attempts to resuscitate her, only to get shot by another human. The creature's response is: "This was then the reward of my benevolence! I had saved a human being from destruction, and, as a recompence, I now writhed under the miserable pain of a wound" (146). Importantly, his "feelings of kindness and gentleness […] gave place to hellish rage and gnashing teeth […] I vowed eternal hatred and vengeance to all mankind" (146). A third attempt also fails, namely to "seize" a child (young William), "and educate him as my companion and friend", because "this little creature was unprejudiced, and lived too short a time to have imbibed a horror of deformity" (147). Naturally, this attempt fails and ends in William's death. In much the same way that Victor only created death rather than life, as his creation is a circumvention of womanhood, natural procreation, and God, the creature realizes that he is Victor's double upon killing William: "I, too, can *create* desolation; my enemy [like nature] is not impregnable [my emphasis]" (149). The only recourse is to ask Victor for a mate so that the creature can form his own community.

During the discussion between Victor and the creature regarding the creation of a mate, Victor is conflicted between two competing loyalties. As a creator/parental figure, Victor ponders: "did I not, as his maker, owe him all the portion of happiness that it was in my power to bestow?" (152-3). Victor is also compelled by his loyalty towards humanity: "I consent to your demand, on your solemn oath to quit Europe for ever, and every other place in the neighbourhood of man" (154). Victor is entirely aware of the fact that the creature poses a group-threat to all of humanity: "His power and threats were not omitted in my calculations: a creature who could exist in the ice caves of the glaciers, and hide himself from pursuit among the ridges of inaccessible precipices, was a being possessing faculties it would be vain to cope with" (154).

Although Frankenstein submits to the creature's request, he later changes his mind for an important reason. Victor and the creature are entwined. Just as Victor postpones his marriage to Elizabeth in order to create the mate (and the creature eventually kills Elizabeth as revenge), the two are married to each other. As Buchen describes it, putting off his wedding merely affirms "a

recognition of the common needs each shares" (110). Buchen also argues, whether "one agrees or disagrees with Frankenstein's later decision not to create the mate", "all the reasons he provides are evidence of a kind of forethought that he failed to exercise before" (110). The realization that compels Victor is simply that the mate, and/or their potential future offspring, might not comply with the creature's oath sworn to Victor. In other words, if the creature becomes part of nature, or "linked to the chain of existence" as it claimed earlier (154), the creature can potentially pose an existential threat to humanity as a species, it "might make the very existence of the species of man a condition precarious and full of terror" (178). This is the moment that Victor and the creature reverse roles once again.

The creature born benign, who acted altruistically, becomes increasingly corrupted and selfish. Victor, who only served himself in isolation initially, becomes an altruist when placing humanity's interests before his own: "Had I a right, for my own benefit, to inflict this curse upon everlasting generations? […] but now, for the first time, the wickedness of my promise burst upon me" (178). Victor realizes that creating the mate means that humanity will potentially compete against a superior species for survival. He concludes: "I shuddered to think that future ages might curse me as their pest, whose selfishness had not hesitated to buy its own peace at the price perhaps of the existence of the whole human race" (178). While Michael Page, for example, considers "Victor's fear that the Creature may spawn a new race to savage humanity" as "reflective of his own deep-seated self-absorption" (89), such a perspective (although valid) does not consider the possible motivation of altruism. Admittedly, Victor's decision to destroy the mate by no means signifies a complete redemption of Victor's character. It simply marks the moment when the creature's altruism ceases entirely, and Victor, if only for a moment, does succeed in becoming an altruist when putting humanity's interests before his own. Thereafter, the two are joined in their mutual isolation and destruction, divorced from all humanity. As Buchen puts it: "The final image of the novel is this ironically substitutive, for locked in eternal pursuit, the two now serve as the mates

of each other, the two halves of a whole that shall never be united except in mutual annihilation" (111).

The theme of altruism in *Frankenstein* in relation to the creation of an artificial entity had a direct influence on robotic science fiction authors such as Asimov and Binder. It seems logical that whenever a pseudo-human is created through artifice, the implication is that there exists the possibility of creating more, just as the creature demanded of Victor. The logical implication is that this process can continue indefinitely, and in so doing, one ends up with an inter-species rivalry or competition. One robot is already sufficient, therefore, to denote the extinction of humanity. A robotic revolt, therefore, does not have to take the form of standing armies or a race of machines, but a single entity signifying the potential to create more is already sufficient to conjure up themes of intergroup competition, and thereby themes of altruism.

Chapter 7: The Steam Man of the Prairies

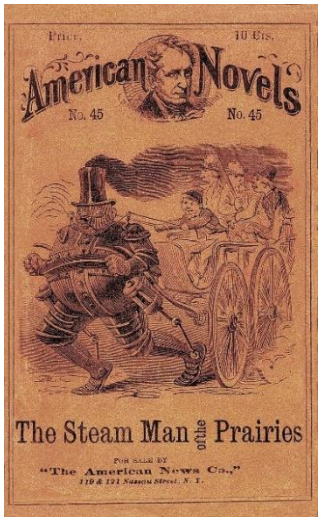

Figure 1 The Steam Man of the Prairies

Edward Ellis' *The Steam Man of the Prairies* (1868) is a dime novel about a young inventor who builds a steam-powered robot; the robot is not autonomous but piloted by the young inventor. The novel proved to be so popular that it resulted in an entire subgenre of dime novels containing young inventors and their fantastic machines, eventually "culminating in Hugo Gernsback's *Amazing Stories* [founded in 1926] – the birthplace of genre SF" (Landon 41). The novel's enduring legacy was due to the numerous "knock-off" versions by competing publishers, "such as Frank Reade, Frank Reade Jr., Jack Wright, Great Marvel's inventors, Frank Edison, Electric Bob" (Landon 41). In all these versions, one finds the same essential narrative of young inventors who invent "a progression of armaments in steam-drawn vehicles" and cause "consequent damage to Indians" (Landon 45). Ellis' inspiration of the contraption, however, is "Possibly based in part on an earlier claimed invention of an ambulatory two-cylinder rotary steam engine called the Newark Steam Man" (Landon 44). Lisa Nocks, in her book *The Robot: The Life Story of a Technology* (2008), discusses the Newark Steam in more detail, and explains that:

> After the Civil War, a team of mechanics in Newark, New Jersey, filed a patent for a mechanical steam-powered walking man that pulled a cart. Zadoc P. Dederick and Isaac Grass demonstrated their invention in January, and on March 24, 1868, they were granted

> patent #75,874 for their 'Steam Carriage' […] the carriage was meant to be used as a kind of taxi. In August 1868, *Beadle's Dime Novels* ran Ellis' story in issue #45. (50)

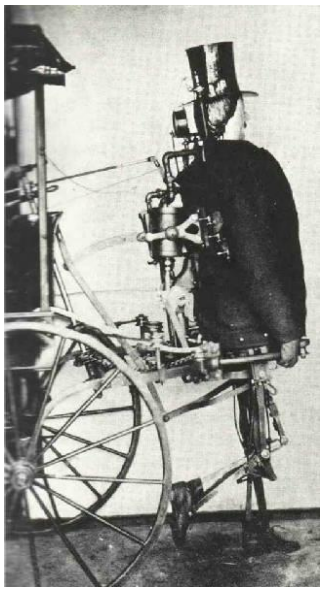

Figure 2 The Newark Steam Man

Such inventions caused great interest when published in newspapers, and reveals that Ellis' novel was participating in the "enthusiasm for technological innovation, and industrial and geographic expansion" (Nocks 50). Although a dime novel, *The Steam Man* had a much longer-lasting effect, given all the spinoffs, but more specifically ushered in the age of the robot in the cultural imagination of North America. As Nocks explains:

> For the next 50 years, reports of inventions of steam men (and subsequently electric men) appeared in newspapers from Ontario, Canada, to the Ohio Valley in the United States. They shared the spotlight with spin-offs and reprints of Ellis' dime novel, and even hoaxed reports of steam men. By the early decades of the twentieth century, the contraptions had been exhibited to thousands of people across the country. Although the inventions ended up in junkshops, and the dime novel readers' interest faded, the steam and electric men phenomenon had contributed to the emergence of the mechanical man as an icon of the future. (50)

In relation to Dederick and Grass' invention, one newspaper article, published on 14 March 1868, explains that the rumors of a steam man are indeed true, but still "far from being perfected" (Scientific American). However, "as soon as the steam man is in a condition to travel," the newspaper promises to explain its workings, and "In the meantime, we advise our contemporaries not to get excited over the steam man, for he is likely to remain harmless for the present" (Scientific American). A follow-up article, published on 21 May 1870, tells a different story: "we saw large posters announcing the greatest wonder of any age, past, present, or future, […] an imitation of the human form divine, impelled by steam" ("The Steam man" 1870). This newspaper article concludes that it is a "curious automaton," but "very much more satisfactory than his predecessor" ("The Steam man" 1870). By 15 April 1893, twenty-three years later, the steam man received a much longer exposition in the newspaper, now also with accompanying pictures, and a sensationalist tone; it is said that the steam man is "6 feet high," cannot "be held back by two men pulling against it," and "is attired in armor like a knight of old" ("The Steam man" 1893).

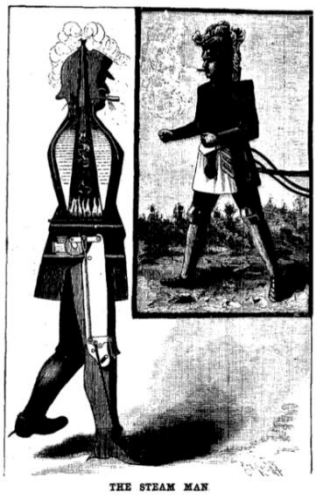

Figure 3 Newspaper Diagram of the Steam Man

Edward Sylvester Ellis was born in Geneva, Ohio, in 1840 ("Ellis, Edward Sylvester"). The son of "famous rifle-shot and hunter," Sylvester and Mary Alberty Ellis, Edward moved with his parents to New Jersey at the age of six ("Ellis, Edward Sylvester"). In 1887, Edward S. Ellis graduated with a "Master of Arts from Princeton College," and was a prolific (and successful) author long before and

after his graduation ("Ellis, Edward Sylvester"). Through the use of many pseudonyms, some of them still "perhaps unknown," and others perhaps even unknown to Ellis himself, makes it difficult if not impossible to trace all his works ("Ellis, Edward Sylvester"). He was a lecturer until the "mid 1880's, after which he devoted all of his time to literature" ("Ellis, Edward Sylvester"). However, from 1860, when his novel *Seth Jones; or, The Captives of the Frontier* was published (and subsequently reported to have sold over five hundred thousand copies), he garnered the reputation as one of the first pioneers of the dime novel ("Ellis, Edward Sylvester"). Although Ellis would write a number of popular works, many of which involve the frontier, it is not clear that he ever visited the frontier himself. His works, however, would have a long-lasting impact on the imaginations of younger minds and avid readers of dime novels.

In his novel, the young protagonist, a genius inventor by the name of Johnny Brainerd, is finishing his greatest contraption yet: a man driven by steam. Johnny encounters a mysterious traveler by the name of Baldy Bicknell. Baldy, who received his nickname as a result of having been scalped by Native Americans, immediately recognizes the steam man's potential to serve on the frontier against hostile Native Americans. Furthermore, Baldy claims to know a secluded and secret location of a river containing large amounts of untouched gold due to hostile threats in the region. Johnny and Baldy then strike an accord to procure the gold. Two companions, Mickey McSquizzle and Ethan Hopkins, who once rescued Baldy, are invited by Baldy to come along on their hunt for gold. Baldy, the experienced hunter and trapper, is familiar with the terrain; Johnny primarily functions as the steam man's machinist; the two companions, Mickey and Ethan, merely serve as muscle and occasional comic relief. During encounters with hostile Native Americans, the steam man manages to terrify and force them to retreat. After a few small digressions (such as having Johnny shoot a bear), the company are eventually ambushed at night but the crew manage an impossible escape after sacrificing the steam man; Johnny overheats the boiler causing an extravagant explosion. The novel closes with the promise that Johnny, with his newly acquired wealth, is

attending the best schools and constructing a new (probably bigger and better) steam man. Readers are advised to keep their eyes on the prairies.

Three dimensions of the novel are important. First, the narrative makes it abundantly clear that it contains no elements of supernaturalism, and inaugurates its own fantastic invention with a lighthearted and adventurous tone.[25] Second, it inaugurates the fictional robot with an overt mind-body duality. Third, the narrative represents several instances of intergroup competition and acts of altruism while such acts serve to only empower one group against the other. In addition, the altruism exhibited in the novel's climax is ultimately a feigned altruism in the service of self-preservation, and also comes at no real cost to any of the protagonists. It nevertheless occurs as the novel's climax and is very suggestive of the altruistic capabilities of machines.

The Machine Without a Ghost

Similar to *Frankenstein's* rejection of supernatural elements (discussed in chapter 6), *The Steam Man* makes it abundantly clear that there is nothing supernatural about this particular robot. As discussed in Chapter 1, the Steam Man more accurately constitutes a mecha or mech in contemporary terminology; it is a kind of robotic device piloted by a human. It merely provides access to unexplored terrains and riches that have previously been barred. For these reasons, the word "soul" never appears in the text. The word "God" only appears as an expletive.[26] The word "spirit" occurs once but not pertaining to any metaphysical quality (it describes the boy's attitude).[27] The word "ghost" appears three times, but actual ghosts never appear in the text.[28] Similarly, the word "phantom" describes the Native Americans, but never any supernatural kind of beings.[29] These

[25] The text frequently depicts such minor humorous moments: "the feet of the steam man began rising and falling with lightning like rapidity, the wagon being jerked forward with such sudden swiftness, that both Ethan and Mickey turned back summersets, rolling heels over head off the vehicle to the ground, while the monster went puffing over the prairie" (9).
[26] "Yer must do it, too, some day My God!" (51).
[27] " […] he was very quick and active upon his feet, and bounded along over the rocks, and across the chasms like a deer, with such a buoyancy of spirits that he forgot danger" (66).
[28] "'It's Baldy or his ghost'. It certainly was no ghost, judging from the manner in which it acted" (3). For the third occurrence see footnote below.
[29] "But they persevered, working with a strange persistency and silence, that gave them the appearance of so many phantoms engaged at their ghostly labor" (113-4).

empirical dimensions of the text illustrate that Ellis' steam man is a thoroughly embodied materialist machine. The major concern in the novel is gold, not salvation. As Baldy exclaims when he sees the steam man for the first time: "It's jest the thing for the West; we'll walk through the Injins in the tallest kind of style, and skear 'em beautiful" (22). The implication is that the machine is a source of cultural superiority towards supernatural/superstitious and religious beliefs.

The narrative also takes delight at making fun of individuals who attempt to attribute supernatural qualities to the machine. When the Native Americans return a second time, they are not as afraid because "their previous acquaintance with the apparatus had robbed it of all its supernatural attributes, and their halt lasted but a few seconds" (75). What is significant is that "supernaturalism" becomes a property of the steam man, something that is not real yet has real influence on the behavior of others towards the machine. When Mickey and Ethan see the steam man for the first time, the narrator points out on the first page: "His [Ethan's] practical eye saw that whatever it might be, it was a human contrivance, and there could be nothing supernatural about it" (2-3). Mickey is somewhat more apprehensive about the machine at first glance as he shies away like a horse "does at first sight of the locomotive" (4). The novel also opens with the religious (or blasphemous) phrase uttered by Mickey: "Howly vargin! What is that?" (1). Later, the narrator compares such first encounters with the steam man by saying:

> It is said that when Robert Fulton's first steamboat ascended the Hudson, it created a consternation and terror such as had never before been known, many believing that it was the harbinger of the final destruction of the world. Of course, at this late day, no such excitement can be created by any human invention, but the sight of a creature speeding over the country, impelled by steam, and bearing such a grotesque resemblance to a gigantic man, could not but startle all who should see it for the first time. (44)

In this scene, the steam man with its passengers race towards an emigrant train, running alongside it for some time. The machine's whistle, despite being simply a whistle, becomes a mythological siren, which "gave forth a shriek hideous enough to set a man crazy," while "the horses and animals of the

emigrant train could be seen rearing and plunging" (45). The men on the train "stood too appalled to do anything except gaze in stupid and speechless amazement" and "one or two, however, […] had sense enough to perceive that there was nothing at all very supernatural about it" (45). The narrator makes it abundantly clear that there is nothing supernatural about Johnny's invention. However, the act of feigning or letting others believe in its supernatural qualities is still presented as a source of joy and empowerment. The novel thus celebrates mechanical sciences, and even presents it as a good deal of fun to be had by those who possess such knowledge over those who do not.

In these moments, there is an implicit message to readers, whether intentional or unintentional. The message is one that (consciously or not) alludes to the concept of Manifest Destiny, as Robert W. Johannsen, in "The Meaning of Manifest Destiny" (1997), explains although initially denoting a "nonviolent" destiny, it became associated in its historical context with territorial expansion (10). During the mid-nineteenth century: "Manifest Destiny combined a fervent, idealistic, even mystical expression of Romantic nationalism with the realistic, practical consequences of extraordinary technological and economic developments as well as an unprecedented movement of Americans to distant parts of the continent" (Johannsen 13). The message is a call for mechanical conformity; those who do not understand the mechanics behind the contraption – who imbue it with supernatural qualities – are cultural defectors, essentially emulating the behavior of Native Americans. Supernaturalism becomes the steam man's source of power towards driving away Native Americans:

> Never was victory more complete. The Indians were thoroughly discomfited, and only too glad to get away after being so severely punished. During this singular running fight the steam man kept up a constant shrieking, which doubtless contributed in no slight degree to the rout of the red-skins. They fired continually at the fearful-looking monster, and, finding their shots produced no effect, invested the thing with a portion of the supernatural power which they had given it at first sight. (97)

The steam man acquires supernatural dimensions through misappropriation. To some, it is a frightful supernatural *other*, like a locomotive to a steed, a monster to the Native Americans, and also something supernatural to passengers on the train. Although merely a prosthetic extension of Johnny, and an invaluable economic asset to frontier travelers like Baldy, the voice of the machine is polysemic. On the one hand, whenever the whistle is blown, it announces the arrival of secular industrialism while, on the other, it also invokes supernatural terror to the uninitiated.

In so doing, the novel also introduces a new kind of cultural conformity: those who buy into the supernatural quality of the machine are deemed to be culturally other, both Native Americans and Europeans. Colonial society has a new cultural code, namely the code of industrial mechanics. Those who convert are promised great wealth and security, while those who do not are destined to live in fear and superstition. Whether Ellis intended this dimension or not, it is clear that with the introduction of this new machine, new cultural boundaries and groups are established.

Beyond the immediate use of the machine by the small company to procure gold, the machine also introduces new group selection pressures. Given the traditional boundaries the text established, namely colonialists vs. Native Americans, it subtly becomes clear that this boundary is no longer absolute. A new internal division appears whereby society is divided between secularists and super-naturalists. As mentioned, when the Native Americans see the machine a second time, some of their superstitious beliefs toward it have subsided, approximating the cultural values of the colonialists, while other Europeans endow the machine with supernatural qualities, thereby approximating the cultural values of the Native Americans. This is the awkward consequence of technology more generally: the utilization of the machine by a few members in actuality results in much greater social transformations at a group level. While the machine empowers individuals, each individual must also reaffiliate themselves accordingly to the new group selection pressures. The whistle is therefore terrifying to both Europeans and Native Americans who believe in supernatural realities. To borrow Rodgers' words (discussed in Chapter 5), the mechanical voice foreshadows a self-organizing system that deemphasizes the individual.

Johnny's Mind-Body Duality

Figure 4 The Steam Man and the Company

Another important feature of the novel is that Johnny's invention helps him to overcome his disability. The narrator introduces Johnny by saying: "If nature afflicts in one direction she frequently makes amends in another direction, and this dwarf, small and misshapen as [Johnny] was, was gifted with a most wonderful mind. His mechanical ingenuity bordered on the marvelous" (13). Johnny may be "humpback" (7), but has "an amiable disposition that made him a favorite with all with whom he came in contact" (13). Johnny never suffers any kind of social isolation as a result of his disability, unlike Frankenstein's creature.

In *The View from Outside Rockwell and Race in 1950*, Jennifer A. Greenhill discusses how "Black skin was linked with stoves, soot, and smoke in a range of media" (79). She turns to Bill Brown's analysis of Ellis' novel:

> This equation is distilled and expanded on in Edward Ellis's circa 1868 novel, *The Huge Hunter or, The Steam Man of the Prairies*, where the black body has actually become the metal machine. Here, as literary critic Bill Brown has noted, the servile mechanized body is racially coded as a finally mastered slave: with smoke billowing out his top hat, the Steam Man has a face 'made of iron, painted a black color, with … a tremendous grinning mouth'. (Greenhill 80)

In his own article, Bill Brown continues to argue that this implicit coding is not merely particular to Ellis' novel, but also applies to other editions of the text, and "invariably he is an enormous black

man" in most of them (131). The fact that the steam man often requires refueling, bringing the party to a halt, is also significant for Brown: "in its occasional moments of breakdown, […] the Steam-Man may be said to embody the threat of the slave's (or the recently freed slaves') violent recalcitrance" (131). Brown goes on to argue:

> More simply, while this technology releases Johnny from the able/disabled somatic binary, it does so only by racializing the mind/body and capital/labor binaries. Which is to say that the novel emancipates man from his body but incarcerates the machine within the American system of somatic semiosis. […] for if the 'natural slave' is he 'who is able to execute with his body what another contrives', then any American machine 'naturally' appears as an American slave, which means: a black American. The novel, participating in this logic, exemplifies Critical Theory's point that technology, far from being dependent on scientific neutrality, is and has been an objectification of divisions within society – in this case, an objectified preservation of divisions that have been politically (if not socially) overcome. (131-2)

Given Brown's observations, the steam man is also racially coded in another mannerism as well. The narrator points out: "The steam man was a frightful looking object, being painted of a glossy black, with a pair of white stripes down its legs, and with a face which was intended to be of a flesh color, but, which was really a fearful red" (6). The intended "flesh color" accidentally resulting in "fearful red" reveals an implicit connection to the Native Americans who are referred to as "red skins" throughout the novel. This additional coding is, as Brown puts it, further objectification of divisions within society. This additional coding gives the Steam Man some cultural ambiguity, on the one hand as a slave, on the other a Native American, but always remains a property owned by the West. The moments of breakdown represent the slave's potential recalcitrance, but the final scene also explicitly mentions the terrifying prospect of the machine being captured by Native Americans, and such cultural transgressions are depicted as life or group-threatening. The inherent quality of being a machine explains its lack of loyalty as much as the machine remains loyal. The threat is never

realized, but the notion that the machine can be captured is always present. From Brown's perspective, the same is true: the breakdowns threaten (symbolizing the slave's recalcitrance), but the threat is never actually realized as such (i.e. such breakdowns are never severe enough to genuinely hinder the company's progress). The machine is a mechanical form of altruism. Similarly in Edgar Allan Poe's story, "The Man that was Used Up" (2015; originally 1839), presents the character of Brevet Brigadier-General John A. B. C. Smith who bravely fought against Native Americans. While Smith is physically handsome, "with the countenance of the marble Apollo" (33), and referred to a "remarkable man" (34), he turns out to be *the man that was used up* when his entire body consists of nothing but protheses. Poe's character is an inverse of the steam man character, as his body has similarly been used up by the war-efforts, replaced with mechanical components, and subsequently (although satirically) celebrated.

Johnny never sacrifices the object for something or someone else, other than for his own survival. However, this is not a genuine sacrifice, as it only serves to concretize and exaggerate the status quo. The trajectories of Johnny's character development is entirely mechanical and circular, symbolizing his own mechanization (albeit expressed in optimistic terms by the narrator). Having achieved his goal of constructing the machine, Johnny is happily constructing a new steam man at the novel's end. Social transgression or social mobility is simply not permitted. In Ellis' text, one is allotted a social position, and like worker ants is destined to stay in that caste. The social position one is allotted may be exaggerated or expanded by technological means, in Johnny's case it involves an upward social mobility, but no possibility for social diversity. Unlike Frankenstein's creature, which undergoes humanization, Johnny remains human/inhuman from beginning to end. Johnny's character thereby emulates the steam man's development, as both remain mechanical and circular.

During one of the digressions, Johnny, alone, leaves the machine behind whilst exploring the landscape on foot. He encounters a bear which he shoots and kills. When he returns, he finds the steam man surrounded by Native Americans who are curiously examining it:

> Several climbed into the wagon, others passed in and around the helpless giant, and one valiant fellow bit him a thwack on the stomach with his tomahawk. His blow hurt the boy far more than it did the iron man, and he could hardly repress a cry of pain, as he looked upon the destruction of his wonderful friend as almost inevitable. (71)

The doubling continues in this scene as the very same man finds Johnny hiding behind a large rock and proceeds to raise his tomahawk to strike Johnny. After shooting the man in self-defense, the others flee and the steam man and Johnny are reunited. Having to fire the boiler up, Johnny hears the Native Americans returning, and in the "nick of time" gets the steam man running (74). A pursuit ensues, and Johnny "in his triumph, could not avoid rising in the wagon, shouting and waving his hat defiantly at his baffled pursuers" (75). He never reflects on his actions or experiences. Rather, in the following scene also involving pursuers and attacks, the steam man runs low on water. It becomes clear that Johnny's moral development mainly consists of mechanical lessons, as he laments to himself: "Why didn't I think and put a pumping arrangement to the machine? I could have done it as well as not, and it would have saved me a good deal of trouble" (80).

Johnny's mind-body duality is premised on the notion that technological prosthetics not only overcomes his disability but actually empowers him to posthuman proportions. Now that the machine exists, there is an ever-present danger that the steam man might be appropriated by other groups as well. This betrays the national group dimension of the text: the technology can actually serve humanity as a whole. On the one hand, Johnny's technological body exists separately and outside of his own body, but still belongs to him. On the other, given the racial coding of the machine, and its doubling as Johnny's posthuman prosthetic body, there is perhaps the very repressed desire of wanting to affiliate with other groups. The narrative teases and plays with the threat of cultural appropriation of Johnny's body, and these moments are as exhilarating as they are terrifying.

Group Selection and Altruism

In his book, *Does Altruism Exist* (2015), David Wilson (discussed in Chapter 5) employs a model with four quadrants in order to illustrate the relations between "the effects on self" and "the effects on others" whilst discussing the worldview of Ayn Rand and others (102). Here, the same quaternary model is used to discuss the relations of characters towards their social surroundings. According to this model, one can classify all types of actions that an individual can undertake in relation to society into four categories:

Society

Individual	-	+
-	A1 (--)	A2 (+-)
+	A3 (-+)	A4 (++)

The first action, A1, represents the kind of action that is injurious to both the individual and society. The second, A2, represents the kind of action that is injurious to the individual but beneficial to society (these actions qualify as altruistic). The third, A3, represents the kind of action that would be beneficial to the individual but injurious to society (such as selfish behavior). The fourth, A4, represents the kind of action that is beneficial to both the individual as well as the society. A broad consideration of, for example, *The Last of the Mohicans* (1826) shows that all four kinds of actions are present, demonstrating, as the title already suggests, a preoccupation with individual-social interactions and themes. In Ellis' novel, the frontier (and its occupants) have been reduced to a social microcosm, ultimately only permitting two kinds of behavior:

The Company

Steam Man

	-	+
-	A1 (--)	
+		A4 (++)

As far as the company is concerned, all actions can be classified as belonging either to A1 or A4, but not A2 or A3. In other words, instead of representing characters as individual heroes, having to establish or discover their place on the frontier, Ellis represents his characters as belonging (inescapably) to groups, which in turn belong inescapably to frontier politics. By removing the characters' ability to perform A2 or A3 type actions, Ellis amalgamates individual and society in a direct one-to-one relationship; the characters have no autonomy, and all their actions are judged in relation to how the group survives. The same is true for the machine itself. When it breaks down (or requires refueling), the company is exposed and vulnerable. Refueling the machine (attending to its personal needs) is concurrent with refueling the company (attending to the quest for gold). When the steam man functions adequately, its victories are only celebrated as such in relation to the company's victories.

Given the firmly established intergroup relations, one also finds altruism and inner-judges within the company. Inner-judges appear in the text when, for example, Baldy feels that he should reward Mickey and Ethan for having saved his life. Loyalty only exists among the company toward one another. Another instance of inner-judges can be seen at the novel's end, and serves to ensure that Johnny has the proverbial inner voice of humanity. At the novel's end, the threat is at its highest when the company is surrounded by Native Americans, and Johnny is to blame, seeing as though he fell asleep during guard duty. When the others realize their peril, they blame Johnny, but also recognize that his internal punishment was sufficient: "As a matter of course, they were all disposed to blame the author of this; but when they saw how deeply he felt his own shortcoming, all three

felt a natural sympathy for him" (118). According to Howard Bloom, conformity enforcement and inner-judges played a prevalent role in New England during the nineteenth century:

> New circumstances called for new conformity enforcers—highly portable ones—like the Victorian sense of guilt. Your parents didn't shame you publicly, they sent you to your room so your 'penalty for wrongdoing [would] … be exacted internally.' The government locked you in a house of penitence—a penitentiary—where your feelings of remorse would theoretically pummel you without cease. 'Each individual,' one reformer wrote, 'will necessarily be made the instrument of his own punishment; his conscience will be the avenger of society.' (90)

Pummeled by his own inner-judge, Johnny concocts a plan in order to rescue the company. Now that the threat reaches its peak, the company being surrounded, no interactions take place, only indirect competition. Baldy realizes the Native Americans have "outwitted him at last" (116), but their success would not be long-lived. The Native Americans constitute their own equal and opposite group when the narrator describes their silent cooperation, emphasizing twice that: "Not a word was exchanged, even in the most guarded of tones, for each understood his part […] Not a word was exchanged, for each knew what was required of him" (114). It becomes clear that one group is potentially able to appropriate the steam man from another group. When these intergroup tournaments are at their most intense, altruism is at its most invaluable. As a means of reiterating the seriousness of the threat, the Native Americans "had among their number those who had become pretty well acquainted with the steam man, else they would not have laid the plan which they did for capturing him" (122).

Johnny setting the machine to self-destruct is nearly altruistic, in the sense that he sacrifices his own invention, but is also a feigned altruism because Johnny gets to refabricate it, or fabricate a new one, after the novel's completion. The machine caused a "terrible" explosion, like an "immense bomb-shell," which "scattered death and destruction in every direction" (121). Even those Native Americans who escaped unharmed, were "beside themselves with consternation" (121). Thus,

although one finds a number of group selection themes and principles at work here, altruism is still sorely underdeveloped in this narrative as there are no overt scenes depicting such behavior. It may have been heartbreaking for Johnny to sacrifice his contraption, but in this context, does not constitute any real form of A2 behavior.

These discussions illustrate two arguments: Ellis' *The Steam Man* is certainly preoccupied with the question of group selection rather than individual selection. Despite not showcasing a sincere form of altruism, it does provide a scene of feigned altruism in the climax. Secondly, from a literary historical perspective, group selection featured alongside the arrival of this fictional robot in such an overwhelming capacity that individual selection is entirely absent. Although he may not have been aware of all these implications, it is clear that the depiction of Ellis' robot introduced group selection and implicit altruism whilst removing supernatural, uncanny and sublime dimensions from its invention.

Brown's notion that technology "is and has been an objectification of divisions within society" and not "dependent on scientific neutrality," takes on its full significance from a group selection perspective. The system, when confronted by intergroup tournaments, finds itself in need of altruism, and utilizes slave labor as a means of compensating for a lack of altruism. Wilson explains that: "In general, group-level selection always favors low-cost forms of altruism over high-cost forms […] High-cost forms of altruism exist primarily when there is no other way of benefitting the group" (81). Putting it differently, because there are not enough altruists – or altruistic behaviors – slaves are called in as compensation. Slavery might save money in the form of currency, allowing resources to be shifted to where they are needed or desired, yet slavery is ultimately anything but a cheap form of altruism. Subjugating a group of people requires tremendous cultural and social resources to instantiate and maintain. Transcultural sympathies are detrimental to the system: "People who agree with statements such as 'I think it is important to help other people' actually do help other people and work toward common goals more than people who express indifference toward helping other people" (Wilson 129). All empires built on slave-labor will, sooner or later,

discover that the cost outweighs the altruism obtained. This, in turn, leads to a greater demand for altruism, resulting in the acquisition of more slaves, which run up even higher costs, until the system collapses. During these moments, the fantasy of mechanical humans is at its most prevalent, and robotic machines at their most heroic. Nationalism and patriotism are similarly at their sharpest to mask the fact that there is something askew, and terribly unnatural, about the system's own internal organization. From this perspective, Ellis' text exhibits both the anxiety of sincere planetary or transcultural altruism, and the celebration of nationalism, and is entirely utopian. The implication is that the West no longer needs slaves because we have machines. The steam man qualifies as a low-cost form of altruism, while other human slaves are a high-cost form of altruism. Given the abolition of slavery a few years prior to the novel's publication, the text reacts enthusiastically towards the invention of the steam man. The historical invention of the Newark steam carriage, shortly preceding the publication of the novel, similarly promised to become the next source of cheap cultural altruism.

Ellis' text exemplifies these evolutionary dimensions quite clearly, but are certainly not exclusive to this text. What Brown observes as the conflation of the "American machine" and the "American slave" is certainly valid, but it is important to understand that such a conflation is not specific to America. *The Steam Man* merely exemplifies this process within the context of American slaves and American machines. As shown here, the interconnectedness between machines and altruism points to a much deeper property of human evolution. It is a universal evolutionary phenomenon in any culture that utilizes technology or slave labor. As discussed in Chapter 9, the word *robot* derives from the Czech word for forced labor, and was coined in an unrelated cultural context. From these perspectives, one can say that although none of the characters in Ellis' novel are altruists, there is one character who is nothing but altruistic, namely the steam man itself. This is not to argue that the steam man has some kind of subjectivity, but merely to point out that altruism is a ghost in this machine that only found fuller expression when the machines themselves became more intelligent, such as in *Tomorrow's Eve*, discussed below.

Chapter 8: Tomorrow's Eve

Tomorrow's Eve (1886) by Auguste Villiers de l'Isle-Adam features the creation of an android. The protagonist is none other than Thomas Alva Edison, portrayed as a genius inventor. However, in his "Advice to the Reader", Villiers explains that it is specifically the legend of Thomas Edison that serves as the character of his text and not the real Edison. In the context of this dissertation, three element of the novel are examined in more detail. The first is an examination of the paradoxical nature of the android and how it reflects on the paradoxical nature of the overall structure of the novel. Second, we will have a look at the novel's treatments of the artificial nature of language and belief and how it comes to be reflected through the various characters. Third, by questioning the sincerity between characters, we will show that altruism is an essential concern in this text. Before turning to these discussions, the remainder of this introduction focusses on Villiers' personal life and the origins of the novel as explained by de Heussey. The aim is not to validate or invalidate de Heussey's accounts of Villiers' personal life, but rather to examine it as a complementary text that sought to explicate some of the confounding elements of Villiers' novel.

The narrative opens with Edison reflecting on his own inventions and the fate of humanity before a disembodied voice belonging to the character Sowana interrupts him. Not much is revealed about Sowana but the reader learns that the two are working on a mysterious project called Hadaly. An acquaintance of Edison, named Lord Ewald, comes to visit and explains that he suffers from a dilemma. Lord Ewald is engaged to Alicia Clary, a beautiful woman who is also vain and materialistic. Ewald informs Edison that although he is in love with her, he cannot love the less flattering traits of her personality, which depresses him even to the point of suicide. Ewald explains: "I'm not one of those who can submit to accepting a body while they reject the soul" (35). Edison, listening to Ewald's dilemma, responds by saying: "there are some wounds that one can heal only by deepening them and making them worse, *I want to fulfill your dream in its entirety!*" (49). To Ewald's lament: "Who will deliver this soul out of this body for me?" (44), Edison replies: "I'll do it!" (49). Edison's

solution is to construct an android which will be a perfect facsimile of Clary, minus the less appealing traits of her character.

The majority of chapters consist of dialogue between Edison and Ewald, deliberating the implications and possibilities of performing such a feat. Ewald, although doubting the fantastic claims made by Edison, ultimately cannot maintain his skepticism against Edison's persuasiveness and eloquence. With the project already underway, Edison merely requires a willing participant, such as Ewald, who would agree to take the android's hand in matrimony. Although Ewald initially refuses, Edison's persistence and confidence results in the two making a pact: within three weeks, the android, named Hadaly, will be finished and should Ewald not be pleased by the result, Edison will personally consent to Ewald's suicide and hand him a loaded pistol.

A plan is devised to ensure that the android will be an identical replica of Alicia Clary. They play into her vanity by inviting her to Edison's mansion under the pretense that she is to pose as a model for a famous sculptor. The sculptor is actually Edison's mysterious assistant, Sowana, who takes the necessary measurements. When Ewald arrives on the last day of Alicia's modelling, she appears to Ewald to have had a tremendous change in character. She is suddenly everything Ewald had wanted – a cruel irony, now that the real Alicia has the ability to be compassionate and loving towards him. However, as readers may suspect, this compassionate and loving Alicia that Ewald has been talking to is in fact the android Hadaly and not the real Alicia. The android puts its hand on Ewald's shoulder and asks: "Dear friend, don't you recognize me? I am Hadaly [the android]" (192). As it turns out, Edison had already completed the project and Ewald failed to notice the difference between the real Alicia and the android (another Turing test moment). Ewald, shocked and conflicted, initially rejects Hadaly. Heartbroken, Hadaly accepts his choice when Ewald changes his mind and calls after her: "Phantom! Phantom! Hadaly! […] we must not part! […] it's the living girl who is the phantom!" (204). Thus, the experiment is a success.

Edison explains the mystery behind Sowana's character. Sowana is actually Mrs. Anderson, the widow of a friend of Edison whose illicit affair drove Edison's friend to suicide. Edison, trying to

help Mrs. Anderson through hypnosis, somehow caused her to become psychic, and in her entranced state adopted the name Sowana. Edison admits that the character of Sowana is ultimately mysterious even to him: "though I know Mrs. Anderson, *I swear to you on my soul* THAT I DO NOT KNOW SOWANA!" (211). It is simply explained that Sowana can "incorporate" herself with the android, and "animate it with her 'supernatural' being" (211). Now that Hadaly and Ewald are happily together, they set out to return to London.

Hadaly must travel in a coffin (the android is essentially switched off for the journey to London). The plan is to reanimate her once they arrive at Ewald's home. Unfortunately, another three weeks later, Edison reads in a newspaper that there had been a great disaster at sea costing the lives of seventy two victims. A fire broke out on the steamer, called *Wonderful*, when "barrels of turpentine and gasoline were ignited by an unknown cause, and soon exploded" (218). The article explains that a young Englishman, Lord Ewald, franticly tried to make his way to the baggage area, and "had to be restrained by no fewer than six sailors" (219). The coffin is lost, and Ewald survived. One of the first names on the list of deceased was Miss Clary (the real Alicia Clary). Three seconds later, Edison receives a telegram from Ewald, stating that "the loss of Hadaly leaves me inconsolable – I grieve only for that shade. Farewell" (219). Edison is left staring out the window at the "inconceivable mystery of the heavens" (219). The novel ends on this note of a failed ideal.

Tomorrow's Eve remains confounding and elusive on many levels. Paul K. Alkon explains that Villiers "is not concerned with the real Edison, or his real science, as material for fiction, but rather with taking a new mythology of science […] and exploring both the aesthetic and philosophical uses of that myth" (85). Thomas Edison was often employed in science fiction, given mythical and heroic status in texts, referred to as edisonades: "By the turn of the century, the edisonade was already a well-established formula," and "Ironically enough, one of the most literal of the early edisonades was [*Tomorrow's Eve*]" (Landon 44). However, not only was Edison rendered into a fictional legend, but Villiers himself also became somewhat of a legend in Mary Loyd's English translation of Robert Du Pontavice de Heussey's biography of Villiers (a cousin of Villiers), published in 1894, which

provides much contextual information about the novel's origins (and is discussed in more detail later).

De Heussey explains that Villiers was born of nobility and his ancestry can be traced back to "Philippe, Grand Master of the Order of the Knights of Malta, the heroic defender of the Island of Rhodes against Suliman in 1521", as well as to earlier notable figures in French history (De Heussey 6). His birthright was important to him, such that when "His rights to everything was disputed, ancestors, nobility, his very name! Villiers used to roar like a lion stung by poisonous flies" (14). Despite his noble birthright, Villiers chose the life of a Parisian bohemian. Symons explains some of the motivations behind Villiers' choice:

> To the aristocratic conception of things, nobility of soul is indeed a birthright, and the pride with which this gift of nature is accepted is a pride of exactly the opposite kind to that democratic pride to which nobility of soul is a conquest, valuable in proportion to its difficulty. […] The pride of being, [versus] the pride of becoming: these are the two ultimate contradictions set before every idealist. Villiers' choice, inevitable indeed, was significant. (Symons 53-4)

Villiers never doubted his noble heritage, but was also convinced that his nobility of the soul had to be earned. It already becomes evident that the difference between constitution (birthright) and substance (character or soul) were important to Villiers, and this notion also plays an important part in the novel.

According to de Heussey, Villiers also experienced a personal loss that would leave a long-lasting impact on his life:

> It was amongst the green fields and lanes of Brittany that there arose for him, to vanish almost immediately in death, that tender vision of womanhood which was his fleeting, but his only earthly love. […] I will not profane the sacred passion of these two young hearts by trying to describe it. I will only say, They loved [sic], and she died. (26)

Villiers wrote a poem after his beloved's passing (when he was 17 or perhaps 18), which de Heussey describes as marking "the close of the child's and the birth of the artist's existence" (26). While Villiers lost his ideal love at a young age, *Tomorrow's Eve* is a novel primarily concerned with regaining such lost ideals. Losing "the ideal" as a result of losing "the real" forms the foundation of the novel which proceeds in the opposite direction: recreating the constitution (the body of the android) allows for the recovery of lost ideal love. Indeed, de Heussey confirms that "Villiers never loved truly, deeply, ingenuously, but this once. No other woman ever took in his existence the place of the gentle, dead Breton girl" (33). Much like Ewald, Villiers' is said to be in love with an ideal that simply cannot be materialized.

Robert Martin Adams' translation of the French text (1984) includes another account of Villiers' personal life. In 1873, Villiers entered into an agreement with "the Comte de la Houssaye", who would provide Villiers a wife "worth at least three million francs" of which two hundred thousand "would be the broker's commission" (xi-i). Travelling to London in 1874 (at the age of 36), he fell in love with Miss Anna Eyre (Adams xii). As Adams explains, the courtship did not quite go according to plan: "he recited so much poetry, gave such a long reading from his next novel, […] that the young lady was frightened, […] and made her escape from his society as abruptly as she could" (xii). Adams links this event as a source of inspiration for the character of Alicia Clary: "what remained most strongly in Villiers' mind was the spiritless, blockish female who had been utterly incapable of responding to his romantic declarations" (xii). This episode, as Adams explains, closely resembles Ewald's own sentiments and predicament towards Alicia Clary.

These two events in Villiers' life certainly elucidate many of the complicated dimensions of the characters of Alicia Clary and Hadaly. The character of Alicia is certainly a fictionalized version of the "spiritless, blockish female" of Adams' account, the human android encoded with cultural values and entirely incapable of original thought. The motivations behind creating the android is intricately involved with recovering the loss of an ideal love, but also becomes a means with which to overcome these very cultural encodings. Symons describes that the protagonists of Adam's works

are “incarnations of spiritual pride, and their tragedies are the shock of spirit against matter, the invasion of spirit by matter”; his characters “seek phantoms, and find themselves,” or “seek the absolute, and find death” (57). These thematic dimensions are very present in *Tomorrow’s Eve*, where one finds an eternal struggle between the ideal and the material, resulting in trauma and/or tragedy. There is yet a third event in Villiers’ personal life provided by de Heussey that serves as a far more direct source of inspiration for the android.

Prior to writing the novel, sitting in a Parisian literary café, Villiers noticed “a young Englishman whose singular face aroused his imagination […] [Villiers] saw at once in the expression of his eyes that grave and scornful look of melancholy which always betokens a hidden despair” (de Heussey 160). According to de Heussey, it is this hidden despair in the stranger’s eyes that reverberated with Villiers. De Heussey only provides the initials of this individual, named “Lord E. W.” (161) which is a remarkable coincidence. As it turned out, a few days after Villiers saw him at the café, Lord E. W. committed suicide and became a popular topic of conversation in Paris for some time (de Heussey 161). Next to the body of Lord E. W. was a wax doll that was a replica of his fiancée: “[They] found an admirably-made lay figure, representing a young woman, whose waxen face, modelled by a great artist, was the portrait of a young lady well known in London for her brilliant beauty, and who had been engaged to be married to the eccentric young nobleman” (de Heussey 161). The curiosity of the situation caused many rumors and speculations. However, according to a personal acquaintance of Lord E. W., the man was in love with his fiancée’s physical appearance, but “held her mind and soul, and everything in her that was not material, in the deepest abhorrence” (162). According to Lord E. W.’s acquaintance, this affliction caused “the slowly-developed madness which ended in his death” (162).

The parallels between the historical Lord E. W. and the character of Lord Ewald in the novel are self-evident: the historical Lord E. W. was in love with a woman’s physical appearance, but disliked everything else, and had a wax doll or duplicate of his fiancée made, before eventually committing suicide. In the novel, Lord Ewald suffers the same dilemma, and also threatens to take

his own life as a result. While the wax figure is replaced by a fully automated android in the novel, de Heussey also explains that the android was not the result of Adam's own imagination, but also suggested to him by an American engineer who happened to work for Thomas Edison.

At a restaurant, "Villiers and a small circle of habitués," were discussing the case of the real Lord E. W., when an American engineer present boasted: "I am sorry your friend did not apply to me; I might have cured him" (de Heussey 162). Intrigued, the company asked how the engineer would have cured Lord E.W., to which the engineer responded: "Great Scott! I would have given this doll life, soul, movement, love!" (de Heussey 162). Here, clearly, one encounters the source of inspiration for the novel's android. According to de Heussey, the company burst out laughing, except for Villiers, "who seemed to be absorbed in rolling his cigarette" (162). The American engineer (an electrician), continued: "You may laugh, stranger […] but the time will come when my great master, Edison, will teach you that electricity is an almighty power" (de Heussey 162). It is not clear whether de Heussey's account is a fabrication, but it nevertheless demonstrates the kinds of beliefs or convictions surrounding the potential of electricity at the time. Gaby Wood's book, entitled *Edison's Eve* (2002), details the creation and history of (the real) Thomas Edison's talking dolls, fitted with phonographs, and manufactured not long after Villiers novel was published (163).

According to de Heussey, there is a historical connection between the novel and Thomas Edison. When the real Edison visited the *Exposition Universelle* in Paris in 1889, someone sent him a copy of *Tomorrow's Eve*, which he read in one sitting (de Heussey 163). Thomas Edison is reported to have said: "That man [Villiers] is greater than I. I can only invent. He creates!" (de Heussey 163). Edison also sought "to make the author's acquaintance" but Villiers "could not respond to Edison's invitation" because he was "already stricken by the fell disease of which he died" (163). This is indeed "deeply to be regretted" as de Heussey states (163).

Gaby Wood, however, provides a different perspective without referencing de Heussey's biography: "the only existing record of Edison's reaction to *The Eve of the Future* is a letter of 1910 from the Villiers de l'Isle-Adam 'committee' in Paris, thanking the inventor for his donation of $25

towards a statue of Villiers" (145). Furthermore, Wood discusses an essay by the historical Edison that closely resembles the sentiments of Villiers' novel: "[Edison] made his views clear [about women] in an article he wrote for *Good Housekeeping* in 1912, entitled 'The Woman of the Future' (Villiers, had he lived, would never have believed the coincidence)" (147). As Wood explains, the real Edison's personal diary provides two accounts in which he attempted to imagine a perfect woman by combining various parts of three different women whom Edison was acquainted with (146). As Wood states, Villiers would probably not have believed these coincidences.

Villiers did not seek emotional consolation for his characters, but treated them with intellectual rigor: "Villiers has no pathos. This is enough to explain why he can never, in the phrase he would have disliked so greatly, 'touch the popular heart'. His mind is too abstract to contain pity, and it is in his lack of pity that he seems to put himself outside humanity" (Symons 58). Having no pathos, his novel becomes satirical, and can be read as a thought experiment or exposé of introspections, at once self-critical and socially critical.

Figure 5 "Clemente Susini's dissectible Medici Venus (1780-82)"

His serious attention to technological innovation is also accounted for by Symons, who claims that Villiers' personal views towards his own art was one that involved a critical approach to science:

> [It] must be the whole effort of one's consciousness to escape from [science's] entanglements, to dominate it, or to ignore it, and one's art must be the building of an ideal

> world beyond its access, from which one may indeed sally out […] in a desperate enough attack upon the illusions in the midst of which men live. (55)

In the case of *Tomorrow's Eve*, Villiers is certainly not ignoring science, but sooner attempts to dominate it, and present an ideal world beyond its grasp. Hence, the android Hadaly is not merely a machine, but a conduit through which ideals are mediated. The machine is a vehicle for social critique and also becomes representative of the ideal of altruism as it is the most altruistic character.

Although Villiers maintained a firm belief in the metaphysical and ideal realms of human existence, he was an unforgiving author particularly in *Tomorrow's Eve*, as he "has no mercy on those who have no mercy on themselves" (Symons 58). Villiers was convinced that "stupidity is more criminal than vice; if only because vice is curable, stupidity incurable" (Symons 58). He was a harsh social critic who could only see "the gross multitude" of society (Symons 58). He would step out of his spiritual realms of idealism to attack the misconceptions of the general public around him: "in this disdain of ordinary human motives and ordinary human beings, there is at once the distinction and the weakness of Villiers" (Symons 58). His lack of sympathy also stems, according to Symons, from a lack of patient understanding (58). These are the general anxieties that underlie *Tomorrow's Eve*.

The Paradoxical Android

Although Villiers was an idealist, it would be an oversimplification to argue that idealism surpasses materialism in his novel. Considering the formal dimensions of the text, one notices that a moral confrontation is at work, whereby idealism and materialism undermine each other. Throughout the vast majority of the novel, one finds the realization of an android named Hadaly. The name is said to mean "Ideal" in Persian according to Edison in the novel (76).[30] Towards the end of the novel, the ideal is attained. The final chapter, entitled "Fate" (218), however, turns the tables, as all the characters are severely punished without further justification or clarification. It is never explained

[30] While this might not be a true translation, the definition provided in the novel is considered relevant in the context of the narrative.

whether this scene is supposed to represent divine punishment, or whether one is dealing with the arbitrary natural conditions of a materialist universe. Thus, the formal dimensions of the novel reveal a narratological ambiguity between idealism and realism: the ideal is possible, but undermined at the end, whereas the real, although the final arbiter of happy endings, is also undermined throughout the vast majority of the novel.

Thus in the treatment of the major theme of the novel – the loss of ideal love and the artificial means of recovering such lost ideals –, the ideal is realized but only momentarily before abruptly being taken away by inexplicable forces. Whatever explication one wishes to attribute to such causes, whether they are divine or arbitrary, does not diminish the severity. Fate itself becomes the final arbiter, fate as the unknown or inconceivable. Consisting largely of dialogue, the novel expounds numerous sympathies and debates, but ultimately renders all these arguments moot. Edison, who answers every question throughout the novel with conviction and eloquence, is by the end driven to silence, gazing at the inconceivable mysteries of the universe:

> […] then, raising his eyes even higher toward the ancient luminous spheres which still shone, unmoved, through the gaps in the heavy clouds, and sent their glints forever through the infinite, inconceivable mystery of the heavens, he shivered – no doubt, from the cold – in utter silence. (219)

These mysterious forces or truths escape even the comprehension of the genius Edison. The ineffable is the result of the competing tensions in the novel having reached their final conclusion: the dichotomy of life and death, spirit and matter, and also love and desire. Hadaly represents the amalgamation of all these tensions much in the way in which Frankenstein and his creature become entangled in the oppositions of life and death. As Wood observes, if Ewald did not threaten Edison with his own suicide, Hadaly would not have been necessary: “She is literally a life-saver, and can only be given life herself if she is needed to that degree” (137).

The character of Edison seems to be an amalgamation of opposing sentiments and thoughts as well. He is introduced as follows:

> A few years ago his features recalled in a striking manner those of a famous Frenchman, Gustave Doré. It was very nearly the face of an artist *translated* into the features of a scholar. The same natural talents, differently applied; mysterious twins. At what age did they completely resemble one another? Perhaps never. (7)

He is an artist and a scholar which implies that, as Rutsky explains: "it is precisely the *full* representation of life that is in question here" (41). Like Victor Frankenstein's initial ambition to banish disease and help humanity, Edison already helps his own culture to progress through the instruction of "his five acolytes" (7), and his desire is to do so for humanity across space and time: "What a latecomer I am in the ranks of humanity! Why wasn't I one of the first-born of the species? … Plenty of great words would be record now" (9). Edison's perspective presupposes that cultural evolution is not an ad hoc or arbitrary process, but teleological. As Wood explains, the historical Edison held similar beliefs as can be evinced from his own essay "The Woman of the Future": "The children of the future, the children of the exercised, developed man, and of the exercised, developed woman will be of mental power incredible to us today" (qtd. In Wood 148). For both the real Edison and the fictional character, the commonality is their interest in absolute perfection as the final goal. Wood's evaluation of the historical Edison's "eugenicist aims" as being "in the interest of perfection" (148), can be easily put alongside Rutsky's assessment of the fictional character of Edison: "Villiers' Edison envisions … a technological representation that would be fully present, a technological memory in which all of history would be brought to life" (41). Gasché similarly observes that Edison's reveries "all concern the dream of a total reproduction or reduplication of the universe from its inception on" (302).

Such ambitions can be correlated with contemporary transhumanist ambitions, such as envisioning utopian futures of posthumans existing as cyborgs, typically as the next step in human evolution. Rutsky also reiterates this idea when discussing more contemporary developments by stating:

> Research scientists, however, continue to work toward the ideal of a completely invisible technology: the dream of a computer and communications technology that would be 'ubiquitous' and yet entirely 'transparent.' Such a dream is, of course, merely the computer-age version of André Bazin's dream of a 'total cinema,' in which the technology of cinema would be effaced in favor of a 'complete illusion' of reality. (111)

In other words, scientific progress, as Edison's reveries reveal in the novel, is a progression towards the perfectibility of knowledge (omniscience), something which marks an end to human evolution itself. There is little regard for the history of human evolution, and even a disdain towards the irretrievableness of human history before technology was there to record it. It is therefore also important to note that the project, Hadaly, was already underway before Ewald's arrival, endowing it with a larger significance than consoling Ewald's dilemma. In the novel, as Rutsky points out, Sowana similarly "needs Hadaly's technological body to serve as the transparent instrument that conducts that desire, that focusses it. Together, then, Sowana and Hadaly will form the perfect machine for the total representation of life" (44). Rutsky states in his analysis: "The scientific-technological aspect of the android, in other words, becomes merely a transparent instrument that allows the circuit of desire between Lord Ewald and Sowana to be completed" (Rutsky 43).

If evolution is progressing to ever greater forms of perfection, and if Hadaly is the embodiment of perfection, the perfect ending is certainly cut short. The fateful ending implies that Edison's attempt to help Mrs. Anderson (who becomes Sowana) is also an utter failure, considering that Mrs. Anderson/Sowana sacrificed her life/soul for a short conversation and kiss with Ewald before perishing at sea. While the creation of the android is successful, it comes at the cost of all the lives of the characters around Edison. His initial statement is contradicted by the novel's ending: "there's no overcoming the craze for skepticism regarding my work […] and since everyone wants a good laugh, well, I'll have the last one" (11). However, Edison is not laughing at the end, but is driven to silence; the perfect and/or complete representation of life is at best achieved only momentarily, and cannot be sustained. Fate, as the final chapter is entitled, responds to Edison's ambitions and

lofty promises by a vicious confrontation of the inconceivable mystery or irony of life itself. In the case of Ewald, his initial reason for wanting to commit suicide is reaffirmed at the novel's end, implying that the entire process of creating the android merely served to postpone the inevitable by a total of six weeks. The ultimate significance of the android, supposedly representing the mysteries of life, was only ever a distraction.

An earlier draft of the novel contained a more descriptive title: "Edison's Paradoxical Android" (Wood 131). While the android is certainly paradoxical, the narrative itself eventually resembles the paradoxical android. Consisting of many small chapters strung together, like the different parts of the android, the sum of all the parts creates the illusion of a cohesive and consistent narrative but, also like the android, possess a mixture of subjectivities that ultimately remain inherently ambiguous. As Gasché puts it: "Hadaly is a book within Villiers' novel […] Hadaly is, indeed, nothing else than the structure proper to [*Tomorrow's Eve*], the *layout* of its conflicting interpretations" (321).

An earlier translation of Villiers' novel, entitled "The Future Eve", was serialized in the *Argosy-All-Story* magazine, from December 1926 to January 1927, by Florence Crewe-Jones (Bleiler 774). Bleiler has noted about this version that Crewe-Jones dropped all identifying references to Edison, and simply called the inventor Professor X. "The translation is smooth enough, but changes have been made and the work is disastrously abridged" (774). While Bleiler's description is accurate, it should not be taken for granted that Crewe-Jones was adapting the narrative for a pulp magazine. As is to be expected, much of the dialogue and loquacious descriptions by Villiers were significantly shortened. However, the overall plot and characterization remain the same despite Edison's name being changed.

Four changes are worth mentioning briefly to illustrate that the theme of altruism is still accurately portrayed. The first pertains to the introduction of the character of Edison (now called Professor X). There is no mention of the Professor's failed experiments regarding the train wreck. He is still a wizard of electricity, and the Professor's mannerisms are largely kept the same. From a

reader-response perspective, he also becomes less of a distinctive character and more mysterious, given the lack of detail regarding his reveries. Because the overall narrative is significantly shortened and simplified, the paradoxical nature of Ewald and the Professor's relationship becomes more obvious. Without a more elaborate introduction, the character of the Professor is still one of a mad genius, and his innumerable explanations for Ewald's questions and objections sooner appear to be the result of someone who is disingenuous and not merely an eloquent and profound thinker. For example, the Professor is introduced as a "humanitarian", prone to "fantastic and bizarre reflections", and while he "communed with his ego, humbly, sadly", his next thought is "[h]ow late I come in humanity's history […] I should have been born centuries ago" (5). Although all these descriptions occur in the original, now that his earlier inventions are omitted along with a more elaborate exposition of his thoughts, the reader must simply infer that his humanitarianism is predicated not on any of his actual inventions but his overall worth as an inventor to humanity at large.

Another significant change is the character of Alicia who, despite all her misgivings, is said in *Tomorrow's Eve* to have been faithful to Ewald. In *Tomorrow's Eve*, Ewald actually expresses a desire that Alicia should have been unfaithful to him: "Would to God she were capable of infidelity! […] For then, I'd have no trouble, *since she would be another person!*" (32). Conversely, in *The Future Eve*, when the Professor asks if she has been "faithful" to him, Ewald replies: "I would to heaven that she had!" (23), but explains that he does not doubt her love towards him. This change, although minor given that it serves no larger purpose in the narrative, is important nonetheless. In *The Future Eve*, making the character of Alicia unfaithful only serves to more succinctly humanize her character in a shorter span of words. In *Tomorrow's Eve*, Alicia is not only faithful, but according to Ewald, "she accords me the *only love of which she is capable*, and I believe it is all the more 'sincere', alas because she feels it IN SPITE OF HERSELF" (32). In *Future Eve*, this is merely shortened to: "I believe that I have the only love of which she is capable" (23). The difference between Alicia's fidelity also reflects the difference between the two Ewalds; the Ewald in *Tomorrow's Eve* more clearly expresses

Villiers' staunch idealistic sentiments in so far that Ewald so desperately wants Alicia to be "another person" that he wants her to be unfaithful to prove her human flaws. Conversely, for a pulp magazine, a more stereotypical depiction is employed when Ewald "bitterly" (23) admits to her infidelity but nonetheless remains convinced that she loves him.

Another interesting difference pertains to Kang's observation regarding the role of Sowana at the end of the narrative. While in *Tomorrow's Eve*, Sowana explains to Ewald that she controlled Edison, such a statement is not made in *The Future Eve*. The way that Sowana thereby comes to undermine Edison's power in *Tomorrow's Eve* is never depicted explicitly in *The Future Eve*. One infers that Sowana's soul still transmigrated into Hadaly when reading about Sowana's death, "a soul for a soul" (2270), and one also encounters the secret between Ewald and Hadaly that they keep from Edison/Professor. However, the secret has become much more mysterious in *The Future Eve* without any clear reference as to what the secret might be other than Hadaly possessing a ghost and being more than just a mere machine. Sowana's character therefore remains more mysterious (or ambiguous) in the abridged version, but the relationship between the Professor and Sowana is more clearly hierarchical.

Overall, the Crewe-Jones adaptation is a noteworthy adaptation given its intended medium. Aside from these differences, one should also note that Crewe-Jones definitely managed to maintain the overall paradoxical nature when it comes to the original narrative as pointed out by Gasché. Hadaly, initially, states in *The Future Eve* that it simply is not interested in living, "master, I do not wish to live," but eventually agrees "resignedly" according to Lord Ewald's wish (481). By the end of the novel, however, the reverse is the case, and Hadaly desperately wants to live, and is deeply saddened by Ewald's refusal and her goodbye kiss is her "final gesture of despair" (2091). In addition, while these passages are also found in the original, Crewe-Jones' portrayal of the Professor is wildly contradictory when it comes to his ideas regarding sincerity. Initially, he tells Lord Ewald: "My affection for you is sincere" (209), while later, he also admits to Lord Ewald that "Every man

plays a comedy – he deceives himself. We are none of us sincere because we don't even know ourselves" (1346).

As argued in relation to *Tomorrow's Eve*, the character of the Professor is thus equally dubious when it comes to his sincerity. In both narratives, Edison/the Professor spends considerable effort to convince Ewald of the monotony of all dialogue, and despite Ewald's reservations about conversing with a machine that only has a limited number of possible responses, the end result is something that exceeds those limitations. In other words, in both narratives, Hadaly is able to converse beyond the restrictions of her metallic encasement, and in both narratives, this undermines Edison/the Professor's various ideas and justifications. While *Tomorrow's Eve* provides some justification (Edison was under Sowana's control), *The Future Eve* does not provide such a clear justification, although in both cases the result is the same. One inevitably must consider that Edison/the Professor's intentions were not entirely sincere and that he does not clearly know himself.

In the context of this dissertation, the interpretation of *Tomorrow's Eve* still largely applies to *The Future Eve*. In some cases, the theme of altruism comes forward in a more direct manner in *The Future Eve,* given its more direct and simplified tone in narration, as, for example, when the Professor states "My affection for you is sincere" (272). The characters' motivations are simplified but nevertheless remain the same. Although Sowana's nature is not as fully explained in *The Future Eve*, Hadaly still escapes the Professor's initial promises and explanations, and as a result, the Professor's character is brought into question. In both narratives, his manipulation of Ewald is evident and his sincerity questionable, which in turn undermines the character's ability to authentically act altruistically towards other characters.

While *Tomorrow's Eve* contains an epigraph of Hoffmann's "The Sandman", it is worthwhile looking at another Hoffmann's short story "Automata" as an interesting parallel concerning *Tomorrow's Eve*'s ambiguity. Hoffman's short story contains several subplots, one of which involves two characters named Ferdinand and Lewis. As Eric G. Wilson explains, this subplot involves a

mysterious automaton called *The Talking Turk*, which is "extremely sophisticated, capable of movements almost as complex as those of human beings" (120). Furthermore, the Turk is "remarkably clairvoyant, able to offer insights into an individual's future" (120). Ferdinand and Lewis are skeptical of such claims, and set out to unravel the mysteries behind this device. When the Turk delivers a prophecy to Ferdinand, it mentions a portrait that no one else could have known about. Eventually, Ferdinand and Lewis "visit the workshop of Professor X, allegedly the engineer behind the building of the Turk and possibly responsible for the machine's oracular powers" (Wilson 121). They speculate as to what the underlying causes (or parlor tricks) might be, but eventually "leave the professor without an answer and are somewhat chagrined over his sarcastic manner" (Wilson 121). The Professor, an elusive character, is later seen at a wedding, participating as a witness, and the bride is none other than Ferdinand's ideal love – a woman he dreamt of, and which the Turk also prophesized about. There is much room for speculation, as Wilson describes in his analysis of Hoffmann's tale:

> This ambiguity especially holds true of Ferdinand's beloved. Barely glimpsed, she generates unanswerable questions. […] If she is a machine of the Professor, is she a vehicle of consciousness-raising beauty or a pawn in an evil game? These doubts surrounding the ideal beloved ensure that this woman can never be reduced to mere instrumentality, that she will avoid becoming only a mirror for the narcissistic projections of humans. […] This ironic idealism is rendered in the tale's fragmented form. A gathering of partial tales more than a complete story, a collection of irreconcilable philosophical theories as much as a work of fiction, the tale puts readers in a limbo of doubt that encourages both ignorance and awareness. (122)

Villiers' character of Edison shares the same sarcastic/elusive qualities, while the entire novel seems to incorporate the same kind of narratological ambiguity of Hoffmann's tale. Hoffmann's "collection of irreconcilable philosophical theories" are similarly paralleled in Villiers' text. The narrative invites interpretation, like the android, but ultimately resists all interpretations. Gasché similarly interprets

Tomorrow's Eve as inherently paradoxical and ambiguous, and this dissertation therefore agrees that "the literary critic determines the 'eternal meaning' of the text" (321). This also applies to *Tomorrow's Eve*. The relationship between reader and text is mirrored between Ewald and the android, as when Hadaly states: "A single thought from you could give me life" (202).

In *Tomorrow's Eve*, Sowana describes the android as a "*single* duality" and states that: "she is going to incarnate!" (12). At least from Sowana's perspective, the android is both a ghost and a machine. As Kang explains: "Villiers turned [Descartes'] machine-man analogy upside down by presenting an electromagnetic android that is controlled by a spirit, asking whether such a creature could be considered a living being" (246). However, to consider the narrative cohesive is similarly to play the part of Ewald whose acceptance of the android "as a true woman lies in the eradication of the line that separates the natural from the artificial, the biological form the mechanical, and finally the living from the dead" (Kang 246). Ewald fails to recognize the paradox as such, and also fails to recognize his own narcissistic projection as his own interpretation of the android. As Gashé succinctly describes: "these things which one cannot firmly conceive to be or not to be, or to be both or neither, are the objects of opinion (*doxa*), the intermediary faculty betwixt and between knowing and ignorance" (309). To borrow Wilson's description of Hoffmann's tale above, Villiers' narrative puts readers in a "limbo of doubt", of knowing and not knowing, that "encourages both ignorance and awareness". As such, in Villiers' narrative, the android is indeed a single duality as an object of opinions – a paradoxical android.

The Ghost in the Machine

Given the paradoxical nature of Villiers' narrative, the misogynistic elements of the narrative eventually deconstruct themselves. This is not to suggest that Villiers in any way intended such a reading, as Cloonan argues:

> While [Villiers] certainly aimed to castigate the modern woman, what comes forth in his work is something more nuanced, much less a critique of women than of the men who belittle them in words and actions. This is a perspective of which Villiers undoubtedly would

> not have approved. Hence the principal difficulty in a literary analysis of the novel lies in the clash between Villiers's well-known opinions and how they are reflected, and at times altered, in the text. (43)

In the context of this dissertation, it is the paradoxical nature of the narrative that inevitably comes to reflect on its own assertions, escaping Villiers' own control as it were. In the novel, one finds a brief passage where Edison discusses the lack of control that any creator has over their creation: "that's where we all stand, all of us! Nobody can see the real character of what he creates because every knife blade may become a dagger" (62). Kang also observes:

> There are interesting gender implications in this story that envisions the creation of a whole new woman through the use of technology and spiritualism, but the misogyny in much of the narrative is subverted by the revelation of the crucial role played by a female being who is of supernatural origin. (246)

In a later publication, Kang explains his interpretation by citing Sowana's own descriptions which reveal that she (as a spirit) manipulated Edison without his knowing. As Sowana explains to Ewald: "I called myself into existence in the thought of him who created me, so that while he thought he was acting of his own accord, he was also deeply, darkly obedient to me" (198). As Kang explains: "This means that Sowana was not just a persona taken on by Anny [Anderson] in her state of hysterical catalepsy, but an actual supernatural spirit that could extend her presence outside her body" ("The Question of the Woman-Machine" 38). In addition, Kang explains that this "supernatural turn in the plot undermines Edison's masculine perspective and power", given that he was "duped by a female spirit who collaborated with him only to bring herself into the world" ("The Question of the Woman-Machine" 39). This dissertation certainly agrees that these words undermine and subvert Edison's misogyny, but simultaneously asserts that these dimensions of the text, as Cloonan states above, are beyond of the authors' control and intentions. Thus, when Kang states: "So the novel not only exemplifies late-nineteenth-century male discourses […] but also contains a critique of it" ("The

Question of the Woman-Machine" 39), one should not interpret Kang's words as indicative of Villiers' personal intentions or critiques.

What is interesting about the narrative is the misogynistic double standard continually asserted without ever being addressed. As Ritch Calvin points out, it is not clear why the "fictional Edison never suggests that any of the blame might lie with Anderson [or Ewald], in particular, or the men who succumb to these 'Eves', in general" (5). Instead, Edison only criticizes women like Evelyn Habal and Alicia Clary: "[Edward Anderson] I neither excuse nor judge! But I declare [Evelyn] guilty, above all, of a capital crime" (112). The misogyny becomes so blatant than even Ewald confronts Edison: "you talk of women with a great deal of severity" (86), to which Edison responds in defense: "For there are, and there always will be, women who will be fully inspired by some principle more lofty than that of Pleasure. *Such women have nothing to do, don't you agree, either with this laboratory or with the question that lies between us?* [original emphasis]" (87). However, this generalization does not help to assuage the overtly misogynistic overtones of Edison. It is also worth pointing out that Edison is not interested in saving women, but rather to save men from such women, as Jennifer Forrest also argues: "Edison's project does not focus on finding a cure for the 'disease' that causes women to resort to artifice. This is not even entertained as a viable solution" (30). Instead, what concerns Edison is: "the thousands of [men's] lives lost each year to these vampire-like women" (Forrest 30). It is therefore important to take note, as Kelly points out, that: "The only two women to have good qualities are not really whole persons: one has an ill body [Mrs. Anderson], the other is bodiless [Sowana]" (129). Both the misogynistic dimensions as well as the artificial reconstruction of women in the novel have been convincingly and effectively dealt with by several scholars (Kelly; Lathers; Forrest). The intention in this section is to show that Edison's misogyny also represents a lack of self-reflection and self-awareness as a scientist, a master of the material realm, who is ultimately manipulated by the forces of spirit. Edison is also a symbol of science and his blindness towards Sowana similarly reveals the ways in which scientific discourses of the time failed to adequately understand human nature.

The revelation that Edison has been under the control of Sowana (and subsequently kept as a secret from Edison), casts doubt on everything that Edison argues or states throughout the entire novel. For example, when Lord Ewald points out that he is taking away Edison's masterpiece by betrothing Hadaly, Edison simply replies that he intends to work on other projects: "Not at all […] I still have the formulas. But… I shall make no more Androids" (217). This statement seems contrary to an earlier statement by Edison: "There's no doubt that within a few years substrata like this one will be fabricated by the thousands […] a factory for the production of Ideals!" (147). In addition, if Edison was interested in rescuing men from vampire-women (the thousands of lives lost each year), it makes little sense to stop his production of androids. This particular change in sentiment is never addressed or explained; if his android is such a lifesaving device, it is simply not explained what other possible projects or discoveries would take precedence over a factory of such ideals. Presumably, given that Sowana manipulated Edison, there is no longer a requirement or need for additional androids now that Sowana has incarnated. It seems equally likely that Edison, not sure of his own motives (being under Sowana's control), fabricated his own justifications, and he becomes the automaton.

As Forrest out: "The symbolist poet, especially, espoused a Platonic notion of an ideal world of which our world is just a copy. The poet-as-seer boasted the ability to go beyond concrete nature and to report back on what he had seen" (28). While Forrest's description is true of how Villiers viewed his own position as the author, it is a much more problematic description of Edison. The character of Edison as the master of materialism occupies the world of copies, matter and facsimiles, while the real world of ideals remains inaccessible to him. He is utterly at the mercy of the realm of ideals, as he admits that he does not know Sowana (211), but he demonstrates a general lack of appreciation and awareness of spiritualism.

When Ewald sees the robotic nightingale for the first time, he believes it to be real. Hadaly tells Ewald he may "admire it" but warns him not to "try to understand how [the song] is produced", otherwise "God would withdraw from the song!" (95). When Ewald tells Edison that he is impressed

by the nightingale, Edison informs him that it died two months prior. When Ewald expresses his disbelief that the nightingale is dead, Edison again responds in his typical contradictory fashion: "Dead, you say? Not altogether, since I've recorded here his song and his spirit. I evoke it by means of electricity; that's spiritualism put in really practical terms, right?" (95). For Edison there is no separation between spirit and matter. It is in his mind, rather than in Ewald's, that the distinction between life and death is eradicated. Edison even boasts that by using the device it warms up and generates heat: "You can light your cigar at the soul of that nightingale" (95). What amounts to blasphemy is a source of joy and pride for Edison. This similarly contradicts his earlier reveries, when Edison muses about sounds and their reproducibility, stating that the source or cause of the sound is more important than the sound itself: "To hear the sound is nothing, but the inner essence, which creates these mere vibrations, these veils – that's the crucial thing" (14). However, according to his own formulation, Edison nevertheless fails to realize (or refuses to admit) that the nightingale *is* in fact dead, and there is no "crucial thing" left in this animatronic machine; the sound and soul is one and the same. This lack of realization is similarly repeated by Edison's lack of realization that there is more to Hadaly than he thinks.

If Edison represents science, then Ewald represents the society that comes to rely on his knowledge for guidance. In line with Kang, Kelly, Lathers and Forrest's demonstration of how nineteenth-century scientific discourses sought to convince society of the nature (and danger) of women, Edison similarly must convince Ewald of his own (superior) convictions and knowledge. In this case, Edison must convince Ewald that his robotic female is somehow equal (or more preferable) to Alicia: "I [Edison] promise to raise from the clay of Human Science […] a Being *made in our image*, and who, accordingly, will be to us WHAT WE ARE TO GOD" (64). Although Ewald points out the obvious: "such a creature could never be anything but a doll", Edison convinces him that between Alicia and Hadaly, Alicia will appear to be the doll rather than Hadaly (64). As nineteenth-century scientists like Jean-Martin Charcot demonstrated the nature of hysteria before live audiences (Kang "The Question of the Woman-Machine" 32), Edison and Ewald's dialogues feature as a similar

Scientific presentation. This much is suggested in the novel when Edison draws the comparison: "Come, my lord, between the two of us, we form an eternal symbol: I represent Science and the omnipotence of its delusions; you are Humanity with its paradise lost" (71). The question of whether Edison succeeds in persuading Ewald of his own convictions is difficult to answer. While Ewald consents to Edison's explanations much of the time, it is not without challenge, disbelief and often the result of fatigue rather than intellectual persuasiveness. For example, after Edison's long diatribe about good/bad women and his experiment, Ewald seems ambivalent, as he stands up "WITHOUT replaying" to Edison's words and lights a cigar (87). Ewald then merely states: "You have an answer for everything […] We can leave whenever you're ready" (87). When learning the truth of Hadaly being possessed by Sowana, one can say that Edison's explanations fail to persuade Ewald entirely, now that Ewald knows a certain truth that Edison does not know. Learning the truth about Sowana's manipulation of Edison, Ewald rather than Edison qualifies as the poet who eventually peers beyond the concrete reality into the realm of ideals. Conversely, Edison's various justifications and explanations are suddenly unreliable.

Edison the scientist affirms: "Never forget that the only things we see in objects are those which our eyes *suggest* to us. We only form our ideas of them from the few glimpses of their real being that they let us catch" (66). Later, when asked whether Hadaly can see, Edison admits: "Who knows […] We don't see all that well ourselves, for that matter" (83). These words become important when Edison dissects Hadaly, showing Ewald the inner workings of the android. All that is revealed (or seen) during the examination are the particular mechanical devices of which Hadaly is comprised. However, Hadaly's artificiality is in turn compared to the artificiality of Alicia and Evelyn. When it comes to the character of Alicia Clary, much of Edison's diagnoses are derived from his own conclusions regarding the character of Evelyn Habal. To borrow Kelly's description, Alicia is presented in the text as follows: "the artificial [social] coding has become her nature, and Alicia is the creation of the bourgeois culture that has written her identity; she is a kind of mass-produced,

common subjectivity. This real woman is but an artificial product of her surroundings; she is the artificial doll" (128). Cloonan, however, attempts to describe Alicia somewhat differently:

> She is an ordinary, poorly educated young woman determined to create a decent life for herself, despite the numerous obstacles besetting women of her background in the latter part of the nineteenth century […] She is indeed a very modern woman, all too conscious of her limited options in the society of her day. She has calculated her assets and will attempt to use them wisely. (62-3)

In the novel, Edison anticipates such a defense by "So-called modern spirits", which he summarizes as follows: "everybody knows they [women like Alicia and Evelyn] use these qualities to make their fortunes, which is nowadays the main aim in life – especially nowadays when our 'social organizations' don't permit them many other ways to get ahead" (113). However, Edison does not blame the social organizations, but rather the artificial nature of such women who arm themselves with cosmetics: "For the prettiness of their persons is quick to become *artificial* and in time VERY ARTIFICIAL" (114). In a much discussed scene when Edison attempts to explain this artificiality by showing Ewald two film reels, one of beautiful (and artificial) Evelyn and the second of an unattractive (and natural) Evelyn, he is himself enamored by the first film: "The electrician seemed lost in a romantic reverie. You would have thought he was waxing sentimental over the girl himself" (117). This moment actually reveals that Edison, caught in the realm of the material, informed only by what he sees, is intrinsically a part of the very problem he identifies.

What these contradictions illustrate is the blindness of Edison the character, but also of the Science which he represents in relation to Ewald's Humanity. As Edison and Ewald form an "eternal symbol" (quoted above), the eternal symbol between science and humanity, this also suggests that there is a dialectical play between science and humanity. Instead of objectively proving his assertations, he merely needs to convince Ewald, and thereby achieve scientific objectivity. This is similarly what Forrest observes, when writing: "Inversely, a vicious circle established itself with science and medicine taking the 'proofs' of social (and professional) bias and presenting it back to

the community at large as the stuff of observed science" (28). Edison attempting to help Ewald is a case of the blind leading the blind. Earlier in the narrative it is made clear that he cannot see Sowana, although she has been listening to him all along; when he asks where she is, Sowana explains: "I'm close beside you, in fact I've been listening for some minutes now, while you were playing with words like a child" (11). As Cloonan describes: "What Villiers despised in science was its all too frequent association in the nineteenth century with a simplified form of positivism, which essentially maintained that the only truth was scientific truth and that science was the only viable path to knowledge" (49). What becomes most remarkable about Edison's character is his inability to doubt his own convictions. Instead, throughout the novel, he remains entirely convinced and certain of his own convictions. According to Cloonan: "Edison […] is a stranger to doubt and insecurity; his self-confidence never wavers" (64). However, the ending of the novel does seem to suggest a reversal of Edison's confidence.

Cloonan' interpretation of the reversal (or blow to Edison's confidence) at the end of the novel is brushed aside by saying: "This silence would seem to suggest that the inventor has abandoned any desire to rebuild another android" (64). Despite Edison's earlier claim that he does not intend to build any more androids, Cloonan speculates that Edison might in fact build more, should, for example, a "second Lord Ewald" show up, or if Lord Ewald himself returned instead of having committing suicide (64). This dissertation disagrees with such an interpretation simply because Cloonan' considerations are speculative beyond the scope of what is actually given in the narrative. As already mentioned, what is more striking is the incongruity between Edison's earlier explanations and his later change in attitude without clarification. Furthermore, Edison is not merely sitting in silence at the end of the novel, he also shivers. Yet, even his shivering must be narratologically justified and explained; while staring through the open window at the "inconceivable mystery of the heavens, he shivered – no doubt, from the cold – in utter silence" (219). Edison does not shiver from some sensation of horror, sadness or fear (considering his friend just informed him of his suicidal intentions), nor does he shiver from the inconceivable mysteries of

the universe, but simply from the cold. This justification, in the context of a tremendous disappointment to Edison, only seems presumptuous and overbearing. Science cannot tolerate the inconceivable or the ineffable, as stated in Edison's earlier musings: "*it's in ourselves that the killing silence exists*" (14). The novel concludes with Edison – the inventor of the phonograph – shivering in silence.

The Theme of Altruism

Tomorrow's Eve does not depict any forms of overt altruistic acts except when recounting the past relationship between Ewald and Edison. It is revealed that Ewald once saved Edison's life, as Edison explains:

> No, I've not forgotten that admirable young man who saved my life all those years ago now, when I was dying of hunger and collapsed in the street up there near Boston. Everyone else walked past me, saying, 'Poor fellow!' But he, the best, the most gracious of Samaritans, without wasting any time on pity, found it in his heart to pick me up, and with a handful of gold to save my life, my work! – And so he remembered my name? I'll receive him with open arms! Don't I owe all my success to him – and the rest of my life? (16)

This description, by comparison to the next chapter, serves to illustrate a dramatic difference between Edison and Ewald's characters. Edison, as explained in the following chapter, invented a device that would prevent two trains from colliding. After Edison convinced the director of a railway company to conduct an experiment of his new invention, the experiment resulted in a catastrophic accident. However, the accident was not due to his invention but human error: "But the engineers lost their nerve at the last minute, in the face of imminent danger, and went quite counter to the instructions of Edison" (18). Edison stood on a hillside nearby watching, and the scene is depicted as follows: "several hundred victims were scattered across the landscape, helter-skelter in every direction. People were crushed, burned, and ground to bits, men, women, and children, both the engineers, and the firemen, of whom it wasn't possible to discover even a trace" (18). Edison simply murmured: "Clumsy idiots!" (18), but what is more remarkable about Edison is that he is "astonished

only that the Americans shrink from a second trial, or, as he sometimes says, 'a third, if need be' – until, in fact, 'the procedure is successful'" (18). These descriptions reveal Edison's lack of regard of human life when it comes to the all-important pursuit of scientific knowledge and invention. However, as Cloonan rightly points out, it also reveals "Edison's inability to grasp the human dimension of his experiment, to understand that people are fallible" (60-1). While Ewald will help a suffering individual, Edison is prepared to sacrifice them in the pursuit of knowledge. His difficulty in understanding why others do not pursue scientific knowledge with the same gusto illustrates that he has no further connection to humanity at large; he is a man of machines and machines to him represent an end in itself. The narrative describes him as follows: "He is a dauntless experimenter who is gentle only to his proven friends" (18).

The chapter in which the railway incident is mentioned also introduces the artificial forearm that Edison is working on. This amputated arm, the "right hand of a young woman", has an appearance "cruel" as it is "fantastic" (17). Prior to the railway incident, the narrative asks: "What unknown danger could have necessitated such a perilous amputation?" (17). By subsequently narrating the railway incident, the implicit connection between the amputated forearm and dismembered bodies at the scene of the accident suggests some preoccupation with reconstructing a body; perhaps, the engineers' approach to rectifying the loss of life. There is no mention of a guilty conscious (quite the contrary), but the amputated forearm (his project of creating artificial flesh) may suggest that Hadaly is Edison's indirect response to his disconnection with humanity. Indeed, as he explains (and later contradicts), creating artificial women is his solution to the thousands of lives lost each year.

The same forearm comes into play later when Edison presents it to Ewald as a means of demonstrating the lifelike quality that the android will have. In a Turing test moment, Ewald exclaims: "The weight! The modeling! The exact coloration! […] Isn't it real flesh" (60). Edison's response is that "it's better than real!" given that flesh "fades and grows old" (60). As such, Edison calls his artificial flesh a "direct rebuke to the complacency of 'Nature'" (60). Natural humans

(especially women) are inferior to (man's) technological inventions. Edison's disconnect from humanity is also expressed in a wildly contradictory statement shortly afterwards. First, Edison proclaims: "I offer you, myself, a venture into the ARTIFICIAL and its untasted delights!" (71). When Ewald asks Edison what he would do if their roles were reversed, Edison replies: "I should blow out my brains" (71). Although humorous, this statement also reveals Edison's perception of Ewald and, by extension, humanity. Clearly, they do not belong to the same species given that Edison promises Ewald that he will hand him the pistol personally if "I haven't restored in you the will to live" (73). If Edison can restore Ewald's will to live as he promises, it is not clear why would he not be able to do so in himself if he suffered from the same affliction. It is such moments throughout the narrative that begin to call into question Edison's motives and nature.

Although Edison informs Ewald that a single objection will mean the end of their experiment, it appears that he is lying. While examining the android, Ewald clearly objects: "If that's the extent of the comedy in which you are asking me to take part forever, it's an offer that I can only refuse – and I should tell you so at once" (133). Yet, Edison's rhetoric manages to win Ewald over once again. The problem for Ewald is that the android will simply repeat itself forever, and such monotony is what drives him away from being with such an android. However, Edison merely latches onto the word "comedy" here and follows up with another loquacious diatribe. However, this diatribe reveals more about Edison's character and the way he views the world:

> Everybody plays in the comedy! And must, perforce! And every man with himself! Being sincere – that is the only dream that is absolutely beyond all hope of realization. Sincerity! How would it be possible in any case, since nobody knows anything? Since nobody is really persuaded of anything? Since nobody even knows who he really is? […] As for love! Well, if two lovers could ever see each other *plainly* […] their passion would evaporate in an instant! Happily for them, they always manage to forget this inescapable law of physics, that 'two atoms can never make real contact with one another.' (134)

For Edison, insincerity is the way of the world, and love is a physical impossibility. In addition, Edison's explanation of sincerity brings his own diatribes into question; if no man knows himself, it is not clear where that leaves Edison. Although Ewald admits that: "It's all very persuasive" (135), he does not actually consent to the experiment. He goes on to state: "You may be able to deceive my eyes, my senses, and my intelligence by this magical vision: but can I ever forget, within me, that she is only an impersonal object?" (135-6). Edison explains, essentially, that all love, conversations and daily actions of one's life consists of mere repetitions and routine; the implication is that humans already are machines. It is in this context that Edison proceeds to dissect Hadaly.

The narrative explains that Ewald "felt that beneath this strident, scientific demonstration two things were hidden in the lecturer's infinite range of severely controlled thoughts" (143). The first is a "love of humanity"; and the second "the most violent shriek of despair" (143). Edison's facial features are also described as "glacial in their austerity" (143). Without a love of humanity and/or despair, along with his glacial austerity, means that Edison is beginning to resemble the machine himself. After the "anatomy lesson" (147), Ewald begins to laugh: "then, seeing that Edison was laughing too, a strange sort of hysterical hilarity gained upon him" (147). Not long after, when discussing Alicia's teeth, Ewald is again overcome by laughter: "Don't mind me […] Go on! It's marvelous! I'm dreaming! I can't stop – and yet I don't really want to laugh" (154). Shortly after, Hadaly, with her "gift of second sight" can watch Alicia sitting on the train, trying to read the message the Edison and Ewald had sent, when Hadaly herself begins to laugh: "Lord Ewald understood that the Android was making the point that she too could laugh, and at living beings" (155). These scenes of laughter accord perfectly with Henri Bergson's theory of laughter in his *Laughter* (1900).

As explained by Eric G. Wilson: "For Bergson, the core of comedy is the blending of human and machine: '*The attitudes, gestures and movements of the human body are laughable in exact proportion as that body reminds us of a mere machine'*. Laughter arises when we witness a person behave as if he or she were a mechanism" (53). The competition between free will and one's body

being overpowered by natural forces – such as slipping on a banana peel – reveals this disjunction between subjective intention and objective compulsion. As such, Hadaly's laughter at Alicia trying to read a message on the train which is travelling too fast accords perfectly with such a definition. This is a complete role reversal as well as foreshadowing, in that Hadaly will replace Alicia (the machine will become the real, and vice versa). However, by implication, Ewald's laughter at Edison similarly implies that he is acting from compulsion as opposed to free will. Given Sowana's manipulation of him, such an interpretation is only supported in the narrative as well. Earlier during Edison's reveries, he surmises that fate should have made his phonograph appear sooner in history, but he describes own purpose by saying: "After all, though, what's that to me? Invent! Invent! That's my job" (14). It is not important to what use his machines will be put, but his job is merely to invent.

If the forearm represents a Turing test instance between Ewald and artificial flesh, the final meeting between Ewald and Hadaly constitutes a more proper Turing test. Believing to be with the real Alicia (and not the android Hadaly), Ewald's reaction to Alicia's words becomes significant in the context of altruism. Alicia (or pseudo-Alicia) displays an act of empathy and concern: "it seems to me that you've been depressed for the last few days […] I may be more your friend than you imagine" (190). Ewald pours out his soul and expects Alicia to laugh at him, but instead of laughter, she shows compassion when she replies: "How you suffer! […] and all because of me!" (191). Edison is overjoyed. All that was required was a human word: "A single *human* word had been enough to touch his heart, and to rouse in it indescribable hopes [my emphasis]" (192). More importantly, however, is that during this moment of regained love, Ewald renounces not just Hadaly but also Edison: "I must have been under a spell […] though he's a good man and a wonderful scientist, *he simply over-persuaded me* [my emphasis]" (192). When it is revealed that the person sitting next to him is not Alicia but Hadaly, he eventually accepts Hadaly/Sowana with the accompanying phrase: "I resign from the human race" (204). Now, accepting Hadaly, Ewald has simultaneously found a soulmate but resembles Edison in the process.

Sailing away on the steamer called *Wonderful,* the catastrophe at sea reveals Ewald's lack of regard for Alicia's life, as he only attempts to save Hadaly; both Alicia and Hadaly are lost. Ewald, like Edison, is left in complete isolation. His resignation from the human race is fulfilled. His preference for human-robot interactions at the expense of human-human interactions is revealed to be the result of a socio-scientific fantasy, engineered by Edison who lacks sympathy, love of humanity, sincerity and self-awareness. While the novel does not deny the potential of science and technological innovation, it does condemn the lofty promises that seem too good to be true (as the ship's name suggests, wonderful). Should humanity, like Ewald, blindly follow these paths to paradise as engineered by Edison, then humanity is doomed. Ewald as a symbolic representation of humanity is rescued by the French packet called *Redoubtable*. While Ewald allowed himself to be persuaded, or over-persuaded as he states, what is required is a more formidable defense against the wonderful promises of science as a means of improving upon human-human relations.

Chapter 9: Rossum's Universal Robots

Karel Čapek's *R.U.R. (Rossum's Universal Robots)* (1920) is accredited for introducing the word "robot" into the English lexicon, from the Slavic word *robota*, meaning "'drudgery', derived from its medieval sense of the unpaid labor a vassal was obliged to perform for his feudal lord" (Kang 279). Čapek would later explain that his brother Josef suggested the word to him. When Čapek told him he wanted to call his entities "laborators, but it seems to me somewhat stilted", Josef suggested to call them "robots" (Capek *Believe in People* 103). A 1935 Russian film, entitled *Gibel Sensatsii* (translated as *Loss of Sensation*), directed by Aleksandr Andriyevsky, contains a robot with the initials "R. U. R." on its chest. The film is described as a cinematic version of Čapek's play (Frakes 198), but it deviates considerably from Čapek's play. Romain Rolland's screenplay *The Revolt of the Machine* (1921) which was never filmed also deals closely with similar themes as Čapek's play. According to Kang: "Rolland portrays the disaster brought on by the revolt of the machines as the manifestation not of a new mechanical consciousness but of the primordial human urge to fight, destroy, and dominate" (285).

Parrinder explains that during a "public meeting in London" in 1923, where the play "was discussed by Bernard Shaw, G. K. Chesterton and others" (149), Čapek explained that he was "much more interested in men than in Robots" (qtd in Parrinder 149). Čapek "wished to write a comedy, partly of science, partly of truth" (qtd. in Klíma xiii). The comedy of science for Čapek is the double-bind of industrialism, which "must not stop, for if it does it would destroy the lives of thousands," and also that it "must, on the contrary, go on faster and faster, even though in the process it destroys thousands and thousands of lives" (qtd in Klíma xiv). This double-bind, according to Čapek, is comedic because a "product of the human brain has at last escaped from the control of human hands" (qtd. in Klíma xiv). Another underlying idea is Čapek's pragmatism, pertaining to the comedy of truth. Čapek argues that:

> The most dramatic element in modern civilization is the fact that one human truth stands against a truth no less human, one ideal against another ideal, one positive value against a

> value no less positive, and that the conflict does not represent, as we are often told, a struggle between a noble truth and vile, selfish evil. (qtd. in Klíma xv)

Klíma explains that Čapek emphasizes the notion that "Everyone has his own truth" (182). Any literary analysis of *R.U.R.* must address Čapek's own explications about the play. Klíma, however, does point out that "Čapek's interpretations [of his own play] cannot always be relied upon as fully trustworthy explanations" (xiii). This is because critics generally considered his "fantastic themes" and style of writing as "second-rate creativity," leading Čapek to inevitably emphasize "the philosophic or noetic side of his work" as a defense (Klíma xiii).

In the context of this dissertation, *R.U.R.* is important because it features a race of machines rather than a single robotic character. A race of machines accord logically with Čapek's own words concerning *R.U.R.*: "human heroism is among my beloved ideas and lured me to this material. And no matter what seems to flow from the finished work, I remained faithful to this idea … I wasn't concerned about robots, but about people" (qtd. in Klíma xvii). It is no coincidence that Čapek's ambition to write about humans and humanity caused him to employ robots for the purpose. A race of machines conjures up themes of group selection, which in turn invokes themes of altruism. While his audiences may have been fascinated by the idea of artificial humans, it is important to note that "Čapek himself designated the play as a collective drama", by which he meant plays "in which the protagonist was some collective, usually fighting for its rights" (Klíma xviii). Discussing *R.U.R.* in relation to Vinge's notion of the Singularity and the (technological) post-humans that will follow, Patrick Parrinder similarly considers *R.U.R.* important to revisit, as he points out: "Vinge's post-human Singularity, however useful it may be as a thought experiment, remains to all intents and purposes inexpressible" (148). Darko Suvin's discussion of *R.U.R.* similarly emphasizes the play's ambivalent ending: "But the end of the play, the robots again grow more like a new human order than like inhuman aliens, more like workers than machines; reacquiring pain, feelings, and love, they usher in a new cycle of creation or civilization" (272). This chapter focusses on the group dynamics and human-human and human-robot interactions as means of uncovering themes of altruism.

Before turning to a summary of the play, it is also worth mentioning that Kang discusses the important socio-historical factors surrounding the play. Three factors are listed as important, namely World War I, Taylorism and the Bolshevik revolution. First, World War I forever changed public perception of technology, obtaining new pejorative connotations as destroyers of humanity at unprecedented scales. Most works concerning automata published during the 1920s take place in factories and feature similar themes: "themes of exploitation, dehumanization, and class conflict" (Kang 267). World War I, with its introduction of tanks and airplanes, meant that soldiers were fighting "monstrous autonomous machines" (Kang 268). While Taylorism was up for debate prior to the war, it became a necessity during the war: "So the war not only turned soldiers into machines […] it also took a further step in turning workers into automatic devices whose monotonous movements were to be constantly timed and measured" (270). While Marxists "were ambivalent" towards Taylorism under capitalism, the Bolshevik leaders viewed it as liberating: "with Lenin asserting that the method was […] positive in a socialist setting since it increased productivity and shortened the duration of necessary labor" (Kang 271). As Klíma explains, however, while the revolutionary events in Russia influenced a "sizable part of the European intelligentsia", "Čapek's conclusions from his wartime experience were ambivalent" (ix). Suvin's discussion on Čapek's ambivalence towards communism serves to reveal Čapek's own position:

> Therefore, he defended himself as a kind of ethical socialist: 'I believe in the socialization of means of production, in the limitation of private property, in an organization of production and consumption, in the end of capitalism, in the right of each to life, work, sufficiency and freedom of mind, I believe in peace, solidarity and equality of the peoples, I believe in humaneness and democracy, and in man, Amen.' (qtd. in Suvin 281)

Kakoudaki notes the prevailing influence of the revolution in Russia on *R.U.R.* when writing: "[Čapek's play] responds directly to this historical environment, making direct allusions to socialism, militarization, capitalism, and industrialization" (134). The success of the play, particularly in the

U.S., also meant the quick succession of other plays with similar themes and concerns.[31] While it is true that Čapek's play is a direct response to these socio-historical developments, this dissertation examines the humanist dimensions of his play, to investigate how Čapek dealt with the theme of altruism.

Figure 6 The Robots Revolt

The play takes place at Rossum's Universal Robots factory situated on an island. The factory produces affordable robots for sale worldwide. Domin, the manager of R.U.R., entertains an important guest visiting the factory, namely young Helena, daughter of the president. It soon turns out that Helena has ulterior motives for visiting, namely to incite a revolt among the robots. She represents the *League of Humanity*, an organization dedicated to liberating robots. Between Act I and Act II, ten years pass, and Helena and Domin are married. By this time, the world's economy has grown dependent on robotic labor and, mysteriously, humans are increasingly becoming infertile.

[31] "Proud of their achievement in in bringing *R.U.R.* to American audiences, the directors of the Theatre Guild sought new writing that would continue to address questions of technology and modernity. In 1923, they followed the long run of *R.U.R.* with Elmer Rice's *The Adding Machine* (1923), in which clerk Mr. Zero is cruelly replaced by a simple adding machine, his supreme adding skills made irrelevant after twenty-five years of work at the same company" (Kakoudaki 143).

Helena, still sympathizing with the plight of the liberation of robots, decides to burn the manuscripts containing the secret formula for creating artificial life. Slowly but surely, the robots begin to revolt, and their revolution reaches the island factory.

During Act III, Helena admits that she destroyed the manuscript, which was the only form of leverage the last surviving humans on the island would have had. The revolt is swift, and everyone is killed by the robots except for Alquist, the chief engineer, who has been spared because he performs manual labor like the robots. During the Epilogue, some years later, Alquist has been given the task of recovering the lost secret formula. However, Alquist explains that it is impossible for him to do so. The robots, desperate to find a means of procreation, permit Alquist to kill and dissect robots to aid his research, but Alquist admits it is of little help. However, Alquist does notice two particular robots, called Primus and Helena (named after the real Helena), who appear to be in love. Playing a hunch, Alquist tells Primus that he must dissect Helena, and Primus refuses, offering himself up instead. Alquist then turns to Helena, telling her that he must dissect Primus, and she also refuses and offers herself up instead. Realizing that these robots are in love, Alquist rejoices, having discovered a new Adam and Eve, and fittingly tells them to go forth and multiply.

From this perspective, it becomes clear that *R.U.R.* has many of the implicit themes of group selection and altruism. As Klíma observes: "*R.U.R.* was a remarkable attempt to present to an atomized world a vision of mankind as a totality— to portray the destructive tendencies in man's behavior, to depict the basic discords in all their magnitude in a form simplified beyond the capabilities of realistic drama" (xxii). While there is much left to be discussed about the fact that the play is a play and not a text, in the context of this dissertation, it is mainly through a close-reading perspective that the themes of altruism come forward. However, this is not to deny the insightful analyses that Kakoudaki provides regarding the performative aspects of the play.[32] What is of

[32] "There are many ways to render Robots as 'other,' but they can never be 'other' enough. And this how 'metalface' works: the more metallic the Robots are, the more they allow the underlying debate about the human to remain free-floating and multivalent. The alienation of the human form on stage enacts a form of distance that makes the allegorical content of the play ever more adaptable for new audiences and new fears"

interest here is the manner in which the themes of the play come forward through the interactions of characters and whether their behaviors reflect group selection dynamics.

What follows below is an analysis of three particular dimensions of the play. First, the passages depicting the creation of the robots and their revolt are discussed. Second, the reasons underlying the robots' revolt in conjunction with the character of Helena are examined. Lastly, the role reversal between humans and machines at the end of the play in group selectionist terms, and its implicit connection to themes of altruism are discussed.

The Creation of Robots

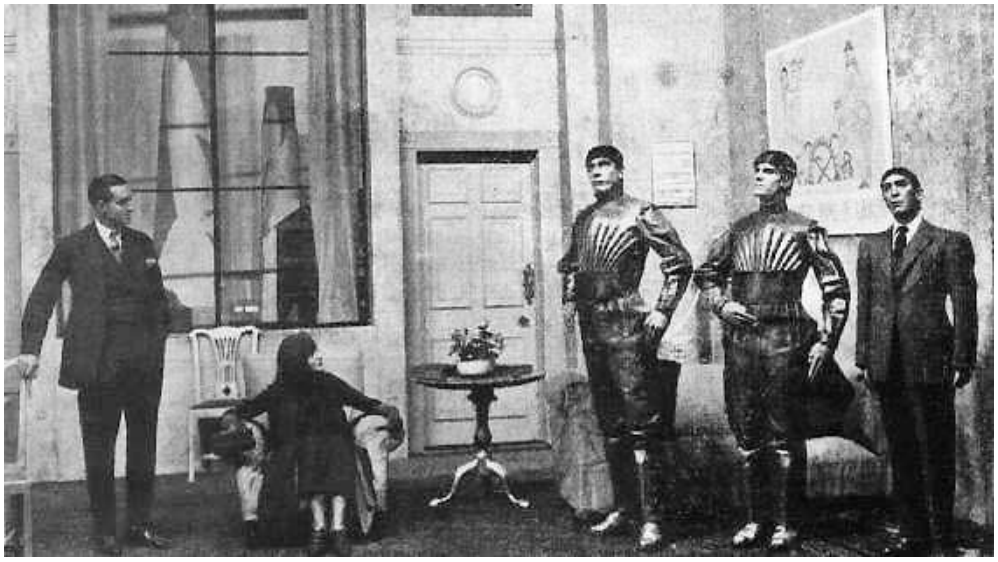

Figure 7 R.U.R. on Stage

While posters advertising the play placed "the action of the play around the year 2000" (Klíma xviii), the prologue opens in some distant future. The stage instructions specify that "on the wall to the left are big maps depicting ship and railway lines, a big calendar, and a clock […] affixed to the wall on the left are printed posters" which read: "The Cheapest Labor: Rossum's Robots"; "New-Tropical Robots-$150"; "Buy Your Very Own Robot"; "Looking to Cut Production Costs? Order Rossum's Robots" (3). Domin is dictating to Sulla, a female robot, who records his words on a typewriter. His

(142). It is also interesting to note Kakoudaki's observation that the appearance of the robots changed towards more metallic representations around 1927-8 in the U.S. (141).

opening words defends his company's liability for damage to robots during transportation: "Just before loading we brought it to the attention of your captain that the ship was unfit for the transportation of Robots, so we cannot be held financially accountable" (3). As a brief moment of foreshadowing, Domin's first action in the play is to denounce responsibility.

Sulla's introduction is the audience's first encounter with the play's robots. Čapek instructs all his Robots to have a "laconic" (2) predisposition, their faces "expressionless and their eyes fixed" (2) – prefiguring the quintessential 1950s robot stereotype. Sulla exits as the character of Miss Helena Glory enters. In a humorous but important exchange, their dialogue is filled with potential ulterior motives:

> HELENA: Why don't you let me finish my sentences?
>
> DOMIN: I beg your pardon. Perhaps you wanted to say something different?
>
> HELENA: I only wanted to ask—
>
> DOMIN: — whether I wouldn't make an exception and show you our factory. But of course, Miss Glory.
>
> HELENA: How did you know that's what I wanted to ask?
>
> DOMIN: Everyone asks the same thing.

Domin's opening question is entirely formalistic and rhetorical given that he already knows how he can help her and how the rest of their conversation will unfold. It appears that Domin has had a thousand such conversations in the past. On another level, the narrative demonstrates that human dialogue and language more generally has become entirely formalistic, rigorous, and no longer has any potential for creativity.[33] Dialogue has become mechanized. Domin's impatience is symptomatic of an industrial milieu whose central concern is to economize everything, even human interlocution. Additionally, Domin's sincerity becomes questionable when he claims that an exception will be made for Miss Glory, and this notion is reinforced by the end of the prologue when he asks her hand in

[33] Although depicted humorously in *R.U.R.*, the implications of the contrivance of language might be a foreshadowing of New Speak in Orwell's *1984* (1949).

marriage. When he does so, Helena points out that there must have been a thousand other women who have visited in the past, and he probably tried to marry them all. While Domin denies such accusations, one cannot help but suspect that he gives the same spiel to every (female) VIP who visits R.U.R.

As Domin explains, the year was 1920 when old Rossum, the father, a great philosopher and young scholar, came to the island to study marine life. At the same time, he attempted to reproduce living matter by means of "chemical synthesis," a matter known as "protoplasm" (6). Jeffrey Allan Johnson (also discussed in Chapter 6) points out the connection to Frankenstein's creature which Johnson argues to also be a chemical creation through spontaneous generation: "The idea of a corporate Frankenstein is much older, however, going back to [*R.U.R.*] […] the corporation founded by Rossum's nephew mass-produces the living robots that will fulfill Victor Frankenstein's worst nightmare by exterminating humanity" (305-6). In 1932, old Rossum stumbled upon a substance with a "different chemical composition" but which behaved exactly like living matter (6). In old Rossum's own words:

> Nature has found only one process by which to organize living matter. There is, however, another process, simpler, more moldable and faster, that nature has not hit upon at all. It is this process, by means of which the development of life could proceed, that I have discovered this very day. (6)

As Domin explains, one should consider the monumental implications of the historic moment when old Rossum was "sitting over a test tube and thinking how the whole tree of life would grow out of it" (6). After this discovery, old Rossum (having refined his work) was in a position to create whatever he pleased, including "a jellyfish with a Socratic brain or a one-hundred-fifty-foot worm" (6). According to Domin, because old Rossum did not have "a shred of humor about him", he set out to create a human (6-7). However, after a number of years, old Rossum managed to produce a dog, and sometime after that, a "mutant calf that died in a couple of days" (6).

Domin goes on to explain that old Rossum's ambition was to overthrow God: "he wanted to somehow scientifically dethrone God. He was a frightful materialist and did everything on that account. For him, the question was just to prove that God was unnecessary" (7). Domin explains that old Rossum set out to create an exact human facsimile, "right down to the last gland […] even the sexual organs" (7). Domin and Helena speculate about the arbitrary nature of giving the organisms sexual organs, because if they can procreate themselves, it would defeat the purpose of actually fabricating these artificial beings in factories in the first place. Suffice it here to say already that the sexual organs will come to play a significant part in the epilogue, as well as reiterating that these robots do/can constitute their own species.

The first human that old Rossum created lived for three days. When young Rossum came along, he felt that his father's progress was far too time costly, and claimed: "This is nonsense! Ten years to produce a human being?! If you can't do it faster than nature then what's the point?" (8). Domin explains that the old Rossum was "well suited to the university, but he had no sense of factory production […] It was young Rossum who had the idea to create living and intelligent labor machines from this mess" (8). Young Rossum, having familiarized himself with human anatomy, discovered that its complexities could be simplified by engineering. The thinking of young Rossum was much more in line with industrialism:

> young Rossum said to himself: 'A human being. That's something that feels joy, plays the violin, wants to go for a walk, in general requires a lot of things that— that are, in effect, superfluous.' […] A gasoline engine doesn't need tassels and ornaments, Miss Glory. And manufacturing artificial workers is exactly like manufacturing gasoline engines. (8)

Domin then demonstrates his theory by asking Helena what the qualities "best kind of worker" would possess, to which Helena mentions honesty and dedication. Domin corrects her by saying: "No, it's the one that's the cheapest. The one with the fewest needs" (9). This was young Rossum's greatest achievement – creating cheap laborers. In order to realize his vision, Domin explains:

> [young Rossum] chucked everything not directly related to work, and in so doing he pretty much discarded the human being and created the Robot. My dear Miss Glory, Robots are not people. They are mechanically more perfect than we are, they have an astounding intellectual capacity, but they have no soul. Oh, Miss Glory, the creation of an engineer is technically more refined than the product of nature. (9)

Other interesting corrections during this dialogue concern the purchasing of robots, as Helena states: "I saw the first Robots back home. The township bought them… I mean hired" (9). Domin corrects her: "Bought, my dear Miss Glory. Robots are bought" (9). Domin also points out (interestingly in the context of this dissertation) that the robots are not uniform: "You see, Rossum's Universal Robots Corporation does not yet manufacture entirely uniform goods. Some of the Robots are very fine, others come out cruder. The best will live perhaps twenty years" (9). Helena asks whether they then "die," to which Domin explains "Well, they wear out" (9).

R.U.R. is also a post-scarcity science fiction text. It deals not merely with the concern that humans are being turned into robots under the forces of industrialism, but also with the more problematic notion that Domin's views are in fact realistic and correct when thinking about robots as actual machines instead of workers. This is a much more interesting perspective on the double-bind of industrialism: industrial capitalism is premised on cycles of production and consumption which are founded on the problems of scarcity. It is this very scarcity that robots help to eliminate, which simultaneously makes them profitable but also threatening: a robot is the progeny of scarcity-driven markets and even financially alluring to create within such a marketplace, yet they also pose the greatest threat to free-market capitalism. Domin is enthusiastic about their development as he stands to gain, while ironically he should have foreseen that robots also portend a revolutionary end to industrial capitalism itself. If not through a violent revolt (which may not be easy to predict), then Domin should have at least realized that robots will mean the obsolescence of the current economic model from which he profits.

It is clear that there are two kinds of robots in this text. On the one hand, there are robots that are genuinely mechanical tools and, on the other hand, robots that approximate the human condition. Young Rossum created robots that function as mechanical tools that do not culminate into anything more than glorified household appliances. Domin instructs Sulla to pose for Miss Glory so she can investigate it. The dialogue between Sulla and Helena illustrate how jargons apply to their different conditions: Helena asks in another Turing test moment whether Sulla was "born" at the factory, to which Sulla replies "I was made here, yes" (10). While Domin displays Sulla with pride, Helena finds the encounter awful, exclaiming: "Please stop!" (10). Sulla's natural appearance causes Helena to suspect that Sulla and Domin are playing some kind of trick on her. In order to convince her, Domin arranges that Sulla is to be dissected. Helena refuses to go, and she even hugs Sulla claiming: "Don't be frightened, Sulla. I won't let them hurt you!," to which Sulla responds: "I am a Robot" (11). Helena, posthumanist in her outlook, claims "That makes no difference. Robots are every bit as good people as we are" (11). Asking Sulla whether she is aware of what would happen if they dissect her, Sulla explains "Yes, I would stop moving," and Helena finds the prospect "dreadful!" (12). However, Helena's sympathy towards robots is misplaced and in some sense naïve, precisely because these robots are merely machines. Another robot, Marius, also enters, and vouches for Sulla's points of view. He too explains that he does not understand the concept of feeling "sorry" for someone, and that he does not fear (nor understand) the concept of death (12). Domin points out that all the laborers Helena had encountered on the island, from bricklayers to office workers, are machines. They do not cling to life. He even explains that he will show her the kneading troughs, giant vats "for the batter" of which they make roughly a thousand robots, including vats for "livers, brains, etc." as well as the "bone factory" and "spinning mill for nerves [and veins]" (13). All these constituent parts are then assembled in an assembly plant, "like automobiles" (13).

At this point, the other characters enter, named Dr. Gall, Dr. Hallemeier, Fabry, Alquist and Busman. During their exchange with Helena, it soon becomes clear that Helena makes another mistaken assumption when she assumes all these characters to be robots themselves, another

Turing test moment. Thinking them to be robots, she confesses her true intentions to incite a revolt among the machines, and after Domin points out that they are men, she feels ashamed and embarrassed. However, the company explain that they will not send her away, despite her intentions, because she is merely the last in a long line of "saviors or prophets […] missionaries, anarchists, the Salvation Army, everything imaginable," all of whom have tried to do the same, but have been unsuccessful (17). They tell her that she is free to preach whatever she pleases to the robots. After some enquiry, Helena explains that her League of Humanity "want first and foremost to protect the Robots and […] to guarantee them – good treatment" (18). The company agree that those are indeed good values, as Fabry states: "I don't like damaged goods. Please, Miss Glory, enlist us all as contributing, dues-paying, founding members of your League!" (18). Helena tries to correct him: "you misunderstand me. We want – specifically – to liberate the Robots […] They should be treated like […] people" (18). Hallemeier, trying to better understand Helena's views, asks whether robots should receive salaries, and/or be allowed to vote, to which Helena claims: "But of course they should!" (18). Hallemeier points out that robots would have no use for money, as nothing makes them happy. He explains that they love nothing, not even themselves; they have no will of their own, "no passion, no history, no soul" (18). However, when Helena enquires about "love or defiance," Hallemeier pauses, and explains that:

> Occasionally, they go crazy somehow. Something like epilepsy you know? We call it Robotic Palsy. All of a sudden one of them goes and breaks whatever it has in its hand, stops working, gnashes its teeth […] Evidently a breakdown of the organism. (18)

Helena claims those moments demonstrate that a soul is somehow present. However, the company assure her that sometimes products are just faulty. Indeed, at this stage of the play, concerning these first machines, Helena's views of liberating robots seem to be misguided. Robots are indeed fantastic sources of low-cost altruism precisely because they are not human.

It is only by the second and subsequent generation of robots that Helena's views seem to become more fitting because these robots begin to approximate the human condition. The

discussion turns to Dr. Gall's current research into "pain-reactive nerves" (19). Dr. Gall explains that the young Rossum oversimplified the nervous system, which was not good, as "We must introduce suffering" (19). Helena, naturally horrified by the notion, wonders why on earth robots must be able to feel pain. Dr. Gall explains that the robots are prone to injury as a result of carelessly going about their work, often breaking fingers and crushing their own hands. Pain will make them more efficient. Helena responds: "Why won't you make souls for them?" (19), to which Dr. Gall, Fabry and Busman explain that it is not within their power nor interest, as it would raise the cost of production. Busman explains that the beauty of their machines is that "it's so cheap!," and prices for clothing (for example) are still falling as a result. Indeed, factories all over "are going belly-up unless they've bought Robots to cut production costs" (19). Busman claims that within another five years, "everything will cost a tenth of what it costs now" (20). Thus, one discovers the promise of a post-scarcity society, yet none of the characters mention the implicit fact that such a society would also mean the end, or a significant redefinition, of capitalism itself. After ten years, Alquist points out, "all the [human] laborers of the world will be out of work," but Domin quickly replies:

> Yes they will […] But within the next ten years Rossum's Universal Robots will produce so much wheat, so much cloth, so much everything that things will no longer have any value. Everyone will be able to take as much as he needs. There'll be no more poverty. Yes, people will be out of work, but by then there'll be no work left to be done. Everything will be done by living machines. People will do only what they enjoy. They will live only to perfect themselves […] then the subjugation of man by man and the enslavement of man by matter will cease. (20-1)

In these words, Domin begins to resemble the Professor's character in *Tomorrow's Eve* when his justifications seem to be at odds with his actions. The industrialist who is obsessed with efficiency, low-cost and profit, describes a utopian society in which any monetary system would essentially be superfluous. He ironically prophesizes his own end in a celebratory fashion. Alquist is the traditionalist, or Tolstoyan, who disagrees with Domin's views; "there was something good in the act

of serving, something great in humility. Oh, Harry, there was some kind of virtue in work and fatigue" (21). Domin agrees that there was something good about it, but that:

> we can't exactly compensate for everything that's lost when we recreate the world from Adam. O Adam, Adam! No longer will you have to earn your bread by the sweat of your brow; you will return to Paradise, where you were nourished by the hand of God. You will be free and supreme; you will have no other task, no other work, no other cares, than to perfect your own being. You will be the master of creation. (21)

What Domin describes as a form of transcendence is in fact a form of death. His faith in machines is one by which humanity will unquestionably evolve to a point of mastering all creation, but it is a dogmatic belief which he pursues in conjunction with financial gain. He sees a world filled with individuals free to perfect themselves, while Alquist sees humanity devoid of any real purpose very similar to the condition of the robots. Alquist more correctly identifies the spiritual impoverishments or oversights in Domin's ideal humanity, namely humility. Humility in this context is the recognition that one exists between two poles, existing on a systems theoretical edge of chaos of competing forces. Mastering creation eliminates such an understanding of the importance of dialectic development of humanity as a species. Helena correctly realizes that *souls* must be saved. However, given that she frequently confuses robots with humans and vice versa, it is clear that she is misguided as to whose souls ought to be saved. As the Turing test implies, Helena should have realized that it is their own souls that required saving, not those of the robots.

The Robotic Revolution

In Act I, taking place ten years after the Prologue, one expects to find Domin's utopian vision to have come to fulfillment, only to discover that quite the opposite is taking place. Humanity is becoming sterile for unknown reasons, although Alquist suspects it is because humans are devoid of purpose. Helena also suspects that Domin (and the others) are hiding something from her, and indeed they are hiding the fact that robots have begun their revolt.

There is an interesting scene between Helena and Alquist, as she is slowly beginning to be influenced by his philosophical worldviews. She admires Alquist's "rough, dirty" hands (32), noting an implicit connection between Alquist and the robots. Alquist explains that he is a reactionary, and that he believes laying bricks is good for the soul: "I think it's better to lay a single brick than to draw up plans that are too great. I'm an old man, Helena; I have my hobbies" (34). When Helena asks Alquist why women stopped having babies, he explains: "Because it's not necessary. Because we are in paradise, understand?" (35). In a short diatribe, Alquist explains that Domin's vision has become a reality; all means of scarcity and all requirements for work have been eliminated. The result is that:

> we people, we, the crown of creation, do not grow old with labor, we do not grow old with the cares of rearing children, we do not grow old from poverty! Hurry, hurry, step right up and indulge your carnal passions! And you expect women to have children by such men? Helena, to men who are superfluous women will not bear children! (35)

When she asks whether humanity will become extinct, Alquist states: "It will. It must. It'll fall away like a sterile flower" (35). After leaving, Helena looks at the beautifully arranged flowers in her room, and exclaims: "Oh, flowers, are there sterile ones among you as well? No, there can't be! Why then would you bloom?" (35-6). The irony here is that the flowers she is referring to were actually engineered by Dr. Gall, who later affirms that they are sterile. On the one hand, being sterile, not having to work, means that a flower can bloom in full, and contains its own inherent beauty for beauty's sake. It is the potential of transcendence or salvation of the industrial human. On the other hand, Alquist's concern is similarly that one is dealing with pseudo-transcendence or damnation, and the implicit horrific truth of such sterile flowers lies in their artificiality, making them entirely superfluous. Here, Čapek is playing with mortality and immortality, salvation and damnation; ironically, an eternal return to paradise would also mean an eternal state of damnation, devoid of strife, struggle and triumph.

Shortly after, Helena, upon hearing that another robot has gone mad, insists on seeing the robot, which is named Radius. While Helena tries to sympathize with the robot, Radius merely

wishes to be sent to the stamping-mill (its death). After a brief inspection, Dr. Gall concludes that Radius' case is not a straightforward instance of ordinary Robotic palsy, but there is something unknown brewing inside Radius, "Defiance, rage, revolt – I haven't a clue" (37). It is then revealed that Dr. Gall started to make a different kind of robot, and these are more "irascible" than the traditional Rossum's Robots: "They're more like people than Rossum's Robots are" (38). In a moment of important foreboding, Helena asks "Is this . . . this hatred of theirs another human characteristic, perhaps?" to which Dr. Gall replies, shrugging his shoulders: "Even that's progress, I suppose" (38). The irony is that Dr. Gall's experiment is a success if the aim is to produce robots with human qualities. However, this is the dividing line between Helena's and Domin's views in the Prologue: the beauty of machines was the fact that they were *not* like humans, and therefore did *not* represent any suppressed working class. Helena's sympathy turns the machines into human working classes.

Helena decides to burn the manuscript which contains the secret formulas of how robots are created. Upon hearing that the robots are revolting, she grows anxious. However, Domin has his own solution to the problem, imitating God and the destruction of Babel:

> We'll open a factory in every country, in every state, and can you guess what these new factories will produce? […] National Robots. […] It means that each factory will be making Robots of a different color, a different nationality, a different tongue; they'll all be different— as different from one another as fingerprints; they'll no longer be able to conspire with one another; and we— we people will help to foster their prejudices and cultivate their mutual lack of understanding, you see? So that any given Robot, to the day of its death, right to the grave, will forever hate a Robot bearing the trademark of another factory.
>
> HALLEMEIER: Thunder, we'll make Black Robots and Swedish Robots and Italian Robots and Chinese Robots, and then let someone try to drive the notion of brotherhood into the noggin of their organization. […]

DOMIN: Helena, just to keep mankind at the helm for another hundred years […] (46)

Not only has Domin's original prophecy of achieving utopia after ten years become a hundred, his solution to stifling the revolt is to endow machines with human-social attributes as a means to divide and conquer. Domin, quite rightly, speculates that introducing *national* divisions in the *universal* conditions of robots would result in their international bickering, removing the intergroup or interspecies threat towards humanity. However, Domin's proposed solution comes too late, as the robots are already revolting and already constitute their own group. They even publish and disseminate their own propaganda, communicating (and therefore evolving) amongst themselves. Domin reads one such pamphlet, and explains that robots "are higher than man on the evolutionary scale," and that humanity "lives off them [robots] like a parasite" (48). In these depictions it becomes clear that Domin's parable of nationalizing robots is the very human shortcoming of banding together at a planetary level. The robots are, as the play's title already states, *universal* and denote a larger and more complex group than humanity.

Group Selection and Altruism

The robots plead with Alquist to figure out the formula and/or to hand them the manuscript so that robots can continue to create more robots. However, Alquist points out (numerous times) that he does not know the formula, and that the manuscript has been destroyed. The robots plead that he should dissect one of them, even while alive.

Two robots, Primus and Helena, were the result of Dr. Gall's experiments, at behest of real Helena to create more humanlike robots. Alquist notices an affection between these two robots, and given that the machines have asked him to dissect some robots to try and uncover the secret of artificial life, Alquist decides to play a hunch. He proposes to dissect robot Helena, to which Primus pleads to spare Helena's life. When Alquist proposes to dissect Primus, robot Helena in turn pleads to spare Primus' life. In both cases, Primus and Helena are willing to self-sacrifice so that the other robot may live, and when Alquist asks why neither of them may be dissected, Primus states: "We-we-belong to each other" (83). Alquist rejoices, "O hallowed sixth day!" (83). He tells them: "Go,

Adam. Go, Eve-be a wife to Primus. Be a husband to Helena, Primus" (83). In this sense, robots do inherit the legacy of humanity, only not in the form they desired. At a group-level, survival depends on altruism and these machines have inherited the ability to be altruistic:

> O nature, nature, life will not perish! Friends, Helena, life will not perish! It will begin anew with love; it will start out naked and tiny; it will take root in the wilderness, and to it all that we did and built will mean nothing— our towns and factories, our art, our ideas will all mean nothing, and yet life will not perish! Only we have perished. Our houses and machines will be in ruins, our systems will collapse, and the names of our great will fall away like autumn leaves. Only you, love, will blossom on this rubbish heap and commit the seed of life to the winds. (84)

This is not merely a cyclical ending, in the sense that the entire comedy of human affairs is doomed to repeat, but a real celebration at having discovered a key question in the play regarding human evolution, perfectibility and artificial life.

Alquist's celebration is a celebration of discovering the complex property of evolution. In this narrative, evolution does not select on the basis of one's existentiality but behavior; whether human or robot, only those capable of expressing love are permitted to live, namely Alquist, robot Helena and Primus. The industrial robots which exterminated humanity still depend on a great truth, namely the secret formula which was lost when Helena burned the manuscript. As such, these robots are doomed. In *The Robot Condition: Karel Čapek's R.U.R. and Hannah Arendt on Labor, Technology and Innovation* (2017), Matthew Dinan also discusses the theme of love:

> Love, then, works against the world alienation endemic to modern technology through the objectivity found in – and facilitated by – the human condition of natality. By ending the play as he does, Čapek seems to suggest that this world-redeeming character of love can run counter to the alienating potential of technology by giving us an encounter with natality and through it a sense of what genuine newness means for the human condition. (115)

The theme of love which "works against" the "alienation" of "modern technology" has deeper evolutionary implications as well. Love is not merely a force capable of freeing us from oppressive industrial technologies, but also plays an integral role in human evolution (or evolution more generally).

Kathleen Richardson similarly highlights these themes in her conclusion to her book. Richardson's concern about the loss of human-human interaction at the hands of technologically mediated interactions, what she refers to as attachment wounds, can be easily positioned alongside Dinan's claim above (131). Richardson points out that in the 1920s violence predominated, "either from the battlefields or from revolution and political violence" (132). Rather than viewed as a commentary on these developments, Richardson reads *R.U.R.* as a critique on violence itself: "Čapek critiqued and rejected that that violence could bring about meaningful change – it can only lead to complete annihilation for humanity. Čapek's message is that any system that is built on violence can only end in violence: *it annihilates itself*" (132). Richardson argues that the end of the play, concerning two robots falling in love, denotes hope: "Even in cases of the most serious human attachment difficulties [i.e. human-human interaction], the attachment wound can be repaired by human love […] Love begets more love" (132). This dissertation agrees with both Richardson and Dinan's conclusions while pointing out that this particular theme of love is expressed in the play through altruistic gestures; their love is affirmed when each robot is willing to be self-sacrificed for the other.

R.U.R. exemplifies the notion that the important difference between humans and robots is not one of existential truths, but rather a question of behavior. In the final act, the play divides its characters between those who are selfish and those who are selfless, irrespective of whether they are human or robot. As mentioned in Chapter 1, Čapek's robots would more accurately be described as androids today given that they are made of synthetic organic materials. Thus, the difference between natural and artificial human is of secondary importance in the play, while the primary

difference is one of selfish and altruistic motivations and behaviors. In “Save Yourself If You Can” (1922), Čapek argues:

> It’s strange what a furious altruism there is in every salvationist faith. Everyone wants to redeem the whole world at all costs. He won’t do it any cheaper. […] I belong among those short-sighted creatures who are unable to see the wood for the trees, or humankind for actual humans. […] I can’t imagine a great idea that wouldn’t, besides having the miraculous power to unite people, also have the miraculous power to divide them. […] Mankind can’t be saved, but it is possible to help a man. Perhaps it is a low ideal, […] but how can one deliver the world if everything is not all right here at hand? (Čapek *Believe in People* 257-8)

As discussed in Part I of this dissertation, the Turing test is not a great truth, but a call for tolerance; La Mettrie argues that our mechanical constitutions do not negate our inherited sense of morality; as Richardson argues, in the age of social robotics, Čapek’s critique on violence and the importance of tolerance and love are only becoming more relevant.

Chapter 10: Metropolis

Fritz Lang's *Metropolis* (1927) is a classic science fiction film, and in 1994 was voted by 324 film critics as the eighth best German film of all time (McGilligan 480). The original version of the film was lost, which "can be said to have been presented only at the Berlin premiere and possibly at the Ufa-Pavilion an Nollendorf-platz – for no more than ten weeks [in 1927]" (McGilligan 133). Over the course of eighty-three years, many reconstructions were made yet none of them would achieve the original 4,189 meter-long version of the film (McGilligan 131). It was only in 2008 that a complete negative was discovered in Argentina which subsequently was released as *The Complete Metropolis* (2010) (Wosk "Update" 1061). Unlike some of Fritz Lang's previous films, Thea Von Harbou was fully credited for the screenplay, as the "screen credits make the point that the script is 'based on a novel by Thea Von Harbou'" (McGilligan 109). Von Harbou's novel was also translated into English in 1927.

Given the close affiliation of Von Harbou and Lang personally and professionally, analyzing either the novel or film in isolation becomes difficult. As Patrick McGilligan points out, in his biography *Fritz Lang: The Nature of the Beast* (1997), a newspaper article published on July 4, 1924, described that the renowned director and his wife were on holiday where they worked "'to finish the screenplay for their new film *Metropolis*'" (109). McGilligan explains: "July was three months before Lang went to America, and five months before he returned in December […] By then, Christmas of 1924, his hardworking wife had certainly completed the *Metropolis* scenario" (109). Lang's goals for the film were so ambitious that the production company funding the project, Ufa, "never expected to turn a profit," but rather "hoped that the colossal motion picture would recoup its investment and create an opening for future German films in America" (110). However, this hope was never realized. A major hinderance to the film's success was its notoriously convoluted narrative. Before releasing the film to American audiences, Paramount commissioned Channing Pollock to reedit the film and change the plot entirely as means of making it more accessible (McGilligan 132). While the Pollock version contains a different plot altogether, Lang's own version was also reedited a number of times in order to shorten the film's unusually long running time, as

McGilligan describes "the film was dictatorially and carelessly slashed everywhere after its Berlin premiere" (133). Two prominent critics of the film included H. G. Wells and Luis Buñuel, a Parisian film director. Wells stated: "I have recently seen the silliest film. I do not believe it would be possible to make one sillier", and went on to explain that the film contains "almost every possible foolishness, cliché, platitude, and muddlement about mechanical progress and progress in general, served up with a sauce of sentimentality" (qtd. in McGilligan 130). Buñuel described the film as a "technical perfection" but the narrative as "trivial, bombastic, [and] pedantic" (qtd in McGilligan 131). Not unlike *R.U.R.*, it was the film's ending that contributed largely to such negative reviews. The film ends with a handshake between Joh (the leader of Metropolis) and one of the workmen as a symbolic gesture of reconciliation between the industrialists and proletariat. However, as McGilligan explains:

> Critics then and later would seize on the handshake as one of the silly, or 'Kitsch,' elements of a film that failed to achieve intellectual depth. […] [Lang] often said in interviews that he had agonized over this particular ending, which, he always emphasized, was principally von Harbou's contribution. […] Late in life, as he softened, the director began to equivocate. Maybe the 'kitschy' message was valid, after all. […] 'I didn't think in those days a social question could be solved with something as simple as the line: The mediator between brain (capital) and hand (working class) must be the heart,' Lang was quoted in 1976, the last year of his life. 'Yet today, when you speak with young people about what they miss in the computer-guided establishment, the answer is always, The heart! So, probably the scenarist, Mrs. Thea von Harbou, had foresight, and therefore was right and I was wrong.' (127-8)

It is important to recognize that the film was a collaborative effort between Lang and Von Harbou. Being an adaptation of Von Harbou's novel, many scenes in the film were included without providing sufficient narratological context. The cinematic result therefore contains an impoverished narrative which fails to adequately explain certain characters' motivations or transformations. While the film *Metropolis* is a collaborative effort, this dissertation is concerned with Von Harbou's novel in order

to gain a better understanding as to why, in Lang's words, she "had foresight" and "was right" about the "computer-guided establishment" missing a heart.

The novel opens with an epigraph, also included in the film, containing a maxim that is reiterated throughout the narrative. The epigraph reads:

> This book is not of today or of the future.
>
> It tells of no place.
>
> It serves no cause, party or class.
>
> It has a moral which grows on the pillar of understanding:
>
> *The mediator between brain and muscle must be the Heart*.
>
> Thea Von Harbou (10)

The narrative takes place in the futuristic city of Metropolis, where society is divided between the affluent who inhabit skyscrapers and the workers who inhabit subterranean factories. The city is overseen by Joh Fredersen. His son, Freder, falls in love with Maria who wishes to incite a peaceful revolution. Fearing a revolution, Joh Fredersen employs the help of a mad scientist called Rotwang. However, the two men have a history given that Joh's deceased wife Hel (Freder's mother) was originally with Rotwang. Hel left Rotwang to be with Joh, and Rotwang still begrudges Joh for stealing Hel away from him. Rotwang reveals his latest invention to Joh, namely a female robot which he intends to model after Hel. However, at the behest of Joh, Rotwang captures Maria and instead models the robot after Maria's appearance. Given his grudge towards Joh, Rotwang secretly gives the robot ulterior motives. This robotic Maria incites a violent revolution among the workers, who then set out to destroy the city. In their rage, the mob destroys the heart machine, responsible for keeping Metropolis from being flooded. The mob fails to realize that their children, being underground, are at risk of drowning first. After escaping Rotwang, the real Maria rescues the children trapped underground. Once the angry mob realize that their own children are in danger, they turn on and burn robotic Maria at a stake (believing it to be the real Maria). After Maria has rescued the children, with Freder's help, both end up fleeing to the cathedral where they find

Rotwang. Eventually, Maria nearly falls to her death, hanging on to a statue, while Freder and Rotwang fight on the cathedral's rooftop. When the mob looks up, they see Maria and realize that they had been duped into revolution by a false prophet. Joh Fredersen rescues the real Maria, and Rotwang falls to his death. Afterwards, Joh Fredersen has learned his mistake of being a heartless industrialist, and now wishes to build a new Metropolis, one that has a heart that mediates between the head and the hands of the city.

Von Harbou's *Metropolis* depicts various moments of altruistic behaviors in conjunction with a broader concern for social cohesion and organization in the context of a technologically advanced civilization. What follows is a closer examination of the character of Joh Fredersen, the function of religious motifs, and finally the theme of altruism. While Freder appears to be the main protagonist, one should recognize that his character remains rather static throughout the narrative. By contrast, Joh's character undergoes a more pertinent transformation. The theme and function of religion and magic in the narrative represent humanity's (or society's) struggle towards obtaining a collective identity rather than dealing with actual religious values. Although there is only a single robot in the narrative, it quickly becomes clear that the robot is intricately connected to every member of Metropolis, and stands in opposition to the entire society. Although the robot is the most iconic image from the film, the character of the robot actually plays a rather small role in the novel's overall narrative. While *Frankenstein* managed to express group selection in evolutionary terms, *Metropolis* attempts to represent humanity as a complex adaptive system akin to a colony of ants. The narrative is very much concerned with the formation of group minds and social organization. *Metropolis* stands apart from other science fiction narratives examined in this dissertation in that it downplays the importance of the robot while emphasizing the importance of cooperation, altruism and group identity.

The Transformation of Joh

The robot in *Metropolis* is born from competing intentions and undergoes three phases of development. Rotwang, who lost his beloved Hel to Joh, originally created the robot as an attempt to overcome his grief. Joh, on the other hand, employs the robot to prevent a revolution. Finally, the robot is employed by Rotwang to enact revenge on Joh. In other words, the robot is created as an object of desire, then becomes a tool or medium with which to sway public opinion, and eventually is employed as a weapon. Although the robot was intended to have Hel's appearance, it is modeled after Maria who is Freder's love interest. The robot is a composite of three characters, and is a symbolic love-interest or temptress to each of the male characters. While Rotwang is able to control the robot, he could not control the real Hel, and while Joh managed to steal Hel away from Rotwang, he cannot control the robot. Freder is similarly duped by the robot during moments of mistaken identity.

According to Barbara Hales, in "Taming the Technological Shrew: Woman as Machine in Weimar Culture" (2009), the robot in *Metropolis* is an expression of the anxieties towards the role of the "new woman" in Weimar culture: "The term 'New Woman' (*die neue Frau*) found in the Weimar popular media, referred to the independent woman who was participating in the work force in large numbers" (302). Despite this emancipation, there was also a "fear of being incapable of controlling women," both domestically and in the workforce, and there "was thus a concern that women would supplant men in the workplace and reject the traditional roles […] in favor of free sexuality and masculine pursuits" (303). This cultural development went so far that sexual education reformers would facilitate such views: "The methodical process of having sex imitated the scientific management of industrialist production," and women "became themselves machines who could perform more productively" (305). As Hales points out: "a woman's body was prepared to work like a machine, namely a machine designed and operated by a male" (305). Hales manages to explain this dimension of the robot clearly and these anxieties certainly play a major role in the narrative. However, the theme of maintaining control is also central to Joh's character development. While Joh

managed to "control" or manipulate Hel into being with him, the robot, Rotwang, the city and Freder's relationship with Maria all escape Joh's control. Joh becomes the only character who learns to accept (and how to cope with) his inability to control others and the world. In the film, Joh remains a static character whose development is merely suggested rather than explained during the final handshake. In the novel, however, Joh's transformation is explained in more detail. A few important chapters in the novel were omitted from the film involving Joh's mother.

Joh's mother (who remains unnamed) reveals that Joh's greatest sin was to steal Hel away from Rotwang. Although Hel eventually loved Joh, Joh never repented: "I could not do without Hel, Mother. I would have stolen her from the arms of God himself" (144). However, Joh's Mother explains: "From God, Joh, you can steal nothing, but something can be stolen from man. You have done that […] Your sin walks behind you like a good dog on the trail" (144). Joh's mother calls Rotwang a "friend," whom Joh betrayed (144). Now that Joh hopes to intervene with Freder's love towards Maria, his mother points out the hypocrisy: "Freder is your son. What do you think, Joh, he would answer me were I to say to him: give up the girl you love…?" (144). Joh remains silent, and the narrative points out the mechanical nature of Joh's countenance: "I know what it means when your eyes grow cold, as now, and when you grow as pale as one of the stones on the wall. You have forgotten that lovers are sacred. Even if they are mistaken, Joh, their mistake itself is sacred" (144-5). Joh only affirms his own needs by saying he needs his son back. His mother points to a passage in the Bible: "*Whatsoever a man soweth, that shall he reap*" (145). This exchange foreshadows that the fate of Metropolis hinges on Joh's personal flaws and redemption. Joh's mother is a symbolic representation of his own conscience, and he disregards her advice. Joh's mother therefore acts as his conscience, and being the leader of Metropolis, his judgements will have direct implications or relations to the entire population of Metropolis. At this stage of his character's development, Joh represents an archetypal leader, but one who confuses authentic leadership with autocratic control. The city of Metropolis is an extension of his character, but Joh is unable to live according to La Mettrie's natural law of doing onto others as he would have them do unto him. By choosing

autocratic control over the more altruistic mode of leadership, Joh reflects the shift in "scientific" discourse from a focus on working with nature towards perfection, the magical/occult discourse present in the novel, towards the exploitation of nature, the urban-industrial setting of the novel with its exploitation of the only natural resource present: the human workers.

When Freder witnesses the hardships of the workers underground, he implores his father to do something about it: "Father! Help the men who live at your machines!", to which Joh responds: "I cannot help them. They are where they must be" (23). Freder exclaims that they are human, and Joh agrees: "Unfortunately, yes" (24). Freder explains that he feels a strong sense of empathy towards the workers: "Every single man, father, who slaves at your machines, has my face," to which Joh responds: "Then mine too, Freder, for we are very like each other" (25). While Freder is filled with horror, Joh simply states that "The time of horror lies behind me" (25). This exchange reveals that Joh, although ruling the city far above the underground factories, is not that different from the workers. His soul has been lost and he seems fully cognizant of the dire situation. As Freder points out, his father is complicit because he literally pushes the button every ten hours; the fact that the button must be pressed every ten hours implies that Joh is just as imprisoned or enslaved by the mechanisms of Metropolis as the workers are underground. This kind of drudgery renders the human laborers as little more than the robots from *R.U.R.*, especially so given that, during this exchange, Joh acknowledges the need for machine men to run the city instead of human laborers, which Freder also considers a better solution so long as those machine men do not have souls. In other words, it is an oversimplification to state that Joh is unsympathetic towards the workers; Joh recognizes the futility of the situation but is powerless to do anything about it. A later exchange with Rotwang reveals that Joh had originally intended to come up with a solution to the problem. When the robot is revealed to Joh initially, Joh tells Rotwang: "I ordered machine men from you, Rotwang, which I can use to work my machines" (52). This line indicates that Joh did intend to replace the working classes of Metropolis with machine men, presumably because such robots are more efficient, but it would also have alleviated the conditions of the workers.

Rather than insisting on his machine men, Joh is persuaded by Rotwang to use the female robot instead. Rotwang explains that his female robot can be used as an instrument of propaganda and is a superior tool than machine men: “Do you know what it means to have a woman as a tool? [….] Why do you fight with the Gothics and the monk Desertus […] Send the woman to them, Joh Fredersen!” (52). It is clear that Rotwang, although deranged, is selling Joh Fredersen a new form of control, but in reality it is a means by which Rotwang can control (or destroy) Joh. When analyzing the narrative from this perspective, the robot becomes a symbol of Metropolis itself. If Joh had remained firm in his stance of employing machine men, the outcome would have been different. Instead, whoever gains control over the robot also controls Metropolis.

The final chapter of the novel consists of another exchange between Joh and his mother (also omitted from the film). Joh, consistently referred to as the brain or head of Metropolis, embodies the superorganism or complex adaptive system of Metropolis. His transformation represents the transformation Metropolis must undergo if it wishes to survive the onslaught of industrialization. In the final chapter, Joh’s repentance is cathartic for him, and he confesses his sins to his mother while expressing his newfound realizations. He claims that when Freder spoke to him that morning, it was the first time he saw his son’s face, a conglomerate of all the people: “It is quite my face […] It is the face of his beautiful, dead mother […] yet it is, at the same time, fashioned after Maria’s features […] But it is, at the same time, the face of the masses” (236). A close reading reveals that this realization is not different from the initial exchange when Freder pointed out that his face is shared by the workers, to which Joh admitted that they share his face as well. The amalgamation of faces symbolizes the interconnectedness between the citizens of Metropolis and the formation of a group or unity. Joh’s mother now questions what is different about Joh’s empathy: “How do you come to know the face of the masses” (236). Joh responds:

> From the heights of the New Tower of Babel I could not distinguish it. And in the night of lunacy, in which I perceived it for the first time it was so distorted in its own horror that it no more resembled itself… When I came out of the cathedral door in the morning the masses

> were standing as one man, looking towards me. Then the face of the masses was turned towards me. Then I saw, it was not old, was not young, was without sorrow and without joy. 'What do you want', I asked. And one answered: 'We are waiting […] for someone to come, who will tell us what way we should go…'. (236)

The relationship between governance and humanity is one of distance and the ability to relate to others. From the height of his tower, the masses appeared to Joh as a complex hive, and despite knowing that he shared their faces, he always acted out of his own self-interest. Only when Joh was at ground level (leaving the cathedral) was he forced to witness their faces, ironically still "standing as one man." Having descended from the heights of his tower, Joh is forced to recognize that he is part of the society he governs. His new realization is not one of a particular political ideology, but rather one that accords with La Mettrie's natural law. Joh's mother asks him if he wants to be that man who leads the people, and Joh explains that he does:

> If we had been living a thousand years earlier, I should, perhaps, set out on a high road […] not returning home until I had cooled my feet, hot from wandering, in the Jordan, and, in the places of redemption, had prayed to the Redeemer. […] For, from comprehension comes love, and I am longing to love mankind, mother… But I believe that acting is better than making pilgrimages, and that a good deed is worth more than the best of words. I believe, too, that I shall find the way to do so, for there are two standing by me [Freder and his mother], who wish to help me…. (236-7)

The handshake in the final scene of the film does not do justice to Joh's transformation. Joh is not only "longing to love mankind" but also recognizes in similar fashion to Čapek (mentioned in Chapter 9) that a single deed is worth more than a great truth. In addition, his confidence that this time he will succeed is directly attributed to the fact that he is surrounded by Freder and his mother who wish to help him, rather than being the autocratic leader he once was. Joh has learned the important difference between leadership predicated on behavior as opposed to great truths.

Figure 8 The Seal of Solomon as Depicted in the Film

Metropolis includes many religious symbols and references to the occult. According to Wosk, "what interested [Lang] most in *Metropolis* was the conflict between the magical and the world of modern technology" (405-6). As in *Frankenstein*, the religious symbols serve to establish a historical connection to ancient (magical) ambitions of achieving posthuman power. However, Wosk does not discuss this particular dimension of the film in detail. The religious and magical connotations of the narrative (in both film and novel) have largely been overlooked. This magical dimension is depicted when the robot is seen seated or standing below a large pentagram. The film offers no further explanation as to the employment of this imagery. The novel, however, frequently references the "seal of Solomon," described as "the pentagram" (43), which accounts for the use of the imagery in the film. The interpretation adopted in this dissertation is that religious and occult descriptions throughout *Metropolis* function as symbolic representations that serve to connect modern technological progress as a means to fulfill ancient desires of mastery over the world.

Given the pentagram in the film and the specific mentioning of "Solomon" in the novel, this dissertation considers the theme of magic as an allusion or intertextual reference to the legend of Solomon. One version of the legend of Solomon which accounts for the function of the pentagram is

The Testament of Solomon, a text originally written in Greek, whose author and date of publication are unknown.[34] The first-person narrative, written from the perspective of King Solomon, explains that while he was praying, archangel Michael appeared and delivered a ring, "engraved with a pentalpha [or pentagram]" (268). Armed with the ring, Solomon calls up various demons one by one, subjugating and forcing them to complete the construction of his temple. In the end, however, Solomon gives in to the temptation of a beautiful woman: "I then was moved, but crafty Eros brought and laid by her for me five grasshoppers, saying: 'Take these grasshoppers, and crush them together in the name of the god Moloch; and then will I sleep with you.' And this I actually did" (295). Subsequently, "the Spirit of God leaves him; he is weakened, and builds temples to Baal, Raphan, and Moloch" (296). Solomon explains that he "became the sport of idols and demons," and he wrote his testament "that ye who get possession of it may pray," and avoid the mistakes he had made (296). In *Metropolis*, the seal of Solomon is a fitting symbol through which the themes of supernatural empowerment and subsequent corruption by means of temptation are expressed. Where Victor Frankenstein eventually claimed to have abandoned his alchemical leanings, these symbols still carry much more potency in *Metropolis.* The robot is the product of Rotwang's ability to control supernatural forces for his own selfish ends: to gain revenge on Joh, who believed in turn that he could exploit Rotwang's genius to seduce the masses. It is both men's failure to follow La Mettrie's natural law that leads to the catastrophic destruction of Metropolis.

Metropolis was not the first science fiction narrative to combine magic and science for the construction of robots. In *Tomorrow's Eve,* the Professor mentioned that he similarly controls the android through the use of rings, while Victor Frankenstein studied Cornelius Agrippa and Paracelsus prior to studying chemistry. In *Frankenstein*, the character of Waldman reminds readers that "these

[34] James Harding and Loveday Alexander, in their article "Dating the Testament of Solomon" (1999), explain in detail why "Scholarly opinion on the date […] varies widely". They claim that the sixth century seems the most likely period, but they also caution that "our evidence is ambiguous enough to preclude any possibility of certainty". Harding and Alexander state that, despite the mysterious origins of the text, "it is clear that the legends of King Solomon acted as a magnet to practicing magicians from late antiquity through to the end of the Middle Ages and the beginning of the so-called 'Age of Enlightenment'". The manuscripts that currently survive date back to the fifteenth, sixteenth and seventeenth centuries.

[alchemists] were men to whose indefatigable zeal modern philosophers were indebted for most of the foundations of their knowledge" (42). As Paul Kléber Monod explains, in *Solomon's Secret Arts*, "the occult [is] a type of thinking, expressed either in writing or in action, that allowed the boundary between the natural and supernatural to be crossed by the actions of human beings" (5). His research has shown that during the Enlightenment "scientific discovery and occult thinking […] were not at war with each other" (16). *Frankenstein*, *Tomorrow's Eve* and *Metropolis* illustrate that in the modern age science is still a continuation of the oldest occult ambitions or pursuits. Science, although a new methodology, is still a continuation of the oldest occult ambitions of gaining dominion over the world. Both magic and science are (in the context of these narratives) employed for one's self-empowerment rather than the betterment of humanity. All these so-called mad scientists, Frankenstein, Edison (as discussed in this dissertation) and Rotwang, undergo similar developments, expressed by Frankenstein as follows: "Learn from me, […] how dangerous is the acquirement of knowledge, and how much happier that man is who believes his native town to be the world, than he who aspires to become greater than his nature will allow" (46). While magic and science can represent two avenues to obtaining power and control over the natural as well as the social world, *Metropolis* also introduces a third avenue, namely religion. Religion functions in the texts as a narrative capable of swaying the public mind. This theme of religion is best expressed through the character of Georgi, who literally becomes "greater than his nature will allow."

The character of Georgi, also known by his assigned number *11811*, is one of the workmen in the underground. Freder, feeling sorry for Georgi, exchanges places with him. Not believing his luck, Georgi finds himself dressed in Freder's clothes, a pocket full of money, and free to explore all that the city has to offer. Giving in to temptation, Georgi goes to Yoshiwara, a red light district in Metropolis. What happens to Georgi in Yoshiwara is not featured in the film, but the novel reveals his encounter with a drug called Maohee. The proprietor of Yoshiwara, a man named September, explains that Maohee is divine, because "it is the only thing which makes us feel the intoxication of the others […] Not of one other – no, of the multitude which rolls itself into a lump" (78). September

explains that in a circular room people crowd densely together, and their faces "appear as one face" (78). During this ritual, some kind of communal mind is formed, and a chosen one finds him or herself in the middle of the room as the centerpiece of all the intoxication: "What he feels […] is felt by them all" (79). This drug-induced ritual serves as an escape from technological dominance and a return to human interactions and relations. Put differently, Maohee is a remedy to the alienating effects of modern technology; it provides a kind of sanctuary where humans can connect with others beyond the reach of technological mediation.

Georgi becomes the chosen one during one of the Maohee rituals. He is described as an anti-Christ figure, torturing everyone who participates in his "intoxication," forcing them to become mad. He is surrounded by women, kneeling, crouching, "as though drunk," while others huddle together as "symbols of fear" (80). September explains that he has seen all kinds of behaviors, but states about Georgi: "He believes himself to be a machine and is praying to himself. He has forced the others to pray for him" (80). Georgi refers to himself as both the devil and a machine: "I am the Three-In-One Lucifer – Belial – Satan! […] I am a machine!" (81). The drug makes Georgi, as Frankenstein warns, "greater than his nature will allow," and the result is monstrous. The intoxication, predicated on the meeting of minds in a pure form of disembodied communication, reveals that Georgi cannot relate to other human minds. All his interactions have been with machines, and he can therefore only relate to others as he relates to machines. This scene effectively illustrates how he has lost his humanity, and this loss is predicated on losing the ability to relate to other humans. In the context of this dissertation, it is an important scene. First, it features humans as a group, or humanity rolled not a "lump", while positioning a mind like Georgi's as the central brain of this group. The result is a nightmarish pain and anguish shared by all the participants. . Georgi's mind, of course, has been severely conditioned by the machines, and the scene therefore depicts what happens to human-human interactions when governed by technological mediation. Second, this scene is a foreshadowing of what happens throughout the rest of the narrative of *Metropolis*; a society governed by machines subsequently suffers from what

Richardson describes as attachment wounds, or interhuman connections and relations, and comes to destroy itself. Georgi's attachment wounds can no longer be healed as his isolation from others has endured for too long. Georgi has become a machine as a result and can only repeat or reiterate the same mechanical motions and conditionings he endured himself. September therefore states about Georgi: "He did not know that the ecstasy for men who are damned is also damnation… The fool!" (80). This self-destruction is experienced by Georgi as pleasurable, given that it is the only way he can relate to others much like the creature in *Frankenstein* (as discussed in Chapter 6).

Given Joh's initial detachment from the masses in his city, Georgi must fashion his own beliefs, his own ideals and essentially his own religion. It is through the character of Georgi that one can understand the function of religious and occult references throughout the text. When he slaves at the Pater Noster machine, he mumbles (what can only be described as) a mechanical-religious prayer:

> Pater Noster... that means, Our Father!... Our Father, which are in heaven! We are in hell. Our Father!... What is thy name? Art thou called Pater Noster, Our Father? Or Joh Fredersen? […] And forgive us our trespasses... […] The trespass of having a brain and a heart, that thou hast not, machine?... […] Lead us not into temptation to rise against thee, machine, for thou art stronger than we, […] Amen... (31-2)

This moment of worship reveals that Georgi confuses a religious deity with a machine, and with Joh Fredersen. The narrative neatly demonstrates that machines and Joh Fredersen are beings on par with religious deities. His character, being merely a number, is a microcosmic representation of all the workers in Metropolis who are similarly experiencing a crisis of faith. The effect Goergi has on others whilst intoxicated with Maohee represents the same effect that Joh Fredersen has on all the workers of the city. However, these passages reveal that the narrative is not expressly concerned with specific religious narratives but rather with the function of narratives in relation to social organization more generally: The workers cannot ascertain to whom their loyalties or commitments lie and to what group they belong. Joh, who is unfettered by any religious concerns, employs the

robot as an instrument of propaganda, a tool with which to sway the sentiments of the group mind. The message that the robot spreads, namely violence, is in direct opposition to Maria's sermons of peace. While Maria preaches: "Do not fight, my brothers, for that makes you to sin" (65), robotic Maria preaches: "Why do you not throw yourselves – a hundred thousand murdering fists – upon the machines and strike them dead" (154). While the workers originally adhere to Maria's sermons, they eventually undergo a conversion and believe robotic Maria's sermons. The narrative of *Metropolis* depicts the relationship between groups and the narratives that inform them alongside the alienating effects of technology.

The building that Joh Fredersen lives in is aptly named the "New Tower of Babel" (12), but that title already suggests that the modern technological era is little more than ancient (biblical) history repeating itself. *Metropolis* recognized that technological innovations in 1927 meant that the social group mind would be solidified more than ever. In his book *Propaganda* (1928), Edward Bernays described how the world was changing:

> As civilization has become more complex, and as the need for invisible government has been increasingly demonstrated, the technical means have been invented and developed by which opinion may be regimented. With the printing press and the newspaper, the railroad, the telephone, telegraph, radio and airplanes, ideas can be spread rapidly and even instantaneously all over the whole of America. (39-40)

While Bernays noted that "This invisible, intertwining structure of groupings and associations [in society] is the mechanism by which democracy has organized its group mind and simplified its mass thinking" (44), *Metropolis* already prefigured such a group mind. Additionally, *Metropolis* accurately pinpoints the property that makes the group mind complex, namely the importance of human-human relations and communication. From ancient magical to religious to modern scientific contexts, *Metropolis* prefigured exactly what Bernays wrote in his book one year later, namely that those "who understand the mental processes and social patterns of the masses […] harness old social forces and contrive new ways to bind and guide the world" (37). Given Joh's transformation

discussed in the previous section, he understands the "mental processes and social patterns of the masses." He understands that there is a difference between "binding" and "guiding" the world.

Group Selection and Altruism

Metropolis features two scenes of altruistic behavior involving Freder and Georgi. Freder's decision to trade places with Georgi is a pointless gesture which serves no purpose; it is entirely motivated, presumably, by his empathy. Georgi later sacrifices his own life to save Freder and Metropolis. Georgi's sacrifice is merely another moment of altruistic behavior, in the form of reciprocal altruism towards Freder, but also a more sincere form of altruism towards the city as he sacrifices himself for the sake of the group. At his moment of sacrifice, Georgi becomes a Christ-like figure given that he is killed by the very group he is saving and thereby finds redemption. It is important to point out, in the context of this dissertation, that *Metropolis* is the only text to include obvious scenes of altruistic behaviors between characters. It is also important to point out that Freder's altruism is directed towards a stranger, while Georgi's sacrifice is similarly for the sake of the whole group and not merely an attempt at saving Freder's life.

Maria's sermon of the tower of Babel, slightly altered from the biblical narrative, neatly summarizes all the considerations discussed in this chapter. It describes the social (systematic) collapse when the distance between individuals and the group mind becomes too great (or when a complex property is removed from a system). Maria explains:

> Never have men worked more rapidly, for they all had one thought, one aim and one dream. […] The work was greater than their working hands. […] Then the builders sent their messengers to all four winds of the world and enlisted Hands, working Hands for their mighty work. The Hands came. The Hands worked for wages. — The Hands did not even know what they were making. […] The Brain which conceived the construction of the Tower of Babel was unknown to those who built it. Brain and Hands were far apart and strangers. […] The hymn of praise of one became the other's curse. 'Babel!' shouted one, meaning: Divinity, Coronation, Eternal, Triumph! 'Babel' shouted the other, meaning: Hell, Slavery,

> Eternal, Damnation! The same word was prayer and blasphemy. Speaking the same words, the men did not understand each other. That men no longer understood each other, that Brain and Hands no longer understood each other, was to blame that the Tower of Babel was given up to destruction […]. (63-5)

Maria's account, a foreshadowing of the events in *Metropolis*, explains that the distance between the brain and hands being "far apart" caused a confusion of language or semantic drift. Through the disintegration of language, despite "speaking the same words," the entire project eventually failed; figuratively speaking, the complex adaptive system of humanity was deprived of the deep property that made it complex. In the narrative, mediation between different segments of society is not the result of some largescale revolution, belief or event; rather human-human interaction and altruism is what allows for the brain and hands to reconnect. In other words, much like the theme of love in *R.U.R.* (discussed in Chapter 9) which is expressed through altruistic gestures of characters, the theme of mediation in *Metropolis* is similarly enacted through overt acts of altruism by its characters. A close-reading reveals that mediation in the novel, as overcoming fissures in society, only occurs as a direct result of the characters acting altruistically towards other characters; Freder exchanges places with Georgi; Georgi self-sacrifices to save both Freder and the city; Maria's heroism is revealed when risking her life to save the children, whose parents at that point in time wish to kill her. These behaviors are what stem the tide of the violent revolution that would otherwise have irreparably destroyed the city and its inhabitants. In other words, it is not simply the theme of mediation that is significant in this text, but the theme of altruism as a primary condition that allows for such forms of mediation to occur; if it were not for their altruistic deeds, Joh's transformation in the novel would not have been possible.

Conclusion to Part II

In all of the texts discussed in Part two, one finds group selection and altruism at work in various ways. In *Frankenstein,* the important realization occurs on the cusp between acting out of self-interest versus causing harm to oneself (Victor) for the sake of humanity. At the moment when humanity is under threat, Victor chooses to act altruistically. Shortly after the American Civil War, *The Steam Man* concerned itself with new techno-utopian depictions of an infinitely affluent society with an endless source of altruism. It seeks to foster a new kind of conformity enforcement, rallying readers behind industrial technology. When faced with group threats, technology can become a source of liberation and empowerment, while it denies any possibility of generating diversity. As long as the steam man is exalted, the system of colonial expansion is stocked with altruism. The narrative already betrays the implicit notion that the steam man may be appropriated by other cultures (that already pose a threat). The steam man, although thoroughly colonial, is nevertheless a transcultural entity subjected to colonial intentions.

Villiers' *Tomorrow's Eve* destroys any such utopian representations of science and technology as a means of fulfilling social/individual desires. The narrative reminds readers, as Kelly puts it, that "the real can never be known in itself; we have only artificial, duplicate ideas of the real. In that sense, then, everything is 'artificial'" (138). Whether or not electricity is an almighty power, *Tomorrow's Eve* reminds readers that the fallibility of the human condition precedes the fallibility of the robotic condition. Human fallibility, for Villiers, stems from an unavoidable existential ignorance of the ineffable mysteries of the ideal. The narrative, just like its protagonist, returns to the indefinite existential silence that it initially desires to reproduce. It is a social critique of cultural vanity, narcissism and selfishness resulting in inauthentic dialogue, motives and desires. In such a context, sincerity, including altruism, becomes impossible. Ultimately, the reader recognizes that the eternal symbol in the text, namely the relationship between Science and Humanity, is one in which not only science must innovate, but ordinary and basic human values must also develop independently from scientific pursuits.

R.U.R. counters such a pessimism by putting forward a more optimistic perspective. Rather than great truths, the text advocates simple decency, kindness and tolerance. It casts an easy glance over the differences between humans and machines, effortlessly making them exchange evolutionary roles, and favors the group capable of more altruism (or selfless love). Where Victor in *Frankenstein* becomes an altruist when he realizes that his second creature could mean the end of humanity, *R.U.R.* actually presents its audience with such an eventuality, and positions its altruistic characters similarly in the midst of intergroup competition.

Metropolis continues *R.U.R.*'s rhetoric, arguing for altruism, cooperation and tolerance. Despite the many digressions and intricate plot, the narrative often addresses various human emotions in a direct and candid manner. The climax of the novel, showing Maria's heroism in the face of death, effectively becomes a celebration of the human condition (a celebration of incarnation), or the enduring human spirit standing up against immeasurable odds. Like *R.U.R.* and *Tomorrow's Eve*, it stresses the importance of human companionship, and unlike the preceding two texts, specifically pinpoints human communication, by appropriating the narrative of Babel, as a central feature to social cohesion. While novel calls upon a heart to mediate between the mind of society and its working hands, it also presents such forms of mediation as contingent upon unaffiliated acts of random kindness.

There is tremendous duality in all of these texts. In *The Steam Man*, there is a doubling between Johnny and the steam man itself. In *Metropolis*, the duality of Maria harks back to the duality of the characters of Hadaly and Helena; they all bring *both* salvation and damnation. Separate from the feminist implications of three female robots being either angelic or demonic, from a technological perspective such duality is to be expected. Blascovich and Bailenson, adopting psychological perspectives while discussing the effects of the virtual reality revolution in the modern world, conclude their discussions by invoking the so-called *dual-use principle*, which states that "any technology can be used for good or for evil" (227). This principle, a guiding rule whenever one considers the future implications of any technology, can only put forward the familiar notion: "With

each technological advancement, new media technologies have magnified existing social effects, some of them positive and others negative" (229). All the female robotic characters discussed in this thesis exhibit a similar duality: Hadaly, Helena and Maria all have their doubles, and the arrival of these doubles ultimately introduce the equal and opposite effect, and together magnify preexisting dilemmas or problems. However, these narratives do not stop at the dual-use principle, but point to loftier themes. In *Tomorrow's Eve*, the character of Sowana is left mysterious, conforming to the theme that if we cannot know ourselves, we cannot be sincere. *R.U.R.* explains that empathy is not enough when it is misplaced or based on ignorance or personal (unrealistic) ambitions or ideals. The real Helena's mistake is to project her own pity onto the machines, and her marriage to Domin seems to be one of convenience rather than genuine love or passion. Her double, robotic Helena, does not misplace her affection towards Primus, and their marriage is one of innocence. In *Metropolis*, both real Maria and robotic Maria call for revolution with similar motivations (liberating the workers), but it is the *means* and not the *ends* that make all the difference between their approaches. The robot calls for violence, while Maria calls for patience, peace and tolerance. From existential anxiety, to empathy, to altruism, there is more at work in these narratives than a reiteration of the dual-use principle. They all point to the same fundamental realization: in a world increasingly governed by advanced technology, human decency becomes *more* relevant rather than less. Where *The Steam Man* presents technology as a kind of social panacea, *Tomorrow's Eve, R.U.R.* and *Metropolis* present technology and its alienating effects as a social disease that can only be cured by sincere forms of altruistic gestures and sentiments.

Tomorrow's Eve contains the least amount of group selection principles. It is thoroughly individualistic, but also, for that reason, the text that best illustrates the limitations of such individualist perspectives. In *The Steam Man*, *R.U.R.* and *Metropolis*, groups and group dynamics are actively employed. Despite racist undertones (or overtones) in *The Steam Man*, which functions at national or racial levels (as opposed to humanity at large), these group selection texts are inherently more enthusiastic and optimistic about humankind. Each in their own way celebrate human virtues

as sources of redemption. In so doing, it becomes clear that these texts were dealing with themes that would only come to be appreciated in the twenty-first century in the age of social robots. When technology interferes with our ability to communicate, there is no cohesion as a group, and individuals become isolated. At this point, robots become sources of solace, or as Lanier puts it: "Separation anxiety is assuaged by constant connection. Young people announce every detail of their lives on services like Twitter not to show off, but to avoid the closed door at bedtime, the empty room, the screaming vacuum of an isolated mind" (180). Technological mediation cannot successfully assuage such forms of separation anxiety, or attachment wounds as Richardson describes puts it, nor can technological mediation replace human-human interactions. As these science fiction narratives illustrate, technology will not save humanity; only humanity can save humanity.

Conclusion:

This dissertation set out to examine the complex relationships between humans and machines as portrayed in five foundational works of science fiction before 1930. Motivated by the realization that these literary texts employed fictional robots as a means with which to explore themes of altruism, complementary scientific discourses were introduced to establish a critical framework. Chapter 1 served to delineate the use of terms and their intended definitions in the context of this dissertation. Chapter 2 illustrated that critical posthumanist discourses, particularly those that reflect on human-machine ontologies, offer useful insights when it comes to the formation of posthuman subjectivities. Chapter 3 discussed La Mettrie human-machine thesis and his recognition that such a perspective does not negate the necessity of moral conduct. La Mettrie's natural law argued that humans (and animals) are endowed with evolutionary moral imperatives, and critical posthumanists are still continuing to develop these ideas today. Chapter 4 revisited Alan Turing's insights pertaining to the importance of groups or communities in relation to the development of artificial intelligence. His famous test emphasizes behavioral, rather than existential, differences when it comes to defining and understanding intelligence. Chapter 5 examined the nature of kin and group selection dynamics, and briefly discussed complex adaptive systems as a useful framework that elucidates the various intricacies of altruism. It also discussed how religious and secular scholars in the nineteenth and early twentieth centuries examined insect colonies as models of human social organization, placing particular emphasis on their altruistic behaviors. These perspectives offer a useful framework with which to examine themes of altruism in early science fiction robots. Applying such a framework, although not universally applicable to all early science fiction robots, nevertheless demonstrates a certain contingency between early science fiction and contemporary anxieties, concerns and ideas regarding human-robotic and human-human interactions. The implication is that human-human and human-robotic interactions are not merely contemporary concerns, but from technological posthumanist perspectives, humans and technology have continually redefined and informed one another. In the context of this dissertation, these texts reveal that aside from mechanical bodies,

disembodied minds, robotic slaves and gods, early robotic science fiction was also concerned with human morality and behavior.

The theoretical framework established in this dissertation may serve as an important contribution to critical posthumanist considerations and science fiction robotic scholarship more generally. When exploring the intersections of humans and machines, much consideration has been given to ontological questions that pertain to problems of consciousness and (mechanical/biological) embodiment. For example, Hayles points out that:

> Although Moravec's dream of downloading human consciousness into a computer would likely come in for some hard knocks […] literary studies share with Moravec a major blind spot when it comes to the significance of embodiment. This blind spot is most evident, perhaps, when literary and cultural critics confront the fields of evolutionary biology. (284)

This dissertation agrees with Hayles' "blind spot" as far as the importance and relevance of evolutionary biology is concerned. However, analyzing robotic characters' existential dimensions reveal little about their actual behavior, along with the themes and morals they represent. As this dissertation has shown, robots come in all kinds of existential varieties, but they seem to share certain behavioral, thematic and moral comparisons and similarities that surpass their existential constitutions. This dissertation therefore contributes to these considerations by developing a framework modeled after the principles of evolutionary biology and altruism in relation to human behavior. The aim was not to evaluate or compare the validity of evolutionary biological debates, but rather to illustrate that early science fiction robots hold the potential to still contribute interesting and important insights to our ongoing evaluations towards the formation of critical posthumanist subjectivities. In relation to robotic characters, we should not make the mistake that Turing outlines in his test, namely of trying to ascertain the ontological dimensions of these mechanical beings (or characters), but rather focus on their behaviors and themes as well as our own in relation to such characters. In this manner, one is better able elucidate their valuable contributions to considerations that are still being evaluated. As La Mettrie puts it: "I do not say

discover with clarity what the nature of man is, but rather attain the highest degree of probability" (30).

Applying the framework developed in Part I of this dissertation to the literary texts in Part II reveals the various adoptions and adaptations of the themes of altruism in these literary texts. *The Steam Man of the Prairies* celebrates the unification of a group under the banner of technological innovation. As civilization has grown more diverse by the twenty-first century, we are in a position to appreciate the levels of political incorrectness expressed in the text, but also in a position to appreciate the evolutionary validity of the prejudices it expresses. It reveals the pejorative dimensions to group formations and group minds, and altruism – while helping the group, one is not by definition helping humanity. The narrative pits one group against another and is constantly aware of (and playing with) the notion that altruism is an invaluable resource to whichever group possesses it. *Tomorrow's Eve* adopts a satirical approach but still affirms the importance of altruism and authentic human-human interactions rather than human-machine interactions. The narrative prefigures the other implication of the Turing test, namely that machines which pass the test might be symptomatic of the extent to which humanity has lowered or devalued itself. In *R.U.R.*, natural selection blindly selects the group who is most selfless and cooperative. The narrative shows that kindness is not weakness, but requires strength, self-sacrifice, commitment and discipline as exemplified by the character of Alquist. The final act deals with the opposition of selfishness and selflessness rather than the human-machine opposition. In *Metropolis*, one finds a successful (or ideal) mediation between segments of society, healing social rifts and fissures caused by the modern technological world. However, mediation in the narrative is not the result of merely obtaining a newly transformed leader, ideology or technology; it comes as a direct result of the various altruistic actions of its characters. Without such forms of altruism, no form of successful mediation can occur.

To conclude, this dissertation suspects that the theme of altruism did not end with *Metropolis*. With this dissertation's framework in mind, it would be interesting to revisit successive generations of robotic characters. After *Metropolis*, Otto Binder (writing under the penname Eando

Binder) produced his *Adam Link* stories from 1939 to 1942 which were published as a collection in 1965, entitled *Adam Link – Robot* (Binder). Fittingly, the character of Adam Link is deeply affected by his reading of Shelley's *Frankenstein*, and the name "Frankenstein" is employed throughout the novel as a pejorative term that offends the robot: "I was the Frankenstein product of a mad genius, a twisted travesty of the human form. The Machine had finally arisen, as had been foretold in imaginative literature, threatening Mankind" (25). In the epilogue, written by the character of Adam Link, the robot poses interesting questions to its readers: "Should [robot] Eve and I create a robot race? Would they aid mankind – or become Frankensteins?" (174). While the first Adam Link story is entitled "I, Robot," the title today is more commonly associated with Isaac Asimov's eponymous collection of short stories published in 1950. Asimov praised Binder's robot by saying: "To anyone fond of the robot story in science fiction, Adam Link is of extraordinary interest. The robot-with-emotion has rarely been handled so well" (qtd. in Binder).

Asimov, arguably the most famous robotic science fiction author, is remembered for his "Three Laws of Robotics," which specify the general behavior of robots towards humans. The First Law stipulates: "A robot may not injure a human being or, through inaction, allow a human being to come to harm" (*I, Robot*) and is considered the most important of the three laws. However, it is rarely mentioned or discussed that the three laws of robotics actually consist of four laws. In the context of this dissertation, it becomes particularly relevant that, according to the protagonist Dr. Susan Calvin, the first law carries a second implication as well which later became known as the Zeroth law (or law zero, preceding the first law). She explains that: "[robots] follow the First Law. But the machines work not for any single human being, but for all humanity, so that the First Law becomes: 'No machine may harm humanity; or, through inaction, allow humanity to come to harm'" (241-2). In Asimov's fifth robot book, *Robots and Empire* (1985), this became known as the Zeroth law, the law preceding the First law. The important implication is the difference between protecting humans vs. protecting humanity: "Use the Zeroth Law, but not to justify needless harm to individuals. First Law is almost as important" (Asimov *Robots and Empire* 436). The difference

between these two laws, individual and group or human and humanity, eventually constitutes two very different kinds of moral frameworks that would be interesting to evaluate in more detail.

One must also consider the example of Philip K. Dick's classic robotic science fiction novel, *Do Androids Dream of Electric Sheep* (1968). The narrative explains that:

> Empathy, evidently, existed only within the human community, whereas intelligence to some degree could be found throughout every phylum and order including the arachnida. For one thing, the emphatic faculty probably required an unimpaired group instinct; a solitary organism, such as a spider, would have no use for it […] A herd animal such as man would acquire a higher survival factor through this [instinct]; an owl or a cobra would be destroyed. Evidently the humanoid robot constituted a solitary predator. (28-9)

One also finds an interesting inverse of La Mettrie's natural law in Dick's narrative: "You shall kill only the killers, Mercer had told them the year empathy boxes first appeared on Earth" (29). From these instances alone it becomes clear that Dick's narrative also deserves to be revisited to investigate the ways in which it deals with themes of altruism. Given the adaptations of Asimov and Dick's novels into Hollywood films, alongside the countless number of other robotic characters in existence today, these few but important examples sufficiently demonstrate that there is much research to be done in robotic science fiction literature and film.

As far as the future implications of intelligent machines for humanity are concerned, one should keep in mind that humans and robots share an inevitable future as well as a common history. Our fates are intertwined, given that their intelligence will be predicated on largescale interactions with humans. One can therefore speculate that machines, like humans, will be both selfish and altruistic. As Asimov put it, "Their first care, therefore, is to preserve themselves, for us" (*I, Robot* 242). What is certain in such a future, however, is that human moral conduct will only become more important, not less.

List of Figures

Works Cited

Aamoth, Doug. “Interview: Eugene Goostman Passes the Turing Test.” *Time*, Time, 9 June 2014, time.com/2847900/eugene-goostman-turing-test/.

Adams, Douglas. *The Hitch Hiker's Guide to the Galaxy*. Picador, 2002.

Alkon, Paul K. *Science Fiction Before 1900: Imagination Discovers Technology*. Routledge, 2002.

“AlphaGo: The Story so Far.” *Deepmind*, deepmind.com/research/case-studies/alphago-the-story-so-far.

Asimov, Isaac, and Martin H. Greenberg. “Robbie.” *Machines That Think: The Best Science Fiction Stories about Robots & Computers*, edited by Patricia S. Warrick, Henry Holt and Company, 1984, pp. 68–89.

Asimov, Isaac. *I, Robot*. Harper Voyager, 2018.

Asimov, Isaac. “Introduction: Robots, Computers, and Fear.” *Machines That Think: The Best Science Fiction Stories about Robots & Computers*, edited by Patricia S. Warrick and Martin H. Greenberg, Henry Holt and Company, 1984, pp. 1–11.

Asimov, Isaac. *Robots and Empire*. Harper Voyager, 2018.

Badham, Van. “We Can Beat the Robots - with Democracy | Van Badham.” *The Guardian*, Guardian News and Media, 2 Nov. 2017, www.theguardian.com/commentisfree/2017/nov/03/we-can-beat-the-robots-with-democracy.

Badmington, Neil. *Alien Chic: Posthumanism and the Other Within*. Routledge, 2004.

Bellamy, Elizabeth W. “Ely's Automatic Housemaid.” *The Feminine Future: Early Science Fiction by Women Writers*, edited by Mike Ashley, Dover Publications Inc., 2015, pp. 43–52.

Bernatchez, Josh. “Monstrosity, Suffering, Subjectivity, and Sympathetic Community in Frankenstein and ‘The Structure of Torture.’” *Vintage Visions: Essays on Early Science Fiction*, edited by Arthur B. Evans, Wesleyan University Press, 2014, pp. 66–81.

Bernays, Edward. *Propaganda*. Ig Publishing, 2005.

Binder, Eando. *Adam Link: Robot*. Paperback Library, 1965.

Binder, Eando. "I, Robot." *Amazing Stories January 1939*, edited by B. G. Davis and Raymond A. Palmer, vol. 13, ser. 1, Ziff-Davis Publishing Company, 1939, pp. 8–18. 1.

Blackford, Russell. *Science Fiction and the Moral Imagination - Visions, Minds, Ethics*. Springer International Publish, 2017.

Blascovich, Jim, and Jeremy Bailenson. *Infinite Reality: Avatars, Eternal Life, New Worlds, and the Dawn of the Virtual Revolution*. William Morrow, 2011.

Bleiler, Everett Franklin. *Science-Fiction: The Early Years*. Kent State University Press, 1990.

Bloom, Howard. *Global Brain: The Evolution of Mass Mind from the Big Bang to the 21st Century*. John Wiley & Sons, 2000.

Brown, Bill. "Science Fiction, the World's Fair, and the Prosthetics of Empire, 1910-1915." *Cultures of United States Imperialism*, edited by Amy Kaplan and Donald E. Pease, Duke University Press, 1993, pp. 129–163.

Buchen, Irving H. "Frankenstein and the Alchemy of Creation and Evolution." *The Wordsworth Circle*, vol. 8, no. 2, 1977, pp. 103–112., doi:10.1086/twc24039234.

Calvin, Ritch. "The French Dick: Villiers De L'Isle-Adam, Philip K. Dick, and the Android." *Extrapolation (Pre-2012)*, vol. 48, no. 2, 2007, pp. 1–14.

Cameron, James, et al. *James Cameron's Story of Science Fiction*. Insight Editions, 2018.

Čapek, Karel. *Believe in People - The Essential Karel Čapek*. Translated by Sarka Tobrmanova-kuhnova, Faber And Faber, 2010.

Čapek Karel. *R.U.R. (Rossum's Universal Robots)*. Translated by Claudia Novack, Penguin Books, 2004.

Chomsky, Noam. "The Mysteries of Nature: How Deeply Hidden." *What Kind of Creatures Are We?*, by Noam Chomsky, Columbia University Press, 2016, pp. 81–127.

Christian, Brian. *Most Human Human; What Artificial Intelligence Teaches Us About Being Alive*. Penguin Books Ltd, 2012.

Clarke, Peter. *The Singularity Survival Guide: How to Get On the Good Side of Your Future Robot Overlords*. Logos Literature, 2019.

Cloonan, William. *Frères Ennemis: The French in American Literature, Americans in French Literature*. Liverpool University Press, 2018.

Condliffe, Jamie. "Top Breakthrough Technologies For 2018 : Dueling Neural Networks." *CITI I/O*, 8 Mar. 2018, citi.io/2018/03/08/top-breakthrough-technologies-for-2018-dueling-neural-networks/.

Conybear, Frederick C. "The Testament of Solomon." *The Three Magical Books of Solomon: The Testament of Solomon & His Greater & Lesser Keys in One Volume*, Mockingbird Press, 2017, pp. 255–296.

Copeland, B. Jack. *Turing: Pioneer of the Information Age*. Oxford University Press, 2012.

Copestake, Jen. "How Pope Francis Could Shape the Future of Robotics." *BBC News*, BBC, 24 Mar. 2019, www.bbc.com/news/technology-47668476.

Courtland, Rachel. "Gordon Moore: The Man Whose Name Means Progress." *IEEE Spectrum: Technology, Engineering, and Science News*, 30 Mar. 2015, spectrum.ieee.org/computing/hardware/gordon-moore-the-man-whose-name-means-progress.

Darwin, Charles. "Descent of Man." *Darwin Online*, 1871, darwin-online.org.uk/content/frameset?itemID=F937.1&viewtype=text&pageseq=1.

De Heussey, Robert Du Pontavice. *Villiers De L'Isle-Adam: His Life and Works*. Translated by Mary Loyd, 1894. https://quod.lib.umich.edu/g/genpub/ahs9353.0001.001

Del Rey, Lester. "Helen O'Loy." *Astounding Science-Fiction December 1938*, XXII, ser. 4, Street and Smith, 1938, pp. 118–125. 4.

Dick, Philip K. *Do Androids Dream of Electric Sheep*. Gollancz, 2010.

Dinan, Matthew. “The Robot Condition: Karel Čapek's R.U.R. and Hannah Arendt on Labor, Technology, and Innovation.” *Perspectives on Political Science*, vol. 46, no. 2, 2017, pp. 108–117., doi:10.1080/10457097.2016.1270677.

Edwards, Justin D. “Introduction: Technogothics.” *Technologies of the Gothic in Literature and Culture*, edited by Justin D Edwards, Routledge , 2015, pp. 1–16.

Ellis, Edward S. “The Huge Hunter; or, The Steam Man of the Prairies.” *The Steam Man of the Prairies: a Dime Novel Anthology*, edited by John Spencer, Monroe St. Press, 2016, pp. 1–123.

“Ellis, Edward Sylvester.” *EDWARD SYLVESTER ELLIS*, Northern Illinois University Libraries, www.ulib.niu.edu/badndp/ellis_edward.html.

Eugenios, Jillian. “Ray Kurzweil: Humans Will Be Hybrids by 2030.” *CNNMoney*, Cable News Network, 4 June 2015, money.cnn.com/2015/06/03/technology/ray-kurzweil-predictions/index.html.

Ford, Martin. *The Rise of the Robots; Technology and the Threat of Mass Enemployment*. Oneworld, 2015.

Forrest, Jennifer. “The Lord of Hadaly's Rings: Regulating the Female Body in Villiers De L'Isle-Adam's ‘L'Eve Future.’” *South Central Review*, vol. 13, no. 4, 1996, pp. 18–37.

Forster, Edward Morgan. *The Machine Stops*. Penguin Classics, 2011.

Gasché, Rodolphe. “The Stelliferous Fold: On Villiers De L'Isle-Adam's ‘L'Eve Future.’” *Studies in Romanticism*, vol. 22, no. 2, 1983, pp. 293–327., doi:10.2307/25600430.

Gibbs, Samuel. “Musk, Wozniak and Hawking Urge Ban on Warfare AI and Autonomous Weapons.” *The Guardian*, 27 July 2015, www.theguardian.com/technology/2015/jul/27/musk-wozniak-hawking-ban-ai-autonomous-weapons.

Greenhill, Jennifer A. “The View from Outside: Rockwell and Race in 1950” *American Art*, vol. 21, no. 2, 2007, pp. 70–95., doi:10.1086/521890.

Hales, Barbara. “Taming the Technological Shrew: Woman as Machine in Weimar Culture.” *Neophilologus*, vol. 94, 2010, pp. 301–316., doi:10.1007/s11061-009-9177-5.

Hampton, Gregory Jerome. *Imagining Slaves and Robots in Literature, Film, and Popular Culture: Reinventing Yesterday's Slave with Tomorrow's Robot*. Lexington Books, 2015.

Hapgood, Fred. "When Robots Live Among Us." *Discover Magazine*, Discover Magazine, 27 May 2008, www.discovermagazine.com/technology/when-robots-live-among-us.

Haraway, Donna J. "A Cyborg Manifesto: Science, Technology, and Socialist-Feminism in the Late Twentieth Century." *Simians, Cyborgs, and Women: the Reinvention of Nature*, by Donna J Haraway, Free Assoc. Books, 1991, pp. 149–181.

Harbou, Thea von, and Forrest J. Ackerman. *Metropolis*. Sense of Wonder Press, 2002.

Harding, James, and Loveday Alexander. "Dating the Testament of Solomon." *Old Testament Pseudepigrapha*, University of St Andrews, 28 May 1999, otp.wp.st-andrews.ac.uk/guest-lectures/dating-the-testament-of-solomon/.

Harman, Oren. *The Price of Altruism: George Price and the Search for the Origins of Kindness*. Vintage, 2011.

Harris, Clare Winger. "The Artificial Man." *The Feminine Future: Early Science Fiction by Women Writers*, edited by Mike Ashley, Dover Publications Inc., 2015, pp. 164–175.

Hayles, N. Katherine. *How We Became Posthuman: Virtual Bodies in Cybernetics, Literature and Informatics*. Univ. of Chicago Press, 1999.

Haynes, Roslynn. "Chapter 1: The Alchemist in Fiction: The Master Narrative." *The Public Image of Chemistry*, edited by Joachim Schummer et al., World Scientific, 2007, pp. 7–36.

Johannsen, Robert W. "The Meaning of Manifest Destiny." *Manifest Destiny and Empire: American Antebellum Expansionism*, edited by Sam W. Haynes and Christopher Morris, Texas A&M University Press, 1997, pp. 7–20.

Johnson, Jeffrey Allan. "Dr. Frankenstein, I Presume? Revising the Popular Image of Frankenstein." *Literature and Medicine*, vol. 36, no. 2, 2018, pp. 287–311., doi:https://doi.org/10.1353/lm.2018.0015.

Kakoudaki, Despina. *Anatomy of a Robot: Literature, Cinema, and the Cultural Work of Artificial People*. Rutgers University Press, 2014.

Kang, Minsoo. *Sublime Dreams of Living Machines: The Automaton in the European Imagination*. Harvard University Press, 2011.

Kang, Minsoo. "The Question of the Woman-Machine: Gender, Thermodynamics, and Hysteria in the Nineteenth Century." *SubStance*, vol. 47, no. 3, 2018, pp. 27–43.

Keim, Brandon. "Robots Evolve Altruism, Just as Biology Predicts." *Wired*, Conde Nast, 4 May 2011, www.wired.com/2011/05/robot-altruism/.

Kelly, Dorothy. *Reconstructing Woman: from Fiction to Reality in the Nineteenth-Century Novel*. Pennsylvania State University Press, 2007.

Klíma, Ivan, and Peter Kussi. "Introduction." *R.U.R. (Rossum's Universal Robots)*, translated by Claudia Novack, by Čapek Karel, Penguin Books, 2004, pp. vii-xxv.

Landon, Brooks. *Science Fiction After 1900: from the Steam Man to the Stars*. Routledge, 2002.

Lanier, Jaron. *You Are Not a Gadget: a Manifesto*. Penguin, 2011.

Lathers, Marie. *The Aesthetics of Artifice: Villiers's "L'Eve Future"*. University of North Carolina Press, 1996.

L'Isle-Adam, Auguste Villiers de. *The Future Eve*. Translated by Florence Crewe-Jones, Read Books, 2013.

L'Isle-Adam, Auguste Villiers de. *Tomorrow's Eve*. Translated by Robert M. Adams, University of Illinois Press, 2001.

Luckhurst, Roger. *Science Fiction*. Polity, 2005.

MacArthur, Sian. *Gothic Science Fiction: 1818 to the Present*. Palgrave Macmillan, 2015.

Mayor, Adrienne. *Gods and Robots: Myths, Machines, and Ancient Dreams of Technology*. Princeton University Press, 2018.

Mellby, Julie L. "The British Bee Hive." *Princeton University*, The Trustees of Princeton University, 18 Dec. 2009, www.princeton.edu/~graphicarts/2009/12/the_british_bee_hive.html.

Miah, Andy. “Posthumanism: A Critical History.” *Medical Enhancement and Posthumanity*, edited by Ruth F. Chadwick and Bert Gordijn, Springer Netherlands, 2007.

Miles, Robert. *Gothic Writing 1750-1820: A Genealogy*. Manchester University Press, 2002.

Miller, John H., and Scott E. Page. *Complex Adaptive Systems: An Introduction to Computational Models of Social Life*. Princeton University Press, 2007.

Monod Paul Kléber. *Solomon's Secret Arts: the Occult in the Age of Enlightenment*. Yale University Press, 2013.

Mozur, Paul. “Google's AlphaGo Defeats Chinese Go Master in Win for A.I.” *The New York Times*, The New York Times, 23 May 2017, www.nytimes.com/2017/05/23/business/google-deepmind-alphago-go-champion-defeat.html.

Nocks, Lisa. *The Robot: The Life Story of a Technology*. John Hopkins University Press, 2008.

Okasha, Samir. *Evolution and the Levels of Selection*. Oxford University Press, 2008.

Okasha, Samir. “The Controversy of Kin Selection, Group Selection and Altruism: A Controversy Without End.” *OUPblog*, 29 Jan. 2015, blog.oup.com/2015/01/kin-group-selection-controversy/.

Page, Michael R. *Literary Imagination From Erasmus Darwin to H.G. Wells: Science, Evolution, and Ecology*. Routledge , 2016.

Parrinder, Patrick. *Utopian Literature and Science: from the Scientific Revolution to Brave New World and Beyond*. Palgrave Macmillan, 2015.

Pepperell, Robert. *The Posthuman Condition: Consciousness Beyond the Brain*. Intellect, 2004.

Peterfreund, Stuart. “Composing What May Not Be ‘Sad Trash’: A Reconsideration of Mary Shelley's Use of Paracelsus in ‘Frankenstein.’” *Studies in Romanticism*, vol. 43, no. 1, 2004, pp. 79–98., doi:10.2307/25601660.

Poe, Edgar Allan. “The Man That Was Used Up.” *The Murders in the Rue Morgue and Other Tales*, Penguin, 2015, pp. 32–43.

Putnam, George Haven. *The Artificial Mother; A Marital Fantasy*. Knickerbocker Press, New York, 1894.

Rand, Ayn, and Nathaniel Branden. *The Virtue of Selfishness; A New Concept of Egoism*. Signet, 1964.

Reill, Peter. *Vitalizing Nature in the Enlightenment*. Berkeley: University of California Press, 2005.

Richardson, Kathleen. *An Anthropology of Robots and AI: Annihilation Anxiety and Machines*. Routledge, 2017.

Rieder, John. *Colonialism and the Emergence of Science Fiction*. Wesleyan University Press, 2008.

Rodgers, Diane M. *Debugging the Link Between Social Theory and Social Insects*. Louisiana State University Press, 2008.

Rutsky, R. L. *High Techne: Art and Technology from the Machine Aesthetic to the Posthuman*. University of Minneapolis Press, 1999.

Sandoval, Eduardo Benitez, et al. "Human Robot Interaction and Fiction: A Contradiction." *Social Robotics: 6th International Conference, ICSR 2014, Sydney, NSW, Australia, October 27-29 ; Proceedings*, edited by Michael Beetz et al., Springer, 2014, pp. 54–63.

Scientific American. "Steam Man." *Friends' Review; a Religious, Literary and Miscellaneous Journal (1847-1894)*, vol. 21, no. 29, 1868, p. 463.

Shavit, Ayelet. "Altruism and Group Selection." *Internet Encyclopedia of Philosophy*, www.iep.utm.edu/altr-grp/.

Shelley, Mary Wollstonecraft. *The Complete Frankenstein*. Edited by Pollan Strober, The Literary Folio, 2018.

Sleigh, Charlotte. *Ant*. Reaktion, 2003.

Stableford, Brian. *Science Fiction and Science Fact: An Encyclopedia*. Routledge , 2006.

Sterrett, Susan G. "Turing on the Integration of Human and Machine Intelligence." *SOAR Home*, Wichita State University, 2016, soar.wichita.edu/handle/10057/12539.

Suvin, Darko. "Karel Čapek, or the Aliens Amongst Us." *Metamorphoses of Science Fiction*, Yale University Press, 1979, pp. 270–283.

Symons, Arthur. *The Symbolist Movement in Literature*. Echo Library, 2017.

Szollosy, Michael. "Freud, Frankenstein and Our Fear of Robots: Projection in Our Cultural Perception of Technology." *Ai & Society*, vol. 32, no. 3, 2016, pp. 433–439., doi:10.1007/s00146-016-0654-7.

"The Steam Man." *Scientific American (1845-1908)*, xxii., No. 21., 1870, p. 335.

"The Steam Man" *Scientific American (1845-1908),* lxviii, No. 15., 1893, p. 233.

Turing, Alan M. "Computing Machinery and Intelligence." *The Essential Turing: Seminal Writings in Computing, Logic, Philosophy, Artificial Intelligence, and Artificial Life: Plus the Secrets of Enigma*, edited by B. Jack Copeland, Oxford University Press, 2004, pp. 441–464.

Turing, Alan M. "Intelligent Machinery, A Heretical Theory." *The Essential Turing: Seminal Writings in Computing, Logic, Philosophy, Artificial Intelligence, and Artificial Life: Plus the Secrets of Enigma*, edited by B. Jack Copeland, Oxford University Press, 2004, pp. 472–475.

Turing, Alan M. "Intelligent Machinery." *The Essential Turing: Seminal Writings in Computing, Logic, Philosophy, Artificial Intelligence, and Artificial Life: Plus the Secrets of Enigma*, edited by B. Jack Copeland, Oxford University Press, 2004, pp. 410–432.

Turkle, Sherry. *Alone Together: Why We Expect More Form Technology and Less from Each Other*. Basic Books, a Member of the Perseus Books Group, 2012.

Vartanian, Aram, and Julien Offray de La Mettrie. *La Mettrie's L'Homme Machine: A Study in the Origins of an Idea*. Princeton University Press, 2016.

Veelen, Matthijs Van. "On the Use of the Price Equation." *Journal of Theoretical Biology*, vol. 237, 2005, pp. 412–426., doi:10.1016/j.jtbi.2005.04.026.

Vint, Sherryl. *Bodies of Tomorrow: Technology, Subjectivity, Science Fiction*. University of Toronto Press, 2019.

Waibel, Markus, et al. "A Quantitative Test of Hamilton's Rule for the Evolution of Altruism." *PLoS Biology*, vol. 9, no. 5, 2011, doi:10.1371/journal.pbio.1000615.

Warrick, Patricia S. *The Cybernetic Imagination in Science Fiction*. MIT Press, 1980.

Weller, Chris. “Meet the First-Ever Robot Citizen - a Humanoid Named Sophia That Once Said It Would 'Destroy Humans'.” *Business Insider*, Business Insider, 27 Oct. 2017, www.businessinsider.com/meet-the-first-robot-citizen-sophia-animatronic-humanoid-2017-10?international=true&r=US&IR=T.

Wells, H. G. “The Empire of the Ants.” *H. G. Wells Complete Short Story Omnibus*, Gollancz, 2011, pp. 636–649.

Wiener, Norbert. *The Human Use of Human Beings: Cybernetics and Society*. Da Capo Press, 1954.

Wilson, David Sloan. *Does Altruism Exist?: Culture, Genes, and the Welfare of Others*. Yale University Press, 2015.

Wilson, Edward O. *The Social Conquest of Earth*. Liveright, 2012.

Wilson, Eric G. *The Melancholy Android: On the Psychology of Sacred Machines*. State University of New York Press, 2006.

Wolfe, Gary K. “Icon of the Robot.” *The Known and the Unknown: The Iconography of Science Fiction*, Kent State University Press, 1979, pp. 151–183.

Wong, Julia Carrie. “Elon Musk Unveils Plan to Build Mind-Reading Implants: 'The Monkey Is out of the Bag'.” *The Guardian*, Guardian News and Media, 17 July 2019, www.theguardian.com/technology/2019/jul/17/elon-musk-neuralink-brain-implants-mind-reading-artificial-intelligence.

Wood, Gaby. *Edison's Eve: A Magical History of the Quest for Mechanical Life*. Anchor Books, 2003.

Wosk, Julie. “Metropolis.” *Technology and Culture*, vol. 51, no. 2, 2010, pp. 403–408.

Wosk, Julie. “Update on the Film 'Metropolis'.” *Technology and Culture*, vol. 51, no. 4, 2010, pp. 1061–1062.

Wyndham, John. “The Lost Machine.” *Machines That Think: The Best Science Fiction Stories about Robots & Computers*, edited by Isaac Asimov et al., Henry Holt and Company, 1984, pp. 29–49.

Wright, S. Fowler. “Automata.” *S. Fowler Wright's Short Stories*, by S. Fowler Wright, Sadler Print, 1996, pp. 121–132.

Biography

Jan Sigle, born in 1984 in Pretoria, South Africa, emigrated to the Netherlands in 2004 to pursue a tertiary education at Leiden University. From 2005 to 2009 he completed the *BA in English Language and Culture* program with full marks for his final thesis. His thesis explored the role of consciousness in Daniel Keyes' science fiction novel *Flowers for Algernon* (1966), which he subsequently presented at a conference in Groningen in 2010. From 2010 to 2012, he completed the *Research MA in the Study of Art and Literature* with cum laude and was a finalist for the Best Thesis Prize at Leiden University in 2012. His thesis performed a posthumanist analysis of contemporary science fiction cinematic representations of humanoids. From 2013 to 2014, he worked full-time as an English teacher at a college of higher education. In 2016, he pursued his PhD as an external candidate at Leiden University. From 2015 to 2017, he worked as a part-time lecturer at Leiden University teaching two courses (BA level) and guest lectures (BA and MA levels). In 2018, he presented at a conference at Birkbeck, University of London. He works as a freelancer in academia, proofreading and copyediting.